NTIA Special Publication SP-06-438

Proceedings of the International Symposium on Advanced Radio Technologies

March 7–9, 2006

Patricia J. Raush, General Chair
Kristen E. Novik, Publications

U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary

John M. R. Kneuer, Acting Assistant Secretary
for Communications and Information

March 2006

DISCLAIMER

Certain commercial equipment, components, and software are identified to adequately present the underlying premises herein. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the equipment, components, or software identified is the best available for the particular applications or uses.

INTERNATIONAL SYMPOSIUM
ON
ADVANCED RADIO TECHNOLOGIES

March 7 - 9, 2006 · Boulder, Colorado

Sponsored by:

Institute for Telecommunication Sciences (ITS)
National Telecommunications and Information Administration (NTIA)

National Institute of Standards and Technology (NIST)

U.S. Department of Commerce, Boulder Laboratories

SAFECOM
National Department of Homeland Security (DHS)

RF Globalnet

University of Colorado Department of Interdisciplinary Telecommunications

2004 ISART Technical Committee:

Patricia J. Raush, General Chair, NTIA/ITS

Carolyn G. Ford, Technical Chair, NTIA/ITS

Fredrick Matos, NTIA/OSM

Stephen Berger, TEM Consulting

Tim Brown, University of Colorado

Robert Kubichek, University of Wyoming

Kate Remley, NIST

Robert Stafford, NTIA/ITS

CONTENTS

Page

The Role of the Regulator in Fostering Innovation

Christopher Haslett
Office of Communications (Ofcom)
Phone: +44 (0)20 7981 3750
Email: chris.haslett@ofcom.org.uk

Ofcom, the United Kingdom regulator, is required to secure the optimum use of the spectrum. Whilst the definition of "optimum" when applied to the spectrum could be debated, Ofcom has identified "encouraging innovation" as one of its major themes in pursuit of this requirement. The traditional "command and control" approach to assignment and licensing is not seen as an approach that encourages innovation. However, even prospective, and active, innovators are keen to see that levels of interference do not rise. Ofcom is exploring ways in which the spectrum usage rights granted to licensees can be made less prescriptive, particularly in terms of the technology deployed. The development of such rights should reduce the time for a new technology to be deployed. This involves two major elements: agreeing the nature of such rights to ensure that there is no ambiguity and developing appropriate methods of interference and coverage prediction to ensure that the bounds to these rights are not breached. This presentation gives details of developments made thus far in these areas.

Quality of Service Analysis of Site to Site IPSec VPNs for Real Time Multimedia Traffic

Jesús Pérez, Víctor Zárate, Ángel Montes, Carlos García
ITESM Campus Cuernavaca, Instituto de Investigaciones Eléctricas, Mexico.
{jesus.arturo.perez, vzarate, A00373556}@itesm.mx, cfgarcia@iie.org.mx

Abstract -- *This paper presents a quantitative analysis of the Quality of Service (QoS) offered by Virtual Private Networks (VPNs) based on the suite of protocols IPSec (IP Security) in a Computer Supported Collaborative Learning environment. Testing was done with routers and network nodes which held a communication channel through a videoconference. The objective is to analyze whether the IPSec VPNs are good enough to transmit real time multimedia traffic while protecting the information. Our results show that under stress traffic conditions the IPSec VPN could not offer an adequate QoS to the videoconference in terms of network latency, and that is why new alternative technologies need to be implemented in order to mitigate the degradation. The IPSec VPN technology is evaluated under the latency, jitter and packet loss parameters, which are the basic ones that determine the QoS in a point to point link.*

1. Introduction

Computer networks have evolved to the point where now it is possible to support multimedia applications (voice and video) over corporate networks and the Internet [1]. This has motivated the use of new technologies for the education field, as in the case of the Computer Supported Collaborative Learning (CSCL). The main characteristic of this architecture is the creation of virtual classrooms which allow students and teachers to interact in a real time collaborative way. All of this supported through IP infrastructure.

The use of videoconferences demands Quality of Service (QoS) requirements for voice and video traffic. QoS is defined as the network capacity to offer the best service for a selected traffic flow [2]. This means that voice and video require a minimum delay. Multimedia traffic transmission with a great jitter, low bandwidth and great loss could end up in an unacceptable communication [3].

Voice and video traffic is very sensitive to delay and the service offered by Internet Protocol called "best effort" is not adequate for this kind of traffic.

Nowadays is necessary to include QoS mechanisms over the link; but also protecting the information should be necessary either, especially when transmitting data over Internet or a shared WAN. This is required because Internet is a public network and is susceptible to many attacks. The IP protocol by itself does not protect the data over a public network: packets can be seen within the route towards the destination node, the IP address might be changed and also many other attacks can be mentioned.

In order to prevent and mitigate some attacks, the IPSec protocol suite was developed standard. Without any distinction, IPSec integrates security elements to the IP protocol such as: origin authentication, data integrity, confidentiality, no repudiation and anti-packet repetition [4].

In this context, our work's objective is to evaluate if a standard network infrastructure conformed primarily by routers (like a university campus or a small size network) can maintain the real time voice and video QoS parameters in a good degree while protecting them with an IPSec tunnel.

Many studies have been done to evaluate the IPSec performance but the shown results do not apply to our purposes since our network infrastructure includes routers which create the IPSec tunnels. Also, the data (in our scenario) to transmit is voice and video simultaneously over the same IPSec tunnel; the data is in real time and not buffered, generated by one videoconference using wired media. The results in [5] only focus in voice traffic and not in video traffic. In [6] the test scenario consists in a wireless network and the IPSec tunnel creation is based on desktop nodes. The same happens in [7] where no network layer equipment is included and also the test did not include voice or video traffic. The scenario in [8, 9] includes wireless equipment and no multimedia traffic considered for the results. In [10] the results include streaming voice and video, but this kind of traffic is not real time like the videoconference's traffic since the data is stored in a file before it is sent. In [11] the analysis is done with a different perspective because the evaluation is based on MIPS (Millions of Instructions Per Second) as the metric and not in terms of QoS parameters.

The paper is organized showing in section 2 basic concepts to understand the IPSec VPN operation. In section 3 we explain the QoS parameters. Section 4 shows the bench test where all the testing was implemented (videoconferences). In section 5 we explain our final results, while section 6 shows the conclusions achieved.

2. IPSec Virtual Private Networks (VPNs)

IPSec was designed and created by the IETF as the security architecture for the Internet Protocol IP. It defines the IP packet formats and infrastructure dedicated to provide authentication, data integrity, confidentiality and anti-packet repetition. IPSec might be implemented either in the origin/destination node or in the gateways/firewalls. In our case, the gateways were routers.

IPSec is based on two encapsulation protocols: ESP (Encapsulation Security Payload) and AH (Authentication Header). AH provides origin authentication, data integrity and anti-packet repetition. ESP also provides all characteristics mentioned above and additionally provides confidentiality through data encryption [12] [13].

ESP modifies the original IP packet inserting a new ESP header (after the IP header but before the data payload) and a packet trailer. The ESP header is not encrypted but a section of the trailer and the complete data payload are encrypted (figure 1). The packet authenticated part includes: ESP header, data payload and a trailer section.

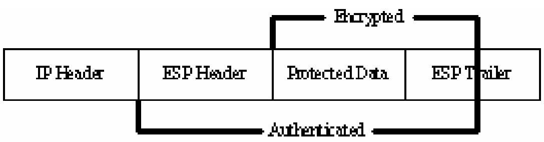


Figure 1. Encapsulated IP packet with ESP

When the destination node receives the IPSec packet, it receives it in clear: the Security Association (SA), the packet sequence number and the hash. This order is because of the same process of receiving that consists in three steps:

1. Sequence number verification.
2. Data integrity verification.
3. Decipher of information.

Before deciphering the information, which is a process that consumes lots of computational resources, the packets need to be checked to see if it did not delay (according with previous received packets) nor repeated. If the packet is valid, the next step is to verify the hash calculation and therefore check that the received information was not modified during transmission by a non authorized user or by a media failure. Finally the information is deciphered using the encryption shared key generated with IKE (Internet Key Exchange Protocol). At this moment the data is ready to be processed.

IKE is a hybrid protocol that gives different services to IPSec such as: IPSec peer authentication, security association agreement and key generation/regeneration for cipher algorithms used by IPSec. IKE negotiates the IPSec SAs. This process requires that IPSec peers get authenticated first with the help of digital certificates or pre-shared keys. After doing this, IKE can take further actions for the negotiation of IPSec SAs.

The next section explains the basic parameters needed to quantify the QoS performance.

3. Quality of Service (QoS)

There are some parameters which are necessary in order to quantify the performance. The usual ones are: latency, jitter and packets loss.

3.1. Latency

For our purposes we considered the latency as the time a packet takes to arrive to the destination network segment (one way latency). We started the count at the moment when packet is put by the origin node in its same network segment and finished the count when the packet arrived to the destination network segment. According to [14] the one way latency for multimedia traffic should not exceed 150ms.

3.2. Jitter

Or also known as packet delay variation, it is the average time that passes elapses between two consecutive packets at the destination node. The higher the jitter, the higher the quality degradation of voice and video will be. It is recommended that jitter should not exceed 50ms.

3.3. Packet Loss

It is the percentage of lost packets during transmission that were not received by the destination node. One of the main causes of packet loss is the network congestion over the links. Packet loss should not exceed 1% of the total transmitted packets.

When the above three parameter recommended limits are exceeded, it does not necessarily mean that the communication will be lost; it means that the quality for voice and video will be degraded in proportion to the exceeded recommended limit.

For this reason, in this paper we try to determine if a standard site to site IPSec VPN with different kinds of traffic is able to hold an appropriate QoS for multimedia traffic and in this way, to observe which are the degradation levels that the network experiments when data is protected with IPSec tunnels.

The next section explains the test scenarios implemented in order to evaluate the performance of the IPSec VPN; it explains the way the routers and the nodes were connected to establish the videoconference, the function of every node and the general aspects considered for the measurements.

4. Test Scenarios and Methodology

The test scenarios were designed according to a real university campus or a small/medium size company network. Our intention is not to simulate a network behavior: we wanted to create a real scenario to measure the performance seen in our topology.

Two scenarios were implemented in order to appreciate the behavior of the real time traffic under different traffic loads using an IPSec VPN: the first one consisted on four nodes; two for the videoconference and the remaining nodes generated TCP-FTP traffic through a large file download of 800Mbytes (figure 2). Videoconf. node A and Videoconf. node B held the videoconference. Node C connected to node E to download the file through a FTP session. We can say that in this configuration the network was an ideal scenario where no congestion occurred.

In test scenario #2 (figure 3) two more nodes were added to generate more traffic including file downloads using the HTTP protocol (with a Web page interface), we also injected high ICPM traffic load through sending large pings (10000 bytes size). In every network segment there was a FTP and HTTP server. Every client was connected with its server to create the connection. Both servers were used to manage the file download but with its correspondent protocol.

In both scenarios the IPSec VPN was implemented within two routers (A and C) creating one IPSec tunnel to protect exclusively the voice and video packets in both ways (figure 2 and 3). All FTP, HTTP and ICPM traffic was not protected with the VPN. Every WAN link was set up to the speed of 1Mbps by using the router's serial interfaces. Two nodes (Videoconf. node A and Videoconf. node B) held a videoconference with VIGO VCON proprietary consoles, cameras and microphones. The protocols for multimedia traffic were G.722 (for voice) and H.263 (for video).

The IPSec cipher specification was AES as the encryption algorithm, HMAC-SHA as the integrity mechanism, ESP encapsulation in tunnel mode, IKE as the key interchange protocol and the IPSec router authentication was made with pre-shared keys.

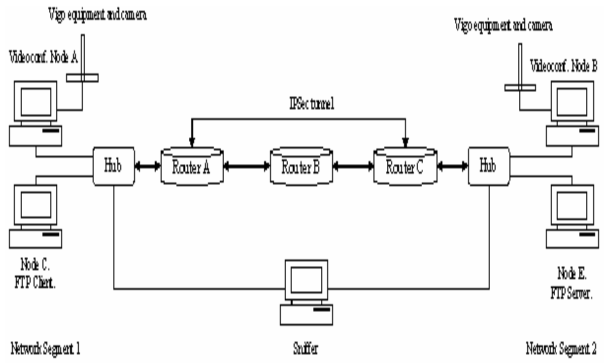


Figure 2. Test scenario #1

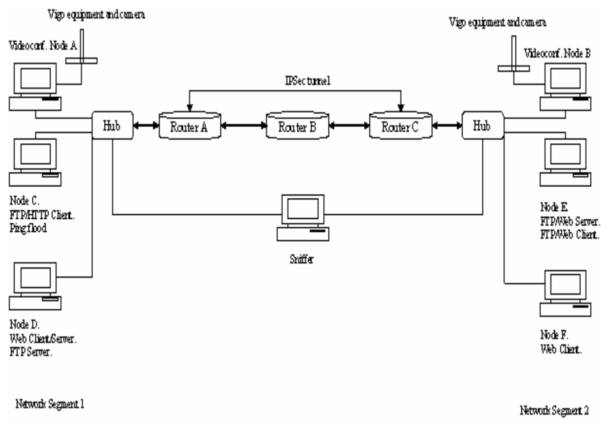


Figure 3. Test scenario #2

Testing consisted on capturing the UDP voice and video packets that traveled from Videoconf. node B to Videoconf. node A with the use of the Etherpeek NX sniffer. We only considered for the results the UDP packets coming from the videoconference. The remaining traffic (FTP, HTTP and ICPM) was discarded since it was just injected for increasing the WAN traffic. The sniffer node captured the incoming and outgoing traffic with two network adapters (NICs).

In order to perform the QoS evaluation, we established twenty real time videoconferences. Every videoconference lasted 2 minutes, during that time all multimedia packets were captured making a total of 40 minutes of videoconference packets captured.

The first ten videoconferences were implemented under test scenario #1, the other ten videoconferences under test scenario #2. In every test scenario, five videoconferences were implemented without the IPSec VPN and five using the IPSec VPN. The results obtained were averaged to reduce the error margin and are shown next.

5. Results Analysis

5.1. Latency

The results of the packets latency are shown on table 1.

Table 1. Results of packets latency

		Latency	
		Voice (ms)	Video (ms)
Scenario 1	With VPN	68.7	71.29
	Without VPN	66.82	63.69
Scenario 2	With VPN	252.18	264.77
	Without VPN	81.26	95.47

In test scenario #1 the latency never exceeded 150ms with and without IPSec VPN. Voice and video packets have close (but different) results producing a good videoconference quality. The main reason of this difference is the encryption process. Encryption demands more CPU and memory space in order to produce protected packets. The AES algorithm did not affected in a considerable way the packet latency but monitoring the router's CPU it used 30% of it. This percentage is vast in terms of resources usage since the router main task is not exclusively encrypting packets, the router also has to route packets towards its destination, create and maintain routing tables and many others.

In test scenario #2 the difference is vast since the traffic load conditions were different. Without the IPSec VPN, the voice and video packets reached around 81 and 95ms of latency respectively. But having the VPN implemented the latency went to 252 and 264ms approximately for voice and video. We attribute this behavior to the traffic load since the multimedia packets have to compete for the access to the serial link. Having a greater traffic load to serve, the packets need to wait more time (buffered) to get access to the medium. The multimedia packets, although they have different needs (latency, jitter and loss), are treated in the same way as a HTTP, TCP or as ICMP packets. There was no preference for this type of traffic.

The encryption process also could be affected since the router's CPU has to process more traffic. Meanwhile the CPU has more tasks to implement and the CPU throughput would be lower, making the encryption to take more time. The encryption process is relevant for increasing the packet latency but it is not the main reason for this latency behavior in test scenario #2.

In general, the video packets took more time to be sent; the reason is because video packets are longer than voice packets. A voice packet occupied 538 bytes, but a video packet average size was 1300 bytes. The serialization delay which is the time that the router takes to put the bits in the medium tends to increase while the packet size gets longer.

5.2. Jitter

The results of the packets jitter are shown on table 2.

Table 2. Results of packets jitter

		Jitter	
		Voice (ms)	Video (ms)
Scenario 1	With VPN	59.96	28.72
	Without VPN	59.97	32.84
Scenario 2	With VPN	60.07	30.28
	Without VPN	60.12	30.98

The jitter for voice packets with and without VPN was higher (almost 10ms) than the recommended 50ms. For the video packets, the jitter with and without VPN remained around 30ms. It can be seen that the encryption process did not influence in the jitter parameter for G.722 and H.263 traffic.

Despite of the 60ms voice jitter average, the end user did not notice a bad voice quality for two main reasons: first, 60ms are not too far away from 50ms and second, the end user listened the voice with hardware dedicated consoles. These consoles have memory for buffering that could compensate the changes in the arrival time. Having buffers for jitter control, the asynchronous arrival time became synchronous, therefore improving the voice and video quality. We can attribute the results similarities in both scenarios due to the constant traffic injection rate and because the FTP and HTTP sessions did not have traffic peaks.

5.3. Packet Loss

The results of the packet loss are shown on table 3.

Table 3. Results of packet loss

		Packet Loss	
		Voice (%)	Video (%)
Scenario 1	With VPN	0	0
	Without VPN	0	0
Scenario 2	With VPN	0	0.01
	Without VPN	0	0

The third important parameter considered was the packet loss. As seen in figure 6 the packets loss was almost null in both test scenarios. The percentage was obtained based on the total amount of packets transmitted by the origin node towards the destination (videoconference nodes).

We can attribute the low increment of packet loss to the IPSec sliding window that detects repeated and out of time packets (similar to the TCP sliding window). All incoming packet must fit inside the sliding window. The traffic load and the encryption process increased the packet latency, therefore there were probabilities that the incoming packets would had not fit in the IPSec sliding window at the correct time. If so, the packet was discarded. Another important factor of the low packet loss was the router's serial interfaces matched speed. There was not speed mismatch between the three routers.

6. Conclusions

Based on the obtained results we can conclude that the QoS in a videoconference using IP infrastructure is affected in two parameters: latency and loss when using IPSec tunnels. The main two reasons of this behavior are the encryption process and the traffic load. Encryption requires great amount of CPU and memory, for this reason the router's manufacturers recommend the installation of VPN accelerator cards which are dedicated exclusively to data encryption/decryption, and in this way, the router is not involved in any other tasks other than routing information, thus increasing the router performance.

With a great amount of traffic, the router buffers the multimedia traffic until the serial link is no longer used. For this reason the latency increments depending on the traffic load. In order to decrease the latency, preferential treatment must be given to this kind of traffic over the remaining traffic.

The jitter parameter was not affected by the VPN. Even though the metric remained a little over the ideal limit with and without VPN, it did not affect the videoconference quality in a visible or audible way. The packet loss percentage changed not much in our test scenarios having or not having the VPN implemented since there was not any interface speed mismatch.

From above reasons, we could deduce that is feasible to implement IPSec VPNs for our collaborative learning architecture over a standard campus network (where the link's speed remain around 1 Mbps) but with medium traffic loads. Additional considerations should be taken when applying to bigger networks such as an Internet service provider, since the behavior is different, having peak times and a higher amount of traffic, resulting in a non-optimum solution. If the videoconference was set in peak times of flow without QoS mechanisms, the quality could be affected seriously.

In highly saturated networks it is necessary to use techniques able to protect and prioritize the information in order to make the traffic transmission secure without affecting the QoS parameters.

7. References

[1] Minoli Daniel, Minoli Emma, "Delivering Voice over IP Networks", Wiley Computer Publishing, 1st edition. 1998. ISBN 0-471-25482-7.

[2] Cisco Systems, "Wireless Quality-of-Service Deployment Guide", 2000, www.cisco.com

[3] Monsour Yishay, Patt-Shamir Boaz, "Jitter control in QoS networks", *IEEE/ACM Transactions on Networking (TON)*, Vol. 9, Issue 4, August 2001.

[4] Naganand Doraswamy, Harkins Dan, "IPSec The New Security Standar for the Internet, Intranets, and Virtual Private Networks", Prentice Hall PTR Internet Infraestructure Series, 1999, ISBN 013011898.

[5] Barbieri R, Bruschi D, Rosti E, "Voice over IPsec: analysis and solutions", *Proceedings of the IEEE Computer Security Applications Conference*, 9-13 Dec. 2002, pp.261 – 270

[6] Wei Qu, Srinivas S., "IPSec-based secure wireless virtual private network"; *IEEE MILCOM 2002 Proceedings* , Vol. 2 , 7-10 Oct. 2002, pp.1107 – 1112.

[7] Khanvilkar S., Khokhar A., "Virtual private networks: an overview with performance evaluation", *IEEE Communications Magazine*, Vol. 42 , Issue: 10, Oct. 2004, pp:146 – 154.

[8] Khatavkar D., Hixon E.R., Pendse, R.,"Quantizing the throughput reduction of IPSec with mobile IP", *Circuits and Systems, MWSCAS-2002*, Vol. 3, 4-7Aug. 2002, pp:III-505 - III-508.

[9] Hadjichristofi G.C., Davis N.J. IV, Midkiff S.F., "IPSec overhead in wireline and wireless networks for Web and email applications", *Conference Proceedings of the IEEE International Performance, Computing, and Communications*, 9-11April 2003, pp. 543 – 547.

[10] Al-Khayatt S., Shaikh S.A., Akhgar B., Siddiqi J., "Performance of multimedia applications with IPSec tunneling", *IEEE Proceedings of International Conference on Coding and Computing*, 8-10 April 2002, pp.134 – 138.

[11] Elkeelany O., Matalgah M.M., Sheikh K.P., Thaker M., Chaudhry G., Medhi D., Qaddour J., "Performance analysis of IPSec protocol: encryption and authentication", *IEEE International Conference on Communications*, ICC, Vol.2, 28April-2May 2002, pp.1164 – 1168.

[12] Kent S. "IP Authentication Header", RFC 2402, IETF Network Working Group, 1998. http://rfc.net/rfc2402.html

[13] Kent S. "Encapsulating Security Payload", RFC 2406, IETF Network Working Group, 1998.http://rfc.net/rfc2402.html

[14] Szigeti Tim, Hattingh Christina, "End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs". Cisco Press, 1st edition, 2004, ISBN 1-58705-176-1.

Introduction to Objective Multimedia Quality Assessment Models

Carolyn Ford, Arthur Webster
Institute for Telecommunication Sciences
National Telecommunications and Information Administration
U.S. Department of Commerce
Boulder, Colorado USA
Phone: (303) 497-3728, fax: (303) 497-6982
cford@its.bldrdoc.gov, awebster@its.bldrdoc.gov

The transmission of multimedia signals over wireless channels has increased exponentially in the past decade. The widespread use of digital technology for the transmission of audio and video signals has led to the need for objective quality assessment methods based on human perception. In particular, the distribution of multimedia signals over wireless links to devices such as laptops, PDAs, and cell phones is widespread. Manufacturers can use objective models to improve products and analyze deployment. Service providers can use objective models to monitor the quality of service they provide. This paper is an introduction to the concepts of multimedia quality assessment models and the design of subjective tests for objective model validation.

1. Introduction

The transmission of multimedia signals over wireless channels has increased exponentially in the past decade. As higher bandwidths become available, applications to fill the bandwidth have increased proportionately. What used to be the realm of analog broadcast television is rapidly becoming dominated by digital transmissions. Digital transmission is used for broadcast applications (e.g., television) as well as point-to-point transmission (e.g., cell phone) of audio and video content.

The widespread use of digital technology for the transmission of audio and video signals has led to the need for objective quality assessment methods based on human perception. Manufacturers can use objective models to improve products and analyze deployment. Service providers can use objective models to monitor the quality of service they provide. Buyers can use objective models to specify requirements, make purchasing decisions and analyze the quality of service received.

Analog methods, such as signal to noise ratio (SNR), while still important, are not sufficient to measure the quality of delivered audio and video signals. Standards for subjective assessment and objective methods for assessing analog signals have been in use for decades, but new subjective and objective methods are needed for both accuracy and ease of use [21]. This paper is an introduction to the concepts of multimedia quality assessment models and the design of subjective tests for model validation.

1.1 Background

Since the early 1990s, the International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) Study Groups 9 and 12, the International Telecommunication Union – Radiocommunication Sector (ITU-R) Study Group 6, and the Alliance for Telecommunications Industry Solutions (ATIS) Performance, Reliability, and Quality of Service Committee (PRQC), and their predecessor organizations[1] have been developing standards for objective measurement of the perceived quality of digital video, speech, and audio. In recent years, perceptual models for the objective measurement of audio quality and video quality have been standardized and put into widespread use. ANSI T1.518-2003 [2] and ITU-T Recommendation P.862 [15] address objective audio quality measurement for narrowband voice and ITU-R Recommendation BS.1387-1 [3] addresses wideband audio. ANSI T1.801.03-2003 [1], ITU-T J.144 [13], and ITU-R BT.1683 [5] provide objective measurement methods of digital video quality for television applications. J.144 and BT.1683 each recommend the same four objective models: the NTIA general model (USA), the British Telecom model (UK), the Yonsei University model (Korea), and the CPqD model (Brazil). These video quality models were validated by the Video Quality Experts Group (VQEG) [www.vqeg.org][2]. Some of these standardized methods are available in commercial products. The NTIA video quality model is available in software with a free evaluation license[3]. It is also described in an NTIA

[1] ITU-R Study Group 6 was formed from ITU-R Study Groups 10 and 11. PRQC was formerly known as T1A1.
[2] The VQEG is an unofficial group formed from video experts of ITU-T Study Groups 9 and 12, ITU-R Study Group 6, industry, and academia.
[3] http://www.its.bldrdoc.gov/n3/video/vqmsoftware.htm

Report [22]. Contact information for the other video models can be found in J.144.

Until recently the development of an objective measure of overall multimedia quality has not been addressed. Multimedia is defined here as the combination of audio and video in the communication of information. Some work has been done combining audio and video quality primarily in subjective experiments [6][7][8][9][10]. The "mapping from the one-way audio and one-way video quality, as derived from audio only and video only subjective experiments, to the one-way overall audiovisual quality [was studied]. Four different laboratories found similar mapping results despite the fact that experimental conditions were quite different." [17]. The laboratories that did this work in the 1990s were Bellcore (USA), KPN Research (Netherlands), NTIA/ITS (USA), and France Telecom/CNET (France). These laboratories did subjective experiments to explore the relationships between audio and video quality. New tests are being conducted to better address audiovisual quality assessment in current multimedia applications.

In 2003 ITU-T Study Group 9 approved Recommendation J.148 entitled "Requirements for an objective perceptual multimedia quality model" [14]. VQEG and the Joint Rapporteur Group on Multimedia Quality Assessment (JRG-MMQA)[4] are currently working on a validation test covering video quality assessment models for multimedia applications. A test plan [19] for VQEG's first phase of multimedia testing (video only) has been approved and testing should begin in 2006. NTIA/ITS is participating in the VQEG tests and is, in addition, conducting independent tests to support our model development.

2. Definitions

Clip - Digital representation of a scene (defined below) that is stored on computer media.

Codec - Abbreviation for a coder/decoder or compressor/decompressor.

Common Intermediate Format (CIF) - A video sampling structure used for video teleconferencing where the luminance channel is sampled at 352 pixels by 288 lines [11].

Feature - A quantity of information associated with, or extracted from, a spatial-temporal sub-region of a video stream (either an original video stream or a processed video stream).

Frame – One complete television picture.

[4] The JRG-MMQA is an official body of the ITU and is formed from members of ITU-T Study Groups 9 and 12.

Hypothetical Reference Circuit (HRC) - A video system under test such as a codec and/or digital video transmission system.

Input Video - Video before being processed or distorted by an HRC (see Figure 1). Input video may also be referred to as Original Video.

Institute for Telecommunication Sciences (ITS) - The research and engineering laboratory of the National Telecommunications and Information Administration, U.S. Department of Commerce.

Mean Opinion Score (MOS) - The average subjective quality judgment assigned by a panel of viewers to a processed video clip.

Multimedia Quality (MMQ) - acronym

Multimedia Quality Metric (MMQM) - An overall measure of video impairment reported by a particular multimedia quality model, either for an individual MM clip (Clip MMQM), or for an HRC (HRC MMQM).

National Television Systems Committee (NTSC) - The 525-line analog color video composite system adopted by the US and other countries (excluding Europe) [18].

Original Video - Video before being processed or distorted by an HRC (see Figure 1). Original video may also be referred to as input video since this is the video input to the digital video transmission system.

Parameter - A measure of video distortion that is the result of comparing two parallel streams of features, one stream from the original video and the corresponding stream from the processed video.

Processed Video - Video that has been processed or distorted by an HRC (see Figure 1). Processed video may also be referred to as output video since this is the video output from the digital video transmission system.

Quarter Common Intermediate Format (QCIF) - A video sampling structure used for video teleconferencing where the luminance channel is sampled at 176 pixels by 144 lines[11].

Scene - A sequence of video frames, with or without accompanying audio.

Video Graphics Array (VGA) – A video format with pixel resolution of 640 x 480.

Video Quality Metric (VQM) - An overall measure of video impairment reported by a particular VQM model, either for an individual video clip (Clip VQM), or for an HRC (HRC VQM). VQM is reported as a single number and usually has a nominal output range from zero to one, where zero is no perceived impairment and one is maximum perceived impairment.

3. Multimedia Quality Assessment

Multimedia quality measurements (i.e. audio and video measurements) characterize either the absolute quality or the degradation in quality of a multimedia signal that has been processed through a system. The system consists of an encoder/decoder pair and a digital channel that connects them. This end-to-end path for a multimedia signal is called an HRC (see definitions). To determine the effect of the system on the multimedia quality, the signal is sampled and calculations are made that have been determined to be significant to perceived multimedia quality. One configuration for making these measurements can be seen in Figure 1.

If the measurement system has access to the full original signal and the full processed signal, the measurement system is referred to as Full Reference (FR). If the measurement system utilizes only an extraction of features from the original signal shared over an ancillary data channel, the system is referred to as Reduced Reference (RR). Figure 1 shows an RR system. Methods also exist which make estimations of quality based upon only the processed signal. This class of measurement method is referred to as No Reference (NR). This paper will describe an example of the RR method. See [12] for a complete description of FR, RR, and NR methodologies.

In the RR measurement method, features are extracted from both the original and processed signals. The features are used to calculate quality parameters. Some of the parameters that are used to produce a quality measurement include the comparison of extracted features to characterize some aspect of the signal (e.g., edge magnitude, edge angle, motion, and local color for video). Features are compared to produce parameters that quantize certain aspects of quality (e.g., blurring, blocking, unnatural motion, color changes).

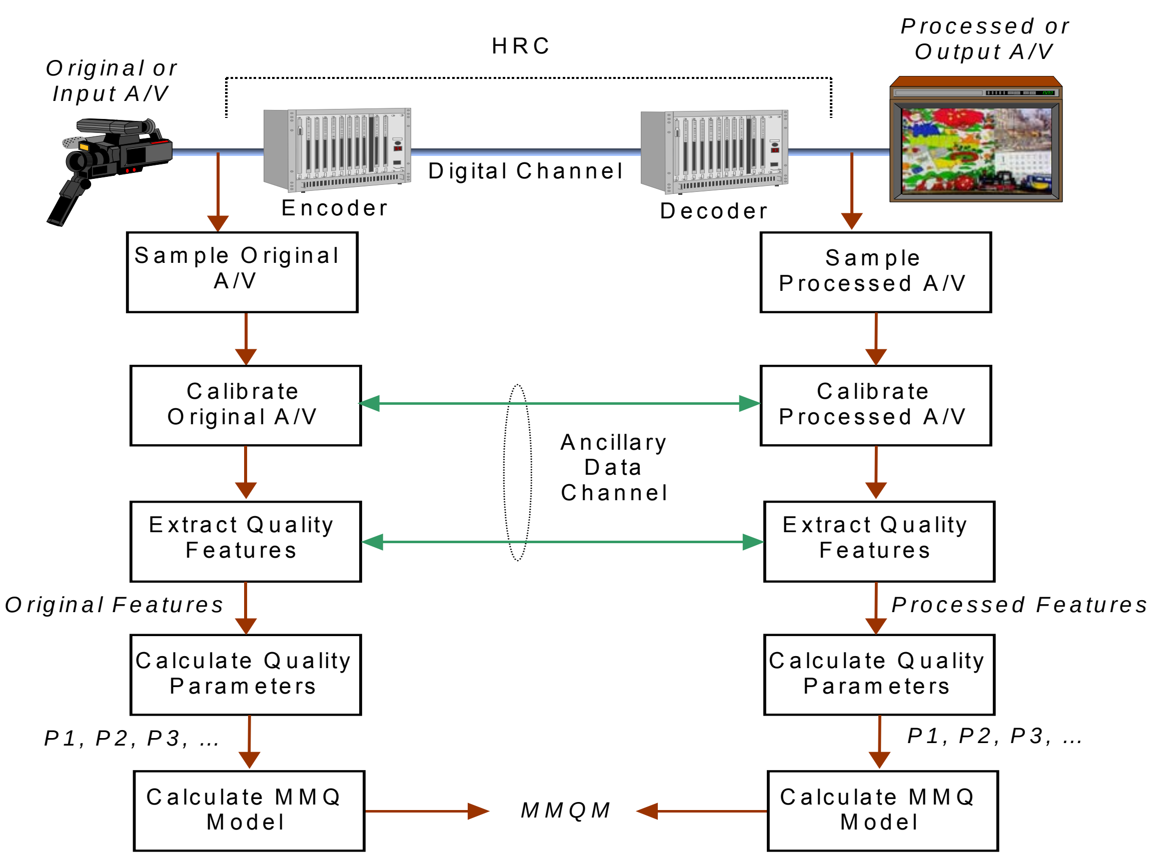


Figure 1. Steps required to compute multimedia quality metric.

3.1 Multimedia Models

The role of an MMQ Model is to predict a human subject's reaction to degradation in quality of a multimedia signal that has been processed through an HRC. Using the measurement process in Figure 1, the quality parameters are calculated. The next step for an MMQ Model is to make a determination of what a person's perceived quality of the processed video would be. This determination is based on the quality parameters and the comparison of parameters calculated from the source sequence with those calculated from the processed sequence.

Figure 1 depicts a multimedia signal being processed as a single entity; however, the channel may affect the audio and video signals differently, so it is necessary to process each signal (audio and video) separately using the process shown in Figure 1. After the audio and video have been processed separately, their parameters must be combined appropriately. This is the challenge of the multimedia quality effort.

3.1.2 Requirements for Multimedia Models [14]

Figure 2 depicts the basic form of a multimedia model. The separate audio and video models provide inputs to the multimedia model. The focus of our research is to define the form of the multimedia quality integration function. The integration function applies specific rules to the information provided by the audio and video inputs. The form of these rules will be based on data derived from subjective quality experiments. The aim is to produce a set of integration rules that enable the multimedia model to accurately predict human quality perception of systems and services under test. Therefore the validity of the model must be shown by comparing the performance of the model against quality ratings obtained from subjective tests for a range of test materials. Subjective testing is discussed in section 4.

The multimedia quality integration function contains three primary input modules. Two modules provide predictions of audio quality and video quality for some multimedia service, and a third provides an indication of the differential delay between the audio and video sources. Standardized models will be used as the quality inputs to the multimedia model. There is ongoing research in developing perceptual quality models for audio and video with full reference, reduced reference, and no reference capabilities. A fourth input allows the model to accommodate any task-dependent influences that may impact quality perceptions. One role of the task input module is to represent the degree of interactivity associated with the multimedia service. This may also be a more comprehensive control, such as choosing specialized audio and video quality models that are best suited to the current task (e.g., a speech quality model suits a videoconferencing task). The multimedia model will be based on generic rules that capture human perceptions of audiovisual quality.

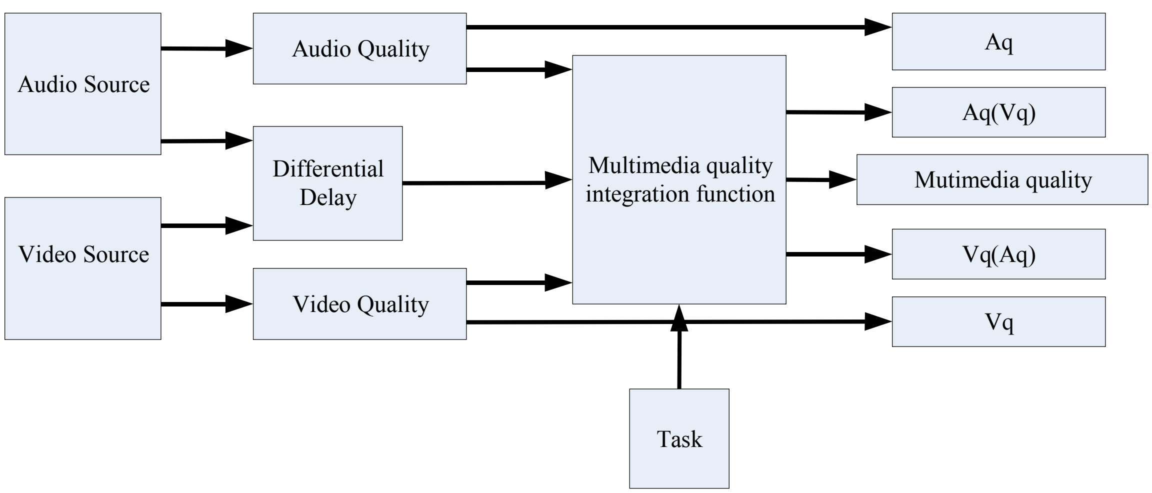


Figure 2. Basic components of a multimedia model

3.1.3 Multimedia Quality Integration Function

The multimedia quality integration function needs to accommodate human perceptual and cognitive processes active in the formation of quality judgments of audiovisual services. Data obtained from subjective quality tests will be used to define the integration function. The role of the integration function is to accept inputs from the audio and video models, and apply some predefined rules to the incoming data to produce a multimedia quality prediction.

The integration function will account for human perceptual responses to multimedia services. In particular, the integration function will apply rules that represent elementary perceptual processes present in subjective quality assessment of multimedia. The complete multimedia model provides five outputs. The primary output is a predicted measure of overall multimedia quality. Four subsidiary outputs provide predictions of perceived quality for the audio (denoted Aq), video (denoted Vq), audio accounting for any influence the video may have (denoted Aq(Vq)), and video perceptual quality accounting for any influence the audio may have (denoted Vq(Aq)).

4 Subjective Testing

As discussed in section 3, objective quality assessments are based on data derived from subjective quality experiments. The validity of any model must be shown by comparing the performance of the model against quality ratings obtained from subjective tests for a range of test materials. This section presents some background on subjecting testing, and describes the design and execution of a subjective test for video for multimedia applications, as an example.

Subjective testing employs human subjects to rate multimedia quality. The results are used to train and test the objective measurements calculated by the models in order to improve audio/video codec quality, to analyze the interaction between network loading and received quality, and to compare the performance of different displays, coders, decoders, and networks. When carefully performed, subjective testing provides “truth data” – people’s actual opinions. Care must be taken, however, or the subjective testing will result in misleading or incorrect conclusions. ITU-R Rec. BT.500 [4] and ITU-T Rec. P.910 [16] address the methodology for subjective testing of video, and the ITU-T has published several of VQEG’s test plans and reports in a 200-page tutorial [20]. The VQEG tutorial addresses all aspects of designing and executing a multimedia subjective test, including test material, test methods, viewing conditions and data analysis. Researchers planning subjective testing are encouraged to review these documents. The ITU Recommendations describe several different options for subjective testing in order to answer different kinds of questions.

4.1 An Example Video Subjective Test

ITS is performing a subjective video test as the first step in the multimedia test process. The test described is the first of a series of multimedia subjective tests that will explore the relationships between the quality parameters for audio and video, as discussed in section 3. The data collected from this subjective test will be used initially to evaluate the effect of video resolution on perceived quality. The results will also contribute to the Vq function parameter in the ongoing multimedia quality assessment modeling effort and multimedia quality integration function research. The subsequent subjective tests will combine video and audio.

This test employs the Absolute Category Rating (ACR) method as documented in P.910 and as amended in the VQEG’s multimedia testplan [19]. The subject rates both processed clips and their associated original clips; however, this method requires the test subject to rate clips individually, with no direct comparison to the original clip. Original and processed clips are combined into a set, and the clips in the set are shown in a randomized order.

4.1.1 Video Clip Selection

ITS collected a large set of high quality original scenes. The original video clips were transmitted through different configurations of video transmission systems to create processed video clips. The resulting set of video clips, original and processed, spanned a broad range of quality from excellent to bad. Each scene was scaled to the resolutions VGA, CIF and QCIF. VGA resolution is approximately that of a television show viewed at full resolution on a PC monitor. CIF resolution is approximately one-fourth the size of VGA, and is typically used by PDAs and some web sites. QCIF resolution is one-fourth the size of CIF and is typically used by cell phones and video on web sites intended for very low bandwidth connections. Each subject was shown the same scene content in each resolution, in varying orders of resolution (VGA-CIF-QCIF, CIF-VGA-QCIF, etc.). Subjects viewed each video clip in turn and indicated their opinion of the video quality.

The processed clips were originally intended to be viewed on a television monitor in NTSC format. These sequences had previously been rated in subjective tests to generate mean opinion scores (MOS). These subjective scores enabled ITS to select original video scenes that spanned a wide range of coding difficulty, from nearly still scenes containing very little detail, to complicated scenes containing multiple scene cuts and fast motion. The video clips were converted to VGA, CIF and QCIF using the approximate area a viewer would see on a television screen. The clips were chosen so that the viewed set would contain clips with a uniform distribution of MOS scores.

Original video sequences should be selected carefully. The content displayed should be visually different, so that subjects do not become bored. Content should cover a wide variety of material spanning a wide range of difficulty. For example, if all of the source scenes are head and shoulder shots with very little motion, then the codecs will respond identically to all content, and little is learned from including multiple scenes. The better approach is to examine the amount of motion and detail in the source scenes, and then choose scenes that contain different amounts of motion and detail. For example, a head and shoulders videoconferencing scene might contain very little spatial detail and be nearly still, while a movie trailer will contain rapid motion, multiple scene cuts, and highly detailed scenery.

Processed video sequences and systems to be tested should be selected carefully to cover a wide range and type of quality degradation. Taken together, all of the original and processed video clips should evenly span a moderate to wide range of quality

4.1.2 Viewing Environment

The test is administered using personal computers (PC's). Video clips are displayed at random on the PC monitor, and the subjects are given unlimited time to rate each clip before the next one is shown.

Any number of identical PC/monitor combinations can be used, since the test is administered individually, and at the test subject's pace. We use two PC/monitor setups. The PC system must be capable of playing the video sequences with no skips or errors introduced by the testing system, and each system used for testing should be identical. For the test under discussion, the systems have the following specifications:
Monitor: 19" LCD with response time of 12 ms and a native resolution of 1280x1024 non-interlaced.
PC: 3.2GHz dual core processor, 2GB Dual Channel SDRAM, 256 MB PCI Express x16 graphics card, 2x80 GB RAID 0 hard drives at 10,000rpm.

Mplayer[5,6] was chosen as the video player, and ITS developed internal software in Matlab™[7], called NTIA Subjective Tester, to control the test flow. After each 8-second clip is played, the dialog box shown in Figure 3 is displayed for subject scoring (each dialog box displays the number of the current clip).

[5] http://www.mplayerhq.hu/

[6] Certain commercial equipment and materials are identified in this report to specify adequately the technical aspects of the reported results. In no case does such identification imply recommendations or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is the best available for this purpose.

[7] http://www.mathworks.com/

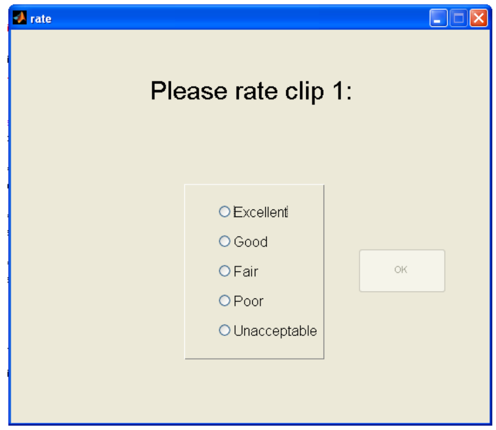


Figure 3. Subjective test scoring dialog box

Once a radio button is selected, the OK button becomes active. After the OK button is selected, the next video clip is played.

ITS is currently executing this subjective test. The results will be reported to VQEG, JRG-MMQA, ITU-T Study Groups 9 and 12, and ITU-R WP6Q.

5. Summary

With the advent of digital transmission of multimedia content, especially over wireless links, methods need to be developed to monitor and measure the quality of the received content. While models and standards exist for objectively measuring the quality of audio and video independently, there is a current need for analogous tools for multimedia – the combination of audio and video. Measurement methodologies need to be adapted from the current standards, and subjective tests need to be conducted in order to develop and train quality assessment models to predict the perceived quality of multimedia signals. Research is currently being conducted to increase the knowledge and standard practice for assessing multimedia quality. These efforts are expected to result in new standards in the ITU to assist industry and users in the measurement of the quality of multimedia applications.

6. References

[1] ANSI T1.801.03-1996, 2003, "American National Standard for Telecommunications - Digital Transport of One-Way Video Signals - Parameters for Objective Performance Assessment."

[2] ANSI T1.518-1998, 2003, "Objective Measurement of Telephone Band Speech Quality Using Measuring Normalizing Blocks (MNBs)"

[3] ITU-R Recommendation BS.1387-1 (2001), “Method for objective measurements of perceived audio quality,” Recommendations of the ITU, Radiocommunication Standardization Sector.

[4] ITU-R Recommendation BT.500-11 (2002), “Methodology for the subjective assessment of the quality of television pictures,” Recommendations of the ITU, Radiocommunication Standardization Sector.

[5] ITU-R Recommendation BT.1683 (2004), “Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference,” Recommendations of the ITU, Radiocommunication Standardization Sector.

[6] ITU-T COM 12-20 (December 1993), *Experimental combined audio/video subjective test method*, Bellcore.

[7] ITU-T COM 12-37 (September 1994), *Extension of combined audio/video quality model*, Bellcore.

[8] ITU-T COM 12-19 (February 1998), *Relations between audio, video and audiovisual quality*, KPN Research, Netherlands.

[9] ITU-T COM 12-64 (November 1998), *Results of an audiovisual desktop video teleconferencing subjective experiment*, USA.

[10] ITU-T COM 12-61 (November 1998), *Study of the influence of experimental context on the relationship between audio, video and audiovisual subjective quality*, France Telecom/CNET.

[11] ITU-T Recommendation H.261, “Video codec for audiovisual services at p x 64 kbit/sec,” Recommendations of the ITU, Telecommunication Standardization Sector.

[12] ITU-T Recommendation J.143 (2000) “User requirements for objective perceptual video quality measurements in digital cable television,” Recommendations of the ITU, Telecommunication Standardization Sector.

[13] ITU-T Recommendation J.144 (2004), “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,” Recommendations of the ITU, Telecommunication Standardization Sector.

[14] ITU-T Recommendation J.148 (2003), “Requirements for an objective perceptual multimedia quality model,” Recommendations of the ITU, Telecommunication Standardization Sector.

[15] ITU-T Recommendation P.862 (2001), “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” Recommendations of the ITU, Telecommunication Standardization Sector.

[16] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications,” Recommendations of the ITU, Telecommunication Standardization Sector.

[17] ITU-T Recommendation P.911, “Subjective audiovisual quality assessment methods for multimedia applications,” Recommendations of the ITU, Telecommunication Standardization Sector.

[18] SMPTE 170M, “SMPTE Standard for Television – Composite Analog Video Signal – NTSC for Studio Applications,” Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.

[19] VQEG Multimedia Testplan, http://www.its.bldrdoc.gov/vqeg/projects/multimedia/index.php

[20] A. A. Webster, ed. “Objective perceptual assessment of video quality: full reference television”, http://www.itu.int/ITU-T/studygroups/com09/docs/tutorial_opavc.pdf

[21] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, “An objective video quality assessment system based on human perception,” *Human Vision, Visual Processing, and Digital Display IV, Proceedings of the SPIE*, Vol. 1913, Feb. 1993, pp. 15-26.

[22] S. Wolf and M. H. Pinson, “Video Quality Measurement Techniques,” NTIA Report 02-392, Jun. 2002.

A Network and Data Link Layer Infrastructure Design to Improve QoS in Voice and Video Traffic

J. A. Pérez, V. H. Zárate { jesus.arturo.perez, vzarate }@itesm.mx Department of Electronics ITESM - Campus Cuernavaca. Temixco, Morelos, 62589 México	J. Jenecek janecek@cs.felk.cvut.cz Department of Computer Science and Engineering Czech Technical University in Prague Prague, Czech Republic

Abstract

Currently, there are a lot of e-learning and collaborative platforms to support distance and collaborative learning, however, all of them were designed just like an application without considering the network infrastructure below. Under these circumstances when the platform is installed and runs in a campus, sometimes it has very poor performance. This paper presents a network and data link layer infrastructure design that classifies and prioritizes the voice and video traffic in order to improve the performance and QoS of the collaborative systems applications. This infrastructure has been designed taking in consideration a typical network of a university campus, so that in this way it can be implemented in any campus. After making the design we have made some tests in a laboratory network demonstrating that our design improves 70-130% the performance of these real time collaborative systems which transmit voice and video.

1. Introduction

There are some applications that support collaborative work; however, this does not imply that they are neither effective nor functional. The mayor issues in those applications are focused in the synchronous collaboration because of the problem of managing the information that flows across the network and the mechanism to ensure the quality of the service.

Applications like the "Elluminate Live Academic Edition" provide some tools to work in real-time with other people across the internet; it provides mechanisms to create virtual conferences based in audio and text with pretty good audio quality and also good application performance, even with a low bandwidth connection. It is important to say that this application is not academic and it requires the purchase of a license to allow people to use it. Also we should note, the technology used to ensure the quality of the communication is private [1].

Other applications, like "Synergeia" [2], "Synergo" [3] and "Blackboard" [4], which are designed to operate under the common structure of the internet (client-server), do not provide efficient tools to communicate with other people across the internet in a synchronous manner. The problem in this case is the lack of control and management in the underlying protocols to achieve the demanded Quality of Service (QoS). These issues force this kind of applications to provide only tools that do not exhaust the bandwidth of the communication channel like text based conferences (chat) and shared blackboards or notepads, tools that do not demand a constant flow of information, just chunks of data (messages).

Anyway, even though if the Elluminate software found a way to guarantee some quality in the voice transfer, it does not ensure an efficient management for all types of data. In fact, none of the applications mentioned previously provide tools to interact in a video conference, or to request video on demand. This is because the video information is more sensible to degradation during the transference and consequently, it is more difficult to control the QoS in this kind of data.

All the analyzed collaborative systems work properly when they are exchanging data through a chat or when using off-line communication like e-mail. However, in those which support real time voice and video transmission, the performance and success of the application is conditioned to the performance of the network below. We have discovered that in many cases modern networks are fast enough to support these applications, but the lack of a proper configuration in routers and switches make the applications suffer from performance issues.

Because synchronous communication is the most difficult to implement [5], in order to guarantee the QoS in data transmission independently of the data type we need to provide a designed framework to manage the data flow in the network and data link layer in the OSI Model.

2. Video and Voice requirements

The audio/video information within a videoconference is segmented into chunks by the application, encoded and compressed, put into a series of data packets and sent over the network to the remote end at basically constant intervals. The data packets may arrive at their destination at slightly varying times, and possibly out of order. In order to keep the "real time" impression of an interactive videoconference, the packets must arrive on time and in time to be re-ordered for delivery through the videoconferencing terminal.

Before proceeding, it is important to involve the five fundamental network problems for videoconferencing and for the transmission of voice over IP (VoIP) [6].

1. **Bandwidth** is the fundamental requirement. There must be enough room in a network path for all of the packets to get through unimpeded. This bandwidth need is symmetric: each end requires to transmit and to receive at the same link speed.
2. **Packet loss** is the amount of packets which fail to arrive correctly to their destination. This is due to insufficient bandwidth, transmission errors or high latency. The packet loss percentage must always be below 1% for voice and 2% for video in order to guarantee an understandable communication.
3. **Latency** is the time delay between an event occurring on one site and the remote end seeing it. Latency is introduced both by the encoding/decoding process, and hence depends on the equipment used, and also by the time it takes packets to traverse the network. A disruption in the image can cause a bad playing in the destination, but a disruption in the voice is more important since it makes the transmission not understandable, so it is considered that the biggest latency allowed in the voice transmission to keep a good quality should always be below the 150 ms value.
4. **Jitter** is the average time variation among each received packet. Jitter should always be below 50ms. Multimedia traffic transmission with a great jitter, low bandwidth and great loss could end up in an unacceptable communication [7].
5. **Policies** are introduced by devices like firewalls and network address translation (NAT) controllers that are generally used to hide or protect network elements from the wider Internet.

In this work we are considering that the network of the Autonomous System (AS) where the designed infrastructure is going to be implemented has enough bandwidth for voice and video traffic. We will focus on creating a configuration to minimize the packet loss, latency and jitter for videoconference traffic.

3. General model for traffic prioritization

In order to maintain the high standards that modern applications and collaborative systems require, traffic should always follow a prioritization scheme in order to guarantee specific bandwidth requirements from real time communications, such as voice and video. This scheme can be represented in the form of a general model which applies to all applications which require special conditions (such as maximum delay) to be met. This general model is represented in Figure 1.

After receiving the IP traffic, the first step would be to mark the incoming frames/packets according to our needs. This marking should be done in the way of traffic *classes*. A different class should be specified for every kind of traffic which should be treated in

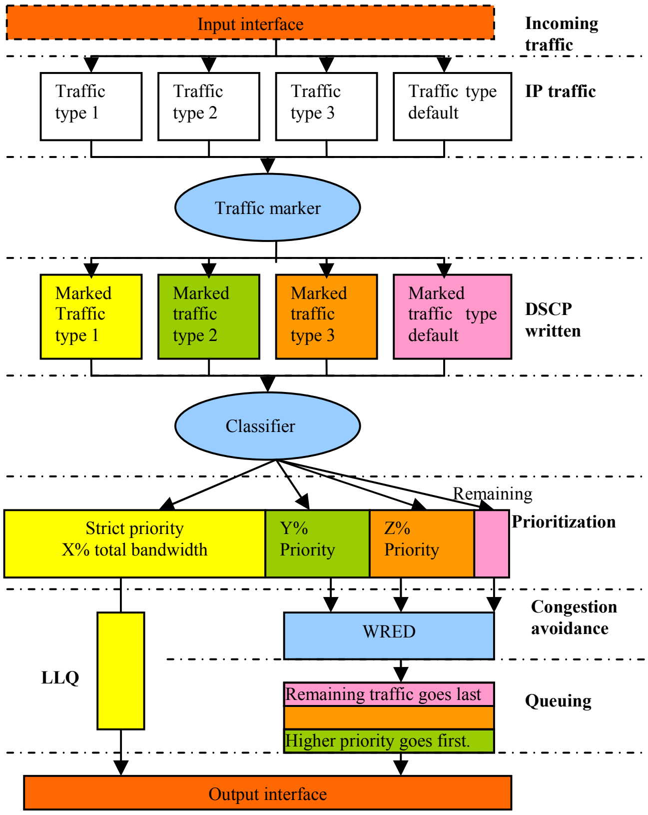


Fig 1. General model for prioritizing network traffic

different way. A common practice is to classify the voice and video traffic in its own class, away from any other type that might cause delays in the processing of the realtime data.

After traffic has been marked, it is ready to be classified according to our own requirements. Since voice and video are the most delay sensitive type of data, they should receive a special treatment in order to avoid delay at all costs. Any other type would be considered as delay tolerant and so, it would be subject of further processing in order to provide the bandwidth only to those applications that do really need it.

The most important traffic class should receive a *strict priority* using Low Latency Queuing (LLQ). LLQ provides traffic the ability to skip directly to the output interface without having to deal with any congestion avoidance technique, reducing its time to go out from the router. By also specifying a reasonable amount of the total bandwidth, we will be guaranteeing that this type of traffic always has the resources it needs to function properly.

As it is shown in the general model (see figure 1), the rest of the traffic classes must go through WRED congestion avoidance mechanisms and queuing. This process would divide the remaining bandwidth according to the specific policies configured for each data class.

Once all conditions are met and all the policies are applied, the now marked and prioritized traffic is sent through the router's outgoing interface to its destination.

Now that we have shown the general model, we will describe in detail what we propose to improve the performance in each layer of our model.

3.1 Improving data link layer

A shared LAN (using hubs) divides the bandwidth between all the available users, so on average we get much less of the nominal bandwidth, plus increasing the risk of packet loss and jitter due to collisions, while a switched network allows full duplex transmission by using microsegmentation. This totally avoids collisions and provides the highest possible available bandwidth for each of the devices connected to the switch. As additional enhancements, switches open up the possibility of using VLANs for increasing security and reducing broadcast domains, while also allowing the use of trunk interfaces for extending the availability of ports for other medical devices.

In order to improve the performance if the switched network, at layer two, we need to configure in the switches their switch mode operation and traffic prioritization with 802.1p.

3.1.1 Switch mode Operation. How a frame is switched from the source port to its destination is a trade off between latency and reliability. A switch can start to transfer the frame as soon as the destination MAC address is received. This switching method is called cut-through and results in the lowest latency through the switch. However, no error checking is available, but considering the type of application, it is more important to transfer frames faster than to lose some frames. So the switch network infrastructure must support cut-through mode instead of store and forward (or fragment-free modes). The switch cut-through command must be entered for each of the switch's port where cut-through mode should be used.

3.1.2 Traffic Prioritization with 802.1p. When we are using a switched network inside a university campus, sometimes most of the traffic never cross the routers interfaces, if VLANs are used around the campus and the traffic is sent among users of the same VLAN the traffic will never cross the router interfaces, so, there is not a way to prioritize the traffic with layer 3 priorities, for this reason we need to add to our designed infrastructure layer 2 priorities.

The IEEE 802.1p is an extension of the IEEE 802.1Q (VLANs tagging) standard. The 802.1Q standard specifies a tag that appends to an Ethernet MAC frame. The VLAN tag has two parts: The VLAN ID (12-bit) and Prioritization (3-bit), as it is shown in the figure 1. The prioritization field was not defined and used in the 802.1Q VLAN standard. The 802.1P defines this prioritization field.

Using frame tagging as the standard trunking mechanism, as opposed to frame filtering, provides a more scalable solution to VLAN deployment. VLAN frame tagging is an approach that has been specifically developed for switched communications and gives the possibility of using the prioritization field.

The 802.1P standard also offers provisions to filter multicast traffic to ensure it does not proliferate over layer 2-switched networks. The 802.1p header includes a three-bit field for prioritization, which allows packets to be grouped into various traffic classes. It can also be defined as best-effort QoS (Quality of Service) or CoS (Class of Service) at Layer 2 and can be implemented in network adapters and switches without involving any reservation setup. 802.1p traffic is simply classified and sent to the destination; no bandwidth reservations are established.

IEEE 802.1p establishes eight levels of priority. Although network managers must determine actual mappings, IEEE has made broad recommendations. The highest priority is seven, which might go to network-critical traffic such as Routing Information Protocol (RIP) and Open Shortest Path First (OSPF) table updates. Values five and six might be for delay-sensitive applications such as interactive video and voice. Data classes four through one range from controlled-load applications such as streaming multimedia and business-critical traffic - carrying SAP data, for instance - down to "loss eligible" traffic. The zero value is used as a best-effort default, invoked automatically when no other value has been set.

In order to apply the prioritization to be applied throughout the network, switches and NICs that support 802.1p are needed. Considering these priorities we should establish in our switches the five and six value for our traffic in this way, we will enable to prioritize network traffic inside our network campus even though this traffic do not cross router interfaces.

Incoming frames can be examined for a pre-existing priority value, which is then mapped to the 802.1p-specific priority value (according to a matrix provided in the 802.1p specification). The 802.1p priority value can then be assigned to an outbound frame on another medium using this same matrix, providing a standard and topology-independent priority-mapping service. The matrix is used only when a frame is transmitted from one medium or technology to a different one since the priorities are not the same in each medium. In this way also 802.1p allows to provide priority to technologies that for nature did not supported, like Ethernet.

Depending on the switch model, it may be necessary to first activate QoS with the mls qos command. This command is required on both the Catalyst 3550 and the Catalyst 6500. The Catalyst 2950 has QoS enabled by default.

```
switch(config)#mls qos
switch(config-if)#mls qos trust cos
switch(config-if)#mls qos cos 0
```

When configuring the switch you can decide to trust or not in the CoS value of the end device with the configuration above any 802.1Q/p frames that enter the switch port will now have its CoS passed, untouched, through the switch. If an untagged frame arrives at the switch port, the switch will assign a default CoS to the frame before forwarding it. By default untagged frames are assigned a CoS of zero.

In the most of cases it is better not to trust in the CoS value of the end device and force to the switch to assign an specific priority for a specific switch port. Specially it is useful when we know that we will connect a high priority devices to these ports, the better configuration for these ports considering they will received voice traffic is:

```
switch(config)#mls qos
switch(config-if)#mls qos cos 5 override
```

For the rest of the ports which are not being prioritized, the same configuration can be applied, except the CoS value will be reset to the lowest priority, the zero value.

Following these brief rules will create a faster infrastructure in the data link layer to allow faster voice and video transmission,— this is sometimes described as "layer 2 quality of service".

3.2 Improving network layer

When we are willing to provide QoS for traffic that will flow outside of our own LAN, there is the need to specify priorities at layer 3 in order to obtain the desired latency and bandwidth for specific delay sensitive data.

QoS refers to both class of service (CoS) and type of service (ToS). The basic goal of these is to guarantee specific bandwidth and latency for a particular application [8]. To achieve this, we use the Differentiated Services Codepoint (DSCP) field in the packet header to indicate the desired service. This value provides the necessary marking as suggested by the first step of our general model (Figure 1) when dealing with layer 3 traffic.

DSCP redefines the older IPv4 ToS octet and IPv6 traffic class octet. It is composed by the first six bits in the ToS byte, while the IP Precedence value is created with the first three bits in the ToS value. The IP Precedence value is actually part of the IP DSCP value, so both values can not be set simultaneously. If both values are set simultaneously, the DSCP value overwrites the IP precedence one.

The marking of traffic at layers 2 or 3 is crucial to providing QoS within a network, and the decision of whether to mark traffic at any or both of these layers is not trivial. We suggest deciding after the following considerations are made:

- Layer 2 marking can be performed for non IP traffic. This is the only option available for non “IP aware” switches.
- Layer 3 marking will carry the QoS information end-to-end.

We propose to use both DSCP to mark packets through the routed links of the network and also mark the frames using CoS to allow layer 2 devices to provide the QoS requirements of packet at the data link layer.

It is important to mention that a mapping between layer two QoS (CoS) and layer three QoS (DSCP) is possible, as it is presented by Ubik [9] However, since in this paper we are just trying to improve the QoS inside our Autonomous System, we will only propose tools associated with the network edge.

After marking the packages classification will be needed in order to create different classes of traffic with different priority.

3.2.1 Low bandwidth WAN circuits. If any low speed connections exist in the network, and a high portion of the traffic is from the RTP kind, the most proper protocol to use is the Compressed Real Time Transport Protocol cRTP which reduce the consumed bandwidth by a g.729 voice call since it enables to compress the 40 byte IP/RTP/UDP header to 2 or 4 bytes.

With cRTP the amount of traffic per VoIP call is reduced from 24Kbps to 11.2Kbps, in this way it is possible to double the amount of the calls in one link. In our experiments we will use Cisco equipment, so a basic configuration for CISCO IOS could be:

```
Interface serial 0
   Ip address 192.168.10.1 255.255.255.0
   Ip rtp header-compression
   Encapsulation ppp
```

cRTP is not required to ensure good voice quality. It is a feature that reduces bandwidth consumption. cRTP must be configured on both ends of the link. After Configuring cRTP in the WAN link, if all other conditions are met then the voice quality will be good.

Another important consideration is fragmentation, large packets takes a long time to move across low-bandwidth links (In cases with a WAN link of more than 768 Kbps, the fragmentation feature is not needed), and they can consume the entire VoIP budget. Fragmentation allows to reduce the latency of packets, but the router must also be able to queue based upon fragments or smaller packets instead of the original (prefragmented) packet.

By default with G.729, two 10-ms speech samples are put into one frame. This gives you a packet every 20 ms. This means you need to transmit a VoIP packet out of the router every 20 ms. This means you need to be able to transmit a VoIP packet out of the router every 20 ms.

Blocking directly affects the delay budget, so it is always desirable to keep the blocking delay at 80 percent of your total voice packet size. So in our case we have a 20 ms seconds packet so the maximum blocking delay must be 16 ms. Now, we need to determinate the exact packet fragmentation size for the links we could have in our collaborative environment with the following algorithm:

WAN bandwidth X blocking delay = fragment size in bits

The low bandwidth circuits that we could support are a dial up 56Kbps link or ADSL 256Kbps link, so applying the last algorithm we have:

Fragment size Dial-up link = 56Kbps x 16 ms = 896 bits per second = 112 bytes per second.
Fragment size ADSL link = 256Kbps x 16 ms = 4096 bits per second = 512 bytes per second

As we can see, in the low bandwidth WAN link is it necessary to fragment the packets to 128 or 64 bytes for the dial up connection and to 512 bytes to the ADSL connection.

In order to fragment the packets we can use FRF.12 if we have a frame-relay interface, if we have interfaces that can run PPP, MCML is recommended otherwise we should use IP MTU, even though this last tool can cause many problems since the receiving station's overall performance is affected.

MCML PPP still requires fragments to be classified by IP Precedence, and to be queued by WFQ. A possible configuration could be like this:

```
Router(config)#interface serial 0/0
Router(config-if)#no ip address
Router(config-if)#encapsulation ppp
Router(config-if)#ppp multilink
Router(config-if)#ppp multilink group 1
Router(config-if)#shutdown

Router(config)#interface multilink 1
Router(config-if)#ip address 10.0.100.1
   255.255.255.0
Router(config-if)#fair-queue
Router(config-if)#Ip rtp priority 16384
   16484 50
Router(config-if)#bandwidth 128
Router(config-if)#ppp multilink fragment-
   delay 16
Router(config-if)#ppp multilink
   interleave
```

In this configuration, PPP multilink is configure on the serial 0/0 interface. A new multilink-group 1 interface is then created with IP RTP Priority configured along with MCML PPP and WFQ. Under the serial 0/0 interface, multilink-group 1 maps to interface multilink 1. This enables all the interface multilink 1 attributes to apply to the serial 0/0 interface.

In the Interface Multilink 1 configuration, the fragment-delay of 16 instructs the router to break up any packet into fragments that will not take longer than 16 ms to cross the WAN link. The interleave command will allow PPP to interleave new packets subject to whatever queuing strategy is in place.

For RTP traffic prioritizing at layer 3 over normal bandwidth WAN circuits, our general model proposes the use of Low Latency Queuing (LLQ) to give absolute priority to voice and video traffic over any other traffic over an interface.

3.2.2 Low latency queuing and congestion avoidance techniques. Low latency queuing (LLQ) was designed for being used in realtime applications, such as a videoconference. It brings strict Priority Queuing (PQ) to CBWFQ. Strict PQ allows delay-sensitive data such as voice to be sent directly through the outgoing interface before packets in other queues are sent (as shown in Figure 1). Without LLQ, CBWFQ provides

WFQ based on defined classes with no strict priority queue available for real-time traffic. For CBWFQ, all packets are serviced fairly based on weight and no class of packets may be granted strict priority. This scheme poses problems for voice traffic that is largely delay intolerant, especially delay variation. For voice traffic, variations in delay introduce irregularities of transmission manifesting as jitter in the heard conversation. LLQ provides strict priority queuing for CBWFQ, reducing jitter in voice conversations.

To configure LLQ priority to a class within a policy map, we need to define a class to match the desired traffic. After this, a policy must be created to specify the allowed bandwidth for each class. As an example:

```
R1(config)# class-map match-any
   VoiceTraffic
R1(config-cmap)# match protocol rtp audio
R1(config-cmap)# match protocol rtp video

R1(config)#policy-map StrictPriority
R1(config-pmap)#class VoiceTraffic
R1(config-pmap-c)#strict priority 128
R1(config-pmap-c)#class class-default
R1(config-pmap-c)#fair-queue
```

This would create a class named VoiceTraffic which matches any rtp audio or video packet. Then, the policy would give it a 128kbits bandwidth while also setting the fair-queue as queuing scheme for any other kind of traffic.

When LLQ is not possible to configure, CBWFQ is the best solution, since we can create a specific class and then assign a specific bandwidth that will be enough to guarantee the QoS of the voice traffic. In the following configuration we assign 50% of the bandwidth for the traffic matched by the VoiceTraffic class.

```
Router(config)#policy-map priorityCBWFQ
Router(config-pmap)#class class-default
Router(config-pmap-c)#random-detect
Router(config-pmap-c)#class VoiceTraffic
Router(config-pmap-c)#priority percent 50
```

It is important to show that we also propose (see figure 1) to include a congestion avoidance technique for the rest of the traffic, Weighted Random Early Detection (WRED) with CBWFQ. With the random-detect command we activate WRED, the net result being that the highest priority and lowest bandwidth traffic is preserved, since it starts to drop less important packets once that the net starts to be congested. WRED allows the link to be used more efficiently by selectively dropping packets according to its importance (more packets of lower priority are dropped more than the ones from high priority).

Following the designed rules that we have previously explained in our WAN Links, creating and prioritizing specific classes will have an important performance improvement.

4. Experiments and Results

The aim of this section is to show the performance improvement that a real AS LAN will have after these procedures are followed.

First, we will deploy a network infrastructure using a default configuration (without any kind of priority neither for voice nor video traffic). After that, we will configure the routers and switches with the model that we proposed in the previous section. We will compare results to determine the level of performance improvement obtained with the proposed network design.

The proposed network topology that represent an AS consists on 3 Catalyst 2600 series routers connected through their serial interfaces configured at a 2 Mb/s link speed (simulating an E1 connection). Each of the edge routers will be connected through their fast

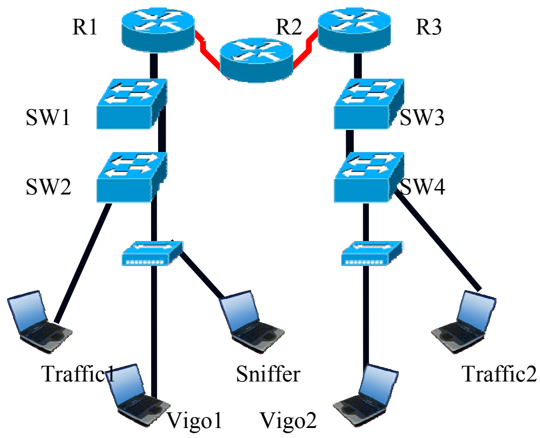


Figure 2. Scenario network topology

ethernet interface with a Catalyst 2900 series switch. Each of these switches connects to one more switch by using its gigabit ethernet trunk interfaces. Finally, the hubs and hosts are connected into these, just as it is pointed out in Figure 2. For each tested scenario we will measure the packet loss, delay and jitter, while testing the data link and network layer.

The routers were configured following a single area OSPF scheme. 802.1q was used on the fast Ethernet interfaces to support the VLAN tagging of the switches

The used IP addresses were as follows:

Used IP Addresses:

R1 Fa0.0 : 192.168.3.1	Trafico1 : 192.168.3.20
R1 S0 : 192.168.1.1	Trafico2 : 192.168.4.20

R2 S1 : 192.168.1.2	Sniffer NIC1 : 192.168.3.10
R2 S0 : 192.168.2.1	Sniffer NIC2 : 192.168.4.10.
R3 Fa0.0 : 192.168.4.1	Vigo1 : 192.168.3.15
R3 S1 : 192.168.2.2	Vigo2 : 192.168.4.15

We will use Vigo videoconference equipment in each end point of the network, while having some other clients generating traffic from protocols like ftp and http. Some other computers will use special software to flood the network with random packets in order to simulate a real scenario. The test for each scenario consists on keeping a videoconference open between two end points of the network while traffic is also being transmitted. We will perform the test of each scenario with and without voice and video priority configurations so that we can measure the improvement percentage.

By recreating a videoconference enabled scenario while also simulating normal network traffic, our testing environment comes very close in terms of reality and thus gives us a much better perception of what would the QoS performance benefit be when applied into a real world case, such as an university campus.

4.1 Endpoints inside the same network – layer 2 priority

The first and simplest scenario describes the typical switched LAN created only by switches. In our simulation, 2 Cisco Catalyst 2950 switches were connected through their Gigabit Ethernet trunk interfaces. For testing, one 3Com 10/100 hub was connected at the Fa0/1 of each of the switches, while also using a traffic generator laptop plugged into the Fa0/2 port of each switch. Both a Vigo videoconference laptop and the sniffer laptop were connected to each of the hubs. The two sniffer cards were inside the same laptop, and each card was connected to a different hub.

A videoconference was established between the 2 Vigo enabled laptops while also injecting traffic from the laptops connected through the Fa0/2 port. All traffic between switches was exchanged through the Gigabit Ethernet trunk interfaces.

The tests ran in our simulated network showed up some slight improvements after applying QoS settings at layer 2. The results were small due to the fact that our layer 2 equipment is able to switch great amounts of data in a very short time, thanks to its fast Ethernet and gigabit Ethernet interfaces. This points out that our attention should be focused into improving the layer 3 prioritization which covers the full AS to where our network is connected.

4.2 Endpoints in different networks – layer 3 priority

In this scenario, the end points are located in different networks, so the traffic will have to go through the router's serial interfaces. In this way we will just evaluate the layer 3 priority. The used network topology can be observed at Figure 2. The sniffer has two cards, each one is connected to a different network.

For this scheme, there are 2 types of router configurations that should be noted. We will refer to them as the *edge routers* and the *middle routers*, being the edge routers the ones that are directly connected to the switches and the middle routers the ones that only use their serial links to communicate the rest of the routers between themselves.

The configuration used for the edge routers were as follows:

```
Router(config)#class-map match-any VOICE-
   VIDEO
Router(config-cmap)# match protocol rtp
   audio
Router(config-cmap)# match protocol rtp
   video
Router(config-cmap)# match protocol rtp
   payload-type "34"
Router(config-cmap)# match ip dscp af41
```

The creation of the VOICE-VIDEO class identifies the RTP traffic commonly used in videoconference. We specify a payload type and a specific dscp value to compare against just to ensure that all our voice/video traffic will be recognized in this class.

```
Router(config-cmap)# class-map match-any
   HTTP-FTP
Router(config-cmap)# match protocol ftp
Router(config-cmap)# match ip dscp af12
Router(config-cmap)# match protocol http
Router(config-cmap)# match ip dscp af11
```

The HTTP-FTP class identifies the traffic we will be using as a 2nd priority. In our tests, our injected traffic is of this kind, so we provide a specific class for it to ensure the router responds as we request.

```
Router(config)# policy-map MARKING
Router(config-pmap)# class VOICE-VIDEO
Router(config-pmap-c)# set dscp af41
```

The MARKING policy is specific of the edge routers. Once in the middle routers, traffic has already been marked so there's no need to do this again.

```
Router(config)# policy-map VOICE-VIDEO
Router(config-pmap)# class VOICE-VIDEO
Router(config-pmap-c)# priority percent
   50
Router(config-pmap-c)# class HTTP-FTP
```

```
Router(config-pmap-c)# bandwidth
   remaining percent 70
Router(config-pmap-c)# class class-
   default
Router(config-pmap-c)# bandwidth
   remaining percent 30
Router(config-pmap-c)# random-detect
```

The VOICE-VIDEO policy gives special treatment to each traffic class specified previously. In the commands entered above, we define a strict 50% traffic priority to all the data matched by our VOICE-VIDEO class. From the remaining bandwidth we chose to give 70% (35% from the total absolute bandwidth) for HTTP and FTP traffic, while giving the rest of the bandwidth to any other kind of traffic not specified in any of our classes.

For each of the edge routers, we applied our policies VOICE-VIDEO and MARKING to the corresponding interfaces. For R1:

```
R1(config)# interface s0/0
R1(config-if)# service-policy output
   VOICE-VIDEO
R1(config-if)# interface fa0/0
R1(config-if)# service-policy input
   MARKING
```

The configurations for the middle router differ from the edge ones, so it doesn't require any marking because R1 and R3 (the edge routers) are doing all the marking themselves. The extra configuration required for the middle router R2 to work was:

```
R2(config)# class-map match-any VOICE-
   VIDEO
R2(config-cmap)# match ip dscp af41
```

The edge routers MARKING policy already set the dscp to af41, so the middle routers can trust this value and only compare the incoming packets against this.

```
R2(config-cmap)# class-map match-any
   HTTP-FTP
R2(config-cmap)# match protocol ftp
R2(config-cmap)# match ip dscp af12
R2(config-cmap)# match protocol http
R2(config-cmap)# match ip dscp af11
```

The same marking was done to the HTTP and FTP traffic, so no need to compare to additional values.

```
R2(config)# policy-map VOICE-VIDEO
R2(config-pmap)# class VOICE-VIDEO
R2(config-pmap-c)# priority percent 50
R2(config-pmap-c)# class HTTP-FTP
R2(config-pmap-c)# bandwidth remaining
   percent 70
R2(config-pmap-c)# class class-default
R2(config-pmap-c)# bandwidth remaining
   percent 30
```

```
R2(config-pmap-c)# random-detect
```

This is the same policy specified in the edge routers.

```
R2(config)# interface s0/0
R2(config-if)#service-policy output
   VOICE-VIDEO
R2(config-if)#interface s0/1
R2(config-if)#service-policy output
   VOICE-VIDEO
```

Finally, we apply the policy to our serial interfaces. In contrast to the edge routers, the same service policy needs to be applied to both serial interfaces. Since we don't process nor do any marking from incoming traffic, we do only need to specify the prioritization for the data already marked.

The edge router marks the header and the middle routers are dedicated to give a preferential or deferential treatment to the marked packets with a given DSCP field [10]. By following the previous steps, we will be successfully marking and prioritizing our traffic through all of our routers. It is important to note that the policies must remain equal through all routers to maintain consistency.

After applying this configuration, the sniffer laptop was set to capture and measure the time differences for a 1-way throughput. The following table shows the differences when applying the commands shown above:

Table 1. Experiment results

	Total packets	**Average delay (ms)**	**Jitter (ms)**
Voice (No QoS)	686	27.910	60.870
Voice (QoS)	705	12.036	60.401
Benefit (%)		**131.88**	**.776**
Video (No QoS)	2328	31.209	18.610
Video (QoS)	2399	17.671	17.940
Benefit (%)		**76.61**	**3.60**

During the tests there were no lost packets at all and, as shown, there is a remarkable improvement in both voice and video (131.88% and 76.61% respectively) after applying the QoS settings. However, lets keep in mind that these results were obtained on a simulated network where lots of traffic was being injected into the fast ethernet interfaces to flow through the serial link, thus forcing the router to apply the prioritization. Under higher data load, the benefits margin would have been even bigger.

Conclusions

In this paper we proposed a general guide for enabling QoS inside an autonomous system composed by several routers and switches in order to provide a more suitable environment for real time traffic used in videoconference. The two created scenarios for simulation of layer 2 and layer 3 infrastructures show up benefits from the implementation of QoS in their policies.

Even though we prioritize the voice and video traffic in our experiments, this model can be applied to any kind of traffic required in collaborative systems.

After running the tests, it's easy to notice the difference between a network with QoS enabled and one without it. The video in both edges appears smoother and the audio is not chopped, no matter what the load in the routers is, as long as the specified priority in the policy maps is enough to handle the video conference demand.

When talking about the urgency to implement QoS at layer 2, we do know that this is not so relevant to keep a good quality conference, since layer 2 only involves devices directly attached into our own switched network, thus providing a connection which depends only on our local hardware, usually fast Ethernet devices. Having a Fast Ethernet switched network provides enough bandwidth for all the devices connected to it, so QoS is not so important as long as the link speed remains constant.

However, when dealing with layer 3, many considerations have to be made since we can not control the traffic coming from other sources. Against this, we must follow the propose model in order to prioritize the outgoing/incoming traffic to be sure that the most important data keeps flowing smoothly without congestions. Inside an AS, this paper provides the required steps to enable QoS in both incoming and outgoing traffic.

The obtained results are a clear sign of the type of improvement which will be obtained in the target AS where these settings are applied (up to 131%). This AS refers to the final network where our collaborative system could be connected, enhancing the quality of their communications while allowing for total control of the traffic flowing through it.

By using a scenario recreating real traffic with the use of http, ftp, pings and udp traffic, our tests come close to reality, showing up that our general model can be successfully applied into a real world scenario while obtaining benefits close to the ones that we got here.

With this general configuration model, we can guarantee an optimal performance inside the AS, translating into a direct benefit to the network where the collaborative systems are set down.

References

[1] Elluminate, I. "Elluminate Live, Academic Edition". Internet Web Site. http://www.elluminate.com. Consultation date: October 2004.

[2] ECOLE. "Synergeia". Internet Web Site. http://www.ecolenet.nl/best/synergeia.htm. Consultation date: October 2004.

[3] Patras, U. "Synergo". Internet Web Site. http://www.ee.upatras.gr/hci/synergo/. Consultation date: October 2004.

[4] Blackboard, I. "Blackboard Portal System". Internet Web Site. http://www.blackboard.com. Consultation date: October 2004.

[5] Avouris, N.; Margaritis, M.; Komis, V. "Real-Time Peer Collaboration In Open And Distance Learning". Proceedings: 6th Hellenic European Conference On Computer Mathematics & Its Applications. Athenas. September 2003.

[6] S. Vegesna, "*IP Quality of Service*", Cisco Press, 2001. ISBN 1-57870-116-3.

[7] Monsour Yishay, Patt-Shamir Boaz, "Jitter control in QoS networks", IEEE/ACM Transactions on Networking (TON), Vol. 9, Issue 4, August 2001.

[8] L. Burgstahler *et al*, "Beyond Technology: The Missing Pieces for QoS Sucess", *Proceedings of the ACM SIGCOMM 2003 Workshops*, Aug 2003, pp: 121-130.

[9] Seven Ubik, Josef Vojtech. *QoS in Layer 2 Networks with Cisco Catalyst 3350*. **CESNET Technical Report 3/2003**.

[10] Fineberg, V., "A practical architecture for implementing end-to-end QoS in an IP network", IEEE Communications Magazine, Vol. 40 , Issue: 1, Jan.2002, pp.122– 130.

A Study on the Acoustic Properties of the Constituent Films in Solidly Mounted Resonators Using Picosecond Ultrasonic Waves

Ta-Ching Li[a,*], Shih-Piao Yu[b] , Le-Ye Tang[c] , Yong-Kang Hsiao[d], Nen-Wen Pu[d], Ben-Je Lwo[e], and Chin-Hsing Kao[b]

[a]School of Defense Science, [b]Department of Weapon System, [c]Department of Electrical Engineering, [d]Department of Applied Physics, and [e]Department of Mechanical Engineering,
Chung-Cheng Institute of Technology, National Defense University, Dahshi, Taoyuan 335, Taiwan.

*Corresponding author E-mail address: g950102@ccit.edu.tw; Tel: +886-3-3900119; Fax: +886-3-3900119

Abstract— Structurally robust, small sized solidly mounted resonators are suitable for mobile and wireless communication system. The aim of this paper is to investigate the acoustic properties of the constituent films in solidly mounted resonators through picosecond ultrasonic waves. These films include molybdenum, tungsten and silicon dioxide; they are commonly employed as constituent layers in Bragg reflectors. In addition, to decrease the parasitic capacitance in resonators, resistivity-adjustable TaN_x films were chosen as a high-acoustic-impedance layer to replace the conventional metallic films, i.e., molybdenum, and their electrical and acoustic properties were studied in detail. The acoustic impedance and velocity of reactively sputtered TaN_x films were measured from the reflectivity response of picosecond ultrasonic waves, and the value of acoustic impedance is higher than molybdenum for films grown at low N_2 concentration.

I. Introduction

As wireless networks grow rapidly in the spectrum from 500 MHz to 10 GHz and the device sizes shrink continuously, the need for integration of the front end of radio frequency becomes more apparent. Great efforts are made to develop miniature filters as spectrum crowding gets worse. One of the most promising filters for the integration is the film bulk acoustic resonators (FBARs) [1,2]. Recently, a new type of solidly mounted resonators (SMRs), which are constructed in a different way from the membrane-type FBARs, have been developed to achieve enhanced robustness. A Bragg mirror made up of a stack of alternate high- and low-acoustic-impedance (Z) quarter-wave thin films replaces the free surface at the bottom of the membrane-type FBARs. The Bragg mirror is solidly mounted on the substrate, and is used for acoustical isolation from the lossy substrate so that a high-Q resonator can be obtained. The required layer thickness and the resulting acoustic reflectivity are strongly dependent on the sound velocity and the acoustic impedance of the film. In this study, we use the high spatial resolution acoustic pulses, i.e., picosecond ultrasonic waves, to measure the velocities of various high- and low-acoustic-impedance films, and to estimate the acoustic impedance of thses films. Furthermore, since Bragg reflector is a natural MOS capacitor with a metal/oxide/substrate structure, lower parasitic capacitance can be achieved by replacing metallic films with high-resistivity films. Thus we also try to investigate the electric resistivity of the films. These results are useful for the researchers and designers of SMRs.

Tantalum nitride (TaN_x) films have received considerable interest in recent years owing to their unique properties including thermal stability [3-5], good optical properties [6], high conductivity [7-10], corrosive resistivity [11,12], and hardness [13,14], making it very attractive for use as structural elements in integrated circuits. Due to their high mass density (therefore high acoustic impedance) and adjustable resistivity (film resistor), TaN_x films are considered a replacer for the present metallic high-acoustic-impedance layers in acoustic wave resonators to overcome the parasitic capacitance problem [15].

Picosecond ultrasonics has proven useful for studying acoustic properties of thin films [16,17]. This technique measures the optical transient reflection change (ΔR) induced by acoustic waves through a pump-probe detection technique. The acoustic waves are excited with a “pump” light pulse generated by an ultrafast laser, and detected by a delayed "probe" light pulse derived from the same laser. The acoustic velocities are next obtained by measuring the echo times of the elastic strain waves (i.e. acoustic waves) reflected from the interfaces or the reflective substrate and re-enter the surface of the transducer film [18,19].

II. Experiments

All test films were deposited on p-type (100) Si substrates using radio frequency (rf) magnetron sputtering from a 50.8-mm-diameter target (99.95%). In addition, to study the adjustable resistivity in TaN films, reactive sputtering was employed. The substrate, which was neither cooled nor heated externally, was kept 60 mm from the target holder. The background pressure was vacuumed initially to 2 ~ 3 $\times$ 10^{-6} torr. The sputtering parameters of deposited films are listed in table I.

The cross-sections of the films were characterized with a scanning electron microscope (SEM, Jeol, model 6340F) with an operating voltage of 5 kV. The film resistivity was calculated from the sheet resistance measured with a four-point probe and the film thickness measured with a surface profilometer (Veeco, model Dektak III). The densities (D) of TaN films were calculated from the formula $D = M/V$, where the mass M was measured by a micro-balance and the volume V was obtained from the fixed area (600 ± 50 mm^2) multiplied by the film thickness.

To optically excite the ultrasonic waves, a self-made mode-locked Ti: Sapphire laser with 50-fs pulsewidth, 800-nm wavelength, and 86-MHz repetition rate was used. The average output power of the laser is about 200 mW, which corresponds to an energy of 2.3 nJ per pulse. The experimental setup with which we made our measurements has been described in detail previously [20].

A very thin Ni layer was deposited on these samples as the light-to-sound transducer. In addition, to ensure the accuracy of acoustic velocity measurement, the area for thickness measurement and the laser probed spot on the sample should be as close as possible. To this end, a mark was first made on the substrate, and we measured both the thickness and the acoustic response in proximity to this mark point.

III. Results and Discussion

In the following plots of reflectivity changes (ΔR) vs. delay time, we use τ to denote the round trip time of acoustic waves, and the velocity of the test film can be calculated from the echo time τ and the film thickness. In addition, the amplitude of ΔR reveals the ratio of high to low acoustic impedances, where acoustic impedance (Z) is defined as the product of density and sound velocity. The results of all test films are described as follows:

Table I. The sputtering parameters for test films.

Film	Pressure (mtorr)	N_2 partial pressure ratio (%)	Thickness (nm)
Mo[a]	7	-	177 ± 5
W[a]	7	-	100 ± 5
SiO_2[a]	7	-	200 ± 5
TaN[b]	3	0	246 ± 5
TaN[b]	3	3.3	200 ± 5
TaN[b]	3	8.5	290 ± 5
TaN[b]	3	25	188 ± 5

a: sputtered with an rf power of 80 W

b: reactively sputtered with an rf power of 200 W

A. The Mo film

Figure 1 shows the results of ΔR vs. delay time for the Mo film. The structure of test sample is depicted in the inset of the figure. A SiO_2 film (lower Z) is used as a reflective layer. Owing to the photoelasticity in Mo, the film itself serves as the light-to-sound transducer. It is obvious that the delay time τ_2 is twice of τ_1, and the amplitude of ΔR at τ_1 is bigger than at τ_2. These are the responses of the successive acoustic echoes bounced back and forth inside the film. According to the thickness of Mo in Table I and the round trip time τ_1, the velocity of the Mo film is about 7300 ± 50 m/s.

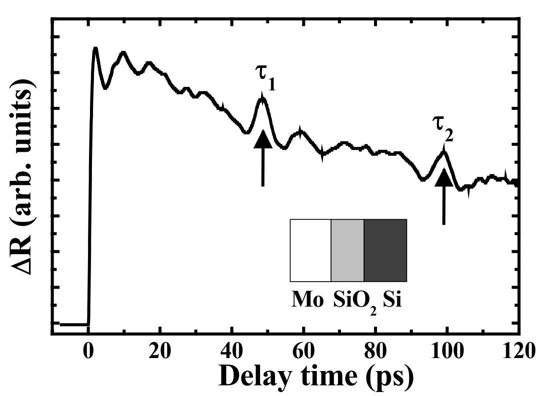


Fig. 1 $\Delta R(t)$ of Mo deposited on SiO_2. The inset is not drawn to scale.

B. The W film

Figure 2 shows the results for Ni/W film. (Ni has higher piezoreflectance [21], so the response ΔR can be easily observed.) The acoustic wave was generated by the Ni transducer. The first response at τ_1 is due to the echo reflected from the Ni/W interface, which shows no phase change because $Z_{Ni} < Z_W$ On the other hand, the second response at τ_2 is caused by the echo reflected from the W/SiO_2 interface, which does show a reflective phase shift because $Z_W > Z_{SiO_2}$. (Therefore, the shape of the response is reversed). According to the thicknesses of Ni (127 nm) and W in Table I, and the round trip times τ_1 and τ_2, the velocities of Ni and W films are about 5450 ± 50 m/s and 6230 ± 50 m/s, respectively.

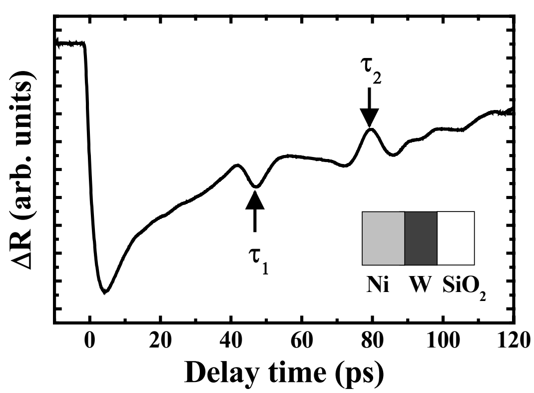


Fig. 2 $\Delta R(t)$ of the Ni/W/SiO_2 film. The thickness of Ni is 127 nm. The inset is not drawn to scale.

C. The SiO_2 film

Figure 3 shows the results of $\Delta R(t)$ for the Mo/SiO_2/W structure. Since SiO_2 has no photoelastic response, a thin Mo layer (about 20 nm) was deposited on SiO_2 film as a transducer. The plot shows two echoes: the one that appears first is smaller and denoted as τ', the other is bigger and denoted as τ_1. This phenomenon is a double-transducer effect [22]. The echoes at τ' and τ_1 are induced by W (underneath SiO_2) and Mo, respectively. This is because considerable amount of the pumping laser light transmits through the thin Mo transducer and the transparent SiO_2, and reaches the W layer. So an extra acoustic wave was generated by W. This is the reason why the delay time τ' is half of τ_1. According to the thickness of SiO_2 and the delay times, the velocity of SiO_2 is about 6450 ± 50 m/s.

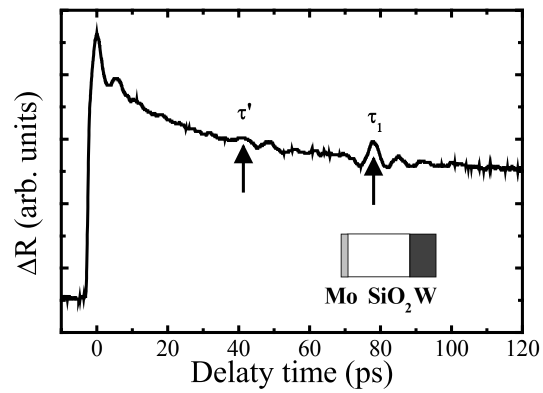


Fig. 3 $\Delta R(t)$ for the Mo/SiO_2/W structure. The inset is not drawn to scale.

D. The TaN_x films

The resistivity of TaN_x films can be changed through reactive sputtering process. Figure 4 shows the average electrical resistivities ($\overline{\rho}$) of the as-deposited TaN_x films as a function of P_{N_2}. The resistivity rises steeply (by 6 orders of magnitude) with P_{N_2}.

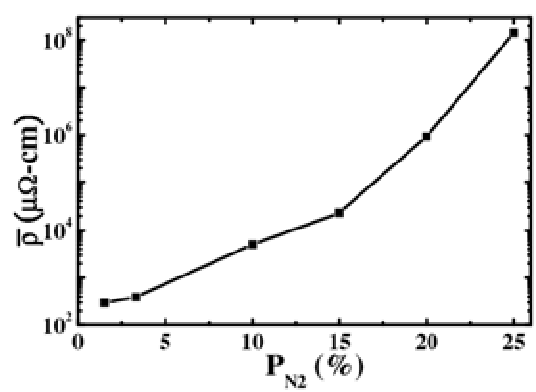


Fig. 4 The average resistivity ($\overline{\rho}$) of the as-deposited TaN_x films as a function of P_{N_2}.

The SEM pictures, which show the various microstructures in TaN_x films except for 0% P_{N_2}, are displayed in Fig. 5. The four nitrogen partial pressure ratios, 0%, 3.3%, 8.5%, and 25%, were chosen because they correspond to four different microstructures, i.e., pure Ta (0%), poly-crystal (3.3%), double-structure (8.5%), and amorphous structure (25%), respectively. In these cross-sectional pictures, an amorphous-structure sublayer sandwiched between the Si substrate and a columnar-structure sublayer can be recognized for 8.5% P_{N_2} (Fig. 5(b)). The thickness of the void-filled amorphous sublayer is an important factor influencing the mechanical and electrical properties. Our previous study [23] found that it has strong dependence on P_{N_2}: as P_{N_2} increases from 3.3%, the thickness of this amorphous part increases gradually, and eventually develops throughout the film.

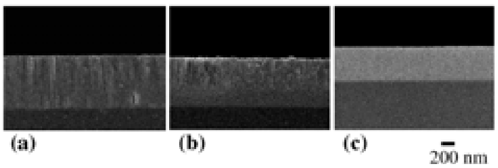


Fig. 5 SEM pictures of TaN_x films deposited at P_{N_2} of (a) 3.3%, (b) 8.5%, and (c) 25%.

The density of TaN_x films was found to decrease with increasing P_{N_2} as shown in Fig. 6. This is because the void-filled amorphous structure is more porous than the poly-crystalline structure. The acoustic impedance of different TaN_x films is also influenced by the various densities.

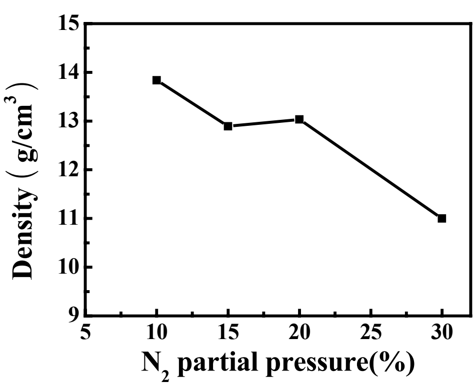


Fig. 6 The density of the as-deposited TaN_x films as a function of P_{N_2}.

To study the effects of microstructure on acoustic impedance (Z) and velocity, the above-mentioned four different structures were investigated with picosecond ultrasonics. Figure 7 shows the results of ΔR vs. delay time for these structures. Among these results, Fig. 7 (a) has the largest magnitude of the echo signal. This is because of the largest mismatch of Z between Ta and Si substrate. Having the same Si substrate, the Z of TaN_x can be estimated from the magnitude and polarity of the acoustic echo. Evidently, the acoustic impedance Z decreases with P_{N_2}.

In Fig. 7 (c), the shape of the echo signal pointed by the dotted arrow is broader than the others; moreover, its echo time (τ') is much shorter than those of the other samples (τ_1). In light of the double structures (for 8.5% P_{N_2}) revealed by SEM pictures, this broader echo is reflected from the polycrystalline/amorphous interface inside the TaN_x film. (The echo reflected from the film/substrate interface is too weak to be detected.) The broader shape was caused by the roughness of this interface. Judging from the polarity of the echo signal in Fig. 7(c), the Z of a polycrystalline structure is larger than that of an amorphous one. Calculated from the measured film thicknesses and the echo times in Fig. 7 (a), (b), and (d), the acoustic velocities for 0%, 3.3%, and 25% P_{N_2} films are $\mathbf{5000 \pm 50}$, $\mathbf{5300 \pm 50}$, and $4600 \pm 50 m/s$, respectively.

From the above data of density and velocity in TaN_x films, we anticipate the acoustic impedance ratio of Mo to TaN_x for 3.3% P_{N_2} to be about 0.98, i.e., the value of Z_{TaN} is larger than Z_{Mo} at low P_{N_2}. Thus we conclude that TaN_x is a promising high acoustic impedance material which can replace the present metallic films.

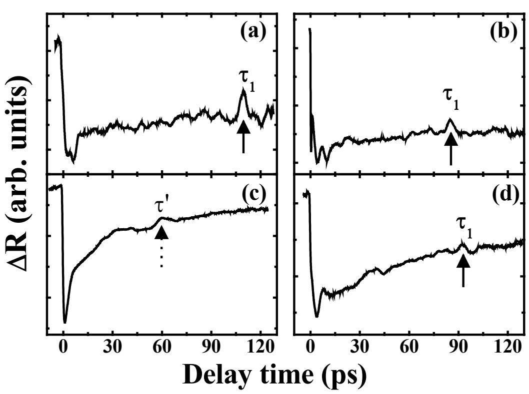


Fig. 7 The reflectance change $\Delta R(t)$ for TaN_x films deposited under (a)0% (b)3.3% (c)8.5% and (d)25% P_{N_2}. τ_1 is the round trip time, and the values are (a)109 ps (b)85 ps (c)unknown and (d) 92 ps, respectively. In (c), τ' is the echo time for the acoustic wave reflected from the polycrystalline/amorphous interface inside the TaN_x film.

IV. Conclusion

In this study, the acoustic properties of the constituent films in solid mounted resonators, which are the most advanced technology for high frequency in the next generation, were investigated using picosecond ultrasonics technique. For these films having different optical properties, different structural designs of test films are required to enhance the signal of reflectivity changes. These results are useful for the designers to improve the performance of solid mounted resonators. In addition, to reduce the parasitic capacity of filters made by SMR, the properties of resistivity-adjustable TaN_x films were studied. We found that the TaN_x films have higher acoustic impedance than Mo at low nitrogen partial pressure ratio. We believe that the simultaneous high electrical resistivity and high acoustic impedance of TaN_x films will be useful for minimizing the parasitic capacitance in acoustic wave devices.

Reference

[1] K. M. Lakin, G. R. Kline, K. T. McCarron, IEEE Trans. Microwave Theory Tech. 41 (12) (1993) 2139.

[2] R. Weigel, D. P. Morgan, J. M. Owens, A. Ballato, K. M. Lakin, K. Hashimoto, C. C. W. Ruppel, IEEE Trans. Microwave Theory Tech. 50 (3) (2002) 738.

[3] J.O. Olowolafe, C.J. Mogab, R.B. Gregory, M. Kottke, J. Appl. Phys. 72 (1992) 4099.

[4] T. Oku, E. Kawakami, M. Uekubo, K. Takahiro, S. Yamaguchi, M. Murakami, Appl. Surf. Sci. 99 (1996) 265.

[5] M.H. Tsai, S.C. Sun, C.E. Sun, C.E. Tsai, S.H. Shuang, H.T. Chiu, J. Appl. Phys. 79 (1996) 6932.

[6] S.M. Aouadi, and M. Debessai, J. Vac. Sci. Technol. A 22 (2004) 1975.

[7] X. Sun, E. Kolawa, J-S. Chen, J. S. Reid, and M-A. Nicolet, Thin Solid Films 236 (1993) 347.

[8] B. Mehrotra, and Jim Stimmel, J. Vac. Sci. Technol. B 5 (1987) 1736.

[9] T. Riekkinen, J. Molarius, T. Laurila, A. Nurmela, I. Suni, and J.K. Kivilahti, Microelectronic Eng. 64 (2002) 289.

[10] Y.M. Lu, R.J. Weng, W.S. Hwang, and Y.S. Yang, Thin Solid Films 398-399 (2001) 356.

[11] M. Nordin, M. Larsson, and S. Hogmark, Surf. Coat. Technol. 120-121 (1999) 528.

[12] S.K. Kim and B.C. Cha, Thin Solid Films 475 (2005) 202.

[13] R. Westergard, M. Bromark, M. Larsson, P. Hedenqvist, and S. Hogmark, Surf. Coat. Technol. 97 (1997) 779.

[14] R. Saha and J.A. Barnard, J. Crystal Growth 174 (1997) 495.

[15] K.M. Lakin, G.R. Kline, and K.T. McCarron, IEEE Trans. Microwave Theory and Technol. 43 (1995) 2933.

[16] H.T. Grahn, H.J. Maris, J. Tauc, IEEE J. Quantum Electron. 25 (1989) 2562.

[17] D.M. Profunser, J. Vollmann, J. Dual, IEEE Ultrason. Symp. (2002) 995.

[18] H.T. Grahn, H.J. Maris, J. Tauc, Appl. Phys. Lett. 53 (23) (1988) 2281.

[19] C. Thomsen, H. T. Grahn, H. J. Maris, J. Tauc, Phys. Rev. B 34 (6) (1986) 4129.

[20] T.-C. Li, N.-W. Pu, B.-J. Lwo, L.-J. Ho, and C.-H. Kao, Thin Solid Films (in press, 2005).

[21] G. L. Eesley, B. M. Clemens, and C. A. Paddock, Appl. Phys. Lett. **50**, 717 (1987).

[22] T.-C. Li, N.-W. Pu, B.-J. Lwo, E-Y Pan and C.-H. Kao, Appl. Phys. Lett. (no published, 2005).

[23] T.-C. Li, B.-J. Lwo, N.-W. Pu, Shih-Piao Yu, and C.-H. Kao, Surf. Coat. Technol. (no published, 2005).

On Controlled Node Mobility in Delay-Tolerant Networks of Unmanned Aerial Vehicles

Daniel Henkel, Timothy X Brown
University of Colorado at Boulder
530 UCB, ECOT 311, Boulder, CO, 80303
Tel.: 303-492-1630, Fax: 303-492-1112
{henk,timxb}@colorado.edu

Abstract

This paper explores the use of controlled node mobility in delay-tolerant networks. Mobile nodes are used to ferry data between otherwise unconnected network nodes. In particular, the problem of efficient route design for aerial data-ferrying nodes is studied. We introduce two new approaches to route design, namely the chain-relay model and the conveyor-belt model. Analytical evaluation of the proposed route designs shows their relative merits depending on node velocity, data rate, and buffer size. Further, we introduce a new problem framework, which departs from the traditional approach of seeing communication as limited to one data rate within a circular region around a node. Our model takes into account multiple data rates varying with the separation distance from a node. Preliminary evaluation suggests performance improvements over existing approaches when combining our route designs with this problem framework.

1 Introduction

Mobile ad-hoc networks (MANETS) are expected to play an increasing role in making ubiquitous computing and communications become a reality. Already mesh networks are increasingly being deployed to give people access to the Internet where no wired infrastructure is readily available. However, with the advent of ever cheaper computing power in ever smaller mobile devices the potential of MANETS goes beyond accessing the Internet. Example applications include search and rescue coordination without fixed infrastructure; scientific observation in high-risk areas; and inter-vehicle communication for collision avoidance.

The data that is used by applications in these networks can be classified as either *real-time* or *delay-tolerant*. Real-time traffic includes voice over IP, video conferencing, or monitoring of critical processes. Delay-tolerant traffic does not carry an urgency and only eventual, reliable reception is important. Examples are email, instant messages, or sensor data of a long-term experiment.

For a wireless ad-hoc network to be able to carry real-time data, there needs to be *contemporaneous, reliable path* from sender to receiver. Such a path is more likely the shorter the sender and receiver separation and the more dense the radio nodes.

Real-time communication is challenging for wireless mobile ad-hoc networks because of inherent variability in wireless connections combined with changing node positions. In the extreme case of sparsely connected, moving nodes, some nodes might not be able to connect with any other node for a long time. In this scenario only delay-tolerant communication is feasible.

The Ad-hoc UAV-Ground Network project (AUGNet) at the University of Colorado lends itself to studying methods to support sparsely-connected, delay-tolerant networks [1, 2]. AUGNet features fast-moving planes which can span large distances in a relatively short time. It is exploring how moving nodes can be used to reduce delay and improve reliability of communication in sparsely connected networks. Controlled plane mobility can increase battery life of ground nodes, enable coordination of large-scale deployments of and among swarming planes, and simplify data collection and distribution among ground nodes.

This paper investigates the role of ferrying in two communication models: the first has a constant data rate up to a maximum distance and zero rate otherwise; the second has a continuously variable rate that decreases with distance and is always non-zero. Several models of ferrying are developed for both a single transfer between two nodes and a transfer between several nodes and a hub. The results are three-fold. First, they reveal which fixed-rate ferrying model is the most suited for a given set of external operating conditions. Second, a delay-optimal scheduling algorithm is derived which assigns ferry time to nodes in a star network. Third, network performance is greatly improved by considering a variable data rate model where the data rate is a continuous function of node separation distance.

The paper proceeds as follows: Section 2 describes our model for delay-tolerant networking and route planning. Section 3 analyzes route design under full knowledge and introduces our new problem framework. Section 4 gives an overview of related work, and Section 5 concludes the paper with an outlook on future work.

2 Modeling Delay-tolerant Networks

2.1 System Components

Task Nodes
Task nodes form the basic sparsely connected ground network. The nodes operate over an area A, called the *survey area*. For our purposes these nodes are assumed to be stationary. Every node has a buffer of size b_{node} bytes and a wireless radio capable of transmitting or receiving at data rate R_{node} in a coverage area forming a concentric sphere around the node with radius r_{node}. For this scenario the data rate is considered constant within the coverage area.

It is assumed that the separation distance l between task nodes is greater than the communication range of the radios, i.e., $l > r_{node}$. Thus there is no connection between nodes and no possibility for routing traffic among them without the help of some other nodes.

A node can send data to a receiver that is at a distance less than r_{node} as long as the receiver is outside of the interference range of any other transmitting node. The interference range is assumed equal to the communication range.

We refer to this model as the fixed radius communication model.

Ferries
Mobile nodes called *transport ferries* are capable of moving around the survey area to any location requested. There are n ferries on the survey area at any given point in time, denoted $F_1, F_2, \ldots, F_n$. They are areal vehicles and the motion can be controlled arbitrarily. The motion is assumed linear with speed v, but ferries can also stop, e.g., for data exchange. Every ferry is equipped with a radio capable of transmitting or receiving at data rate R_{ferry} in a coverage area forming a concentric sphere around the ferry with radius r_{ferry}. The ferry can buffer b_{ferry} bytes of data. For simplicity we are assuming buffer size, communication radius, and data rate are the same for both, task nodes and ferries and so the subscripts will be dropped. Ferries can talk to nodes or other ferries, but not at the same time.

2.2 Performance Metrics

The analytical evaluation of the proposed ferry path models will compare the following performance measures. They are computed for the ideal case of perfect coordination between ferries.

2.2.1 Throughput

The throughput as experienced by the sender or receiver can be calculated as follows.

$$T = \frac{\text{buffers delivered per cycle}}{\text{cycle period}} \quad (1)$$

Herein the cycle period comprises the total time for a ferry to follow a given path and return to the original position and state. It includes ferry travel times plus times to fill and empty buffers. Such a cycle may involve one or more buffers of traffic delivered.

2.2.2 Average Packet Delay

The average delay experienced by one packet can be derived by looking at the best-case and worst-case delays incurred by the system with subsequent averaging[1]. The best-case delay occurs when a packet stored in the source buffer gets loaded onto the ferry last, i.e., just before departure. The worst-case delay occurs when a packet arrives at the source buffer just as the ferry departs, i.e., it gets buffered at the source as the first packet until arrival of the next ferry. Figure 1 illustrates this method.

2.2.3 Buffers

Packet delay includes the ferry travel times, time that a packet spends waiting in ferry buffer while other packets load and unload, and time for buffering at the source. Propagation delay is assumed negligible. Packets are considered small relative to buffer sizes so that buffers

[1]The validity of this approach can be shown for the chain relay and conveyor belt models. For other models it is an approximation.

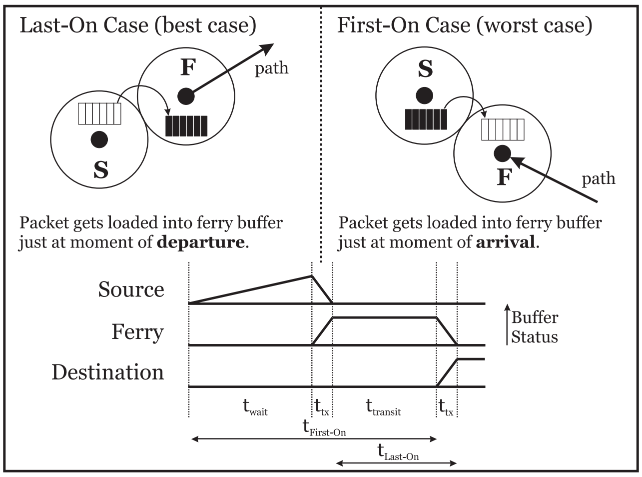


Figure 1: The two boundary cases of packet delay and corresponding timing diagram

smoothly fill and empty. Traffic flows are also smooth and continuous. Thus data is modeled as a fluid for simplicity so that we can focus on the role of mobility.

We assume a FIFO buffer policy. If R is the transmit rate then $\frac{b}{R}$ is the time required to completely fill or empty the buffer.

2.3 Traffic Flow and Ferry Path Models

When looking at traffic flow, there can be three different models: (a) Star model: all traffic flows between one special node (the hub) and the k task nodes, (b) Uni-directional model: traffic can flow between any nodes, but only in one direction, and (c) Bi-directional model: traffic can flow between all nodes in both directions.

We evaluated two different path models for ferry movement, namely the *chain-relay model* and *conveyor-belt model*. The following sections introduce these mobility models in the special case of $k = 1$ nodes communicating with the hub uni-directionally.

2.3.1 Chain-Relay Model

In this model it is assumed that all participating ferries are distributed along a connecting path from source to destination, forming a chain. The source passes available data on to the first ferry, which physically transports it some way along the path, hands it off to the next ferry, and returns to the source. This hand-off procedure is repeated until the last ferry hands the data to the destination. Figure 2 depicts this model.

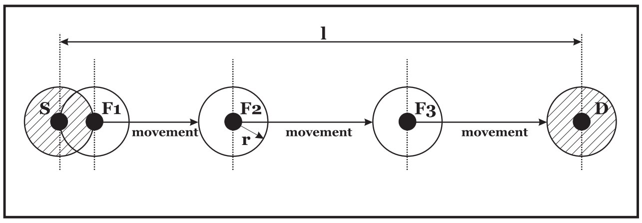


Figure 2: The Chain-Relay Model

Throughput and Ferry Travel Space

The cycle time is the time for F_1 to fill its buffer, travel to within the range of F_2, offload the buffer, and travel back to the source node. The traveling distance is the total separation distance from sender to receiver, l, less $n + 1$ times the communication radius, divided by the number of ferries that operate on the link:

$$d_{neighbor}(n) = \frac{l - (n+1)r}{n}.$$

Therefore

$$t_{cycle} = t_{travel} + t_{tx} = 2\frac{l - (n+1)r}{n\,v} + 2\frac{b}{R}$$

and from (1) the throughput can be stated as

$$T = \frac{b}{2\left(\frac{l-(n+1)r}{n\,v} + \frac{b}{R}\right)}. \tag{2}$$

In this model even if $v \to \infty$ the maximum throughput is $\frac{R}{2}$. Substituting $T = \frac{R}{2}$ into (2) and solving for n yields $n = \frac{l}{R} - 1$. In this case, $d_{neighbor} = 0$, and we reach the "saturation" case of a connected chain of relays[2].

Delay Calculation

We compute the packet delay by averaging over worst possible delay and best possible delay. (a)The First-on

Case (worst-case delay)

The packets in the chain model get handed off between neighbor-ferries and one cycle happens between two neighbors. This cycle time needs to be accounted for in the wait time for the packets. The total delay seen by a single packet from source to destination consists of the following times.

[2]Note that when $d_{neighbor} = 0$ we rely on the interference distance to be equal to the communication distance. In fact, the nodes move a small distance away from the sending node and out of its interference range before communicating with the next node in the relay. A longer interference range would require a longer minimum separation between nodes.

i) packet wait delay until neighbor ferry shows up again and stops at distance r from center of source; this includes actual travel time of ferry as well as off-loading and loading at remote neighbor, plus off-loading time at the source. This is not the average wait time since the ferry had just departed and a full cycle needs to be traveled.

$$t_{wait} = 2\frac{l-(n+1)r}{nv} + \frac{b}{R}$$

ii) ferry loading time. Even though the first packet is loaded immediately into the ferry buffer, the packet still has to wait until all packets are loaded and the ferry is ready to depart.

$$t_{loading} = \frac{b}{R}$$

iii) ferry travel delays to n remote neighbors, including the destination.

$$t_{transit} = n\frac{l-(n+1)r}{nv}$$

iv) buffer exchange time with $n-1$ neighbors. We assume an empty neighbor buffer.

$$t_{exchange} = (n-1)\frac{b}{R}$$

(b)The Last-on Case (best-case delay) A packet experiences the following delays from source to destination.

i) n transit delays between neighbors.

$$t_{transit} = n\frac{l-(n+1)r}{nv}$$

ii) n buffer exchanges between neighbors, including the destination.

$$t_{exchange} = n\frac{b}{R}$$

The average delay of packets from source to destination is the average of best and worst case delays:

$$delay^{chain} = \left(1+\frac{1}{n}\right)\frac{l-(n+1)r}{v} + \left(n+\frac{1}{2}\right)\frac{b}{R} \quad (3)$$

2.3.2 Conveyor-Belt Model

In this model it is assumed that every ferry travels the complete distance from source to destination, then returns to the source. Multiple ferries are sharing the same path, which increases throughput and robustness of the system. Figure 3 depicts the model.

Throughput and Buffer Status

One cycle in the conveyor belt model is given by a complete round-trip from source to destination including buffer transmission times. The corresponding cycle time is thus given by

$$t_{cycle} = t_{travel} + t_{tx} = 2\frac{l-2r}{v} + 2\frac{b}{R}$$

Since there are n ferries hauling data between two nodes, the data delivered in one cycle time is $n \cdot b$. According to (1), the total throughput is

$$T = \frac{n \cdot b}{2\frac{l-2r}{v} + 2\frac{b}{R}} = \frac{n}{\frac{2(l-2r)}{v \cdot b} + \frac{2}{R}}. \quad (4)$$

R is the maximum rate at which a node can transmit. $T = R$ implies that for the node to be transmitting at its maximum data, at least n_{opt} ferries are necessary:

$$n_{opt} = \left\lceil 2\left(1+\frac{l-2r}{v}\frac{R}{b}\right)\right\rceil.$$

When $n > n_{opt}$, ferries arrive faster than buffers can fill. In this case, either (a) ferries get queued at the source or (b) ferries' buffers are filled partially to $\varepsilon \cdot b$, where $0 < \varepsilon \leq 1$. ε is the ratio of the time between ferry arrivals and the time to fill the buffer:

$$\varepsilon = \frac{\frac{2}{n}\left(\frac{l-2r}{v} + \frac{b}{R}\right)}{\frac{b}{R}} = \frac{2}{n}\left(1+\frac{l-2r}{v}\frac{R}{b}\right)$$

Packet Delay Calculation

In this section we calculate the average delay for one individual packet from source to destination analoguous to the chain-relay model.

(a)The First-on Case (worst-case delay) The delay this packet experiences comprises these time components.

i) packet wait delay until the next ferry arrives and stops at a distance r from the center of the source; this includes the actual travel time of the ferry as well as off-loading and loading at the destination, plus off-loading at the source. In the single flow case there is no loading at the destination or subsequent off-loading at the source. For n ferries this delay is on average as the n^{th} fraction of the cycle time less one buffer transfer time. This is not the average wait time since the ferry had just departed and a full cycle needs to be traveled.

$$t_{wait} = \frac{2}{n}\left(\frac{l-2r}{v} + \frac{b}{R}\right) - \frac{b}{R}$$

ii) ferry loading time. Even though the packet is loaded immediately into the ferry buffer, the packet still

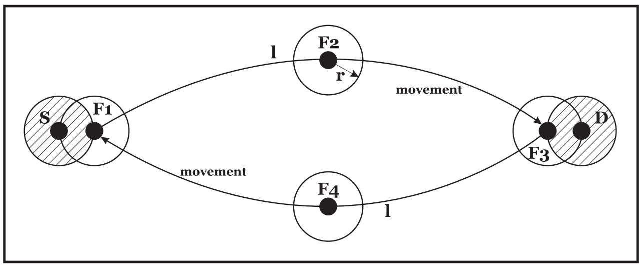


Figure 3: The Conveyor Belt Model

has to wait until all packets are loaded and the ferry is ready to depart.

$$t_{loading} = \frac{b}{R}$$

iii) ferry transit delay to destination. There will be no offloading delay since the packet is the first in the FIFO queue and can be handed to the application immediately.

$$t_{transit} = \frac{l - 2r}{v}$$

(b)The Last-on Case (best-case delay) A packet in the last-on case experiences the transit delay plus the offloading delay at the destination:

$$t_{best} = \frac{l - 2r}{v} + \frac{b}{R}$$

The average delay of packets from source to destination is calculated as the average of best case and worst case delays:

$$delay^{conveyor} = \left(1 + \frac{1}{n}\right)\frac{l - 2r}{v} + \left(\frac{1}{2} + \frac{1}{n}\right)\frac{b}{R} \quad (5)$$

2.4 Evaluation of proposed route designs

Each of our proposed route designs yields benefits over the other, depending on the environment it is deployed in. The following section compares both schemes in regards to throughput and packet delay. Environment variables include ferry speed v, ferry buffer size b, number of ferries n, data rate R, transmission radius r, and separation distance between sender and destination l.

Throughput

In the case of $n = 1$ ferries, both throughputs are equal, i.e.,

$$T = \frac{1}{\frac{2(l-2r)}{v\ b} + \frac{2}{R}}, \quad (6)$$

which validates the model analysis.

For $v \to \infty$ or $b \to \infty$ the chain model throughput becomes $\frac{R}{2}$ and the conveyor belt model yields $\frac{n \cdot R}{2}$.[3] Ferries with high buffer size can deliver more data per cycle in the conveyor belt model, since there are no intermediate buffer exchanges like in the chain-relay model. However, when using a fast radio, i.e., $R \to \infty$, the chain-relay model outperforms the conveyor-belt.

If communication range for each node is reduced, i.e., $r \to 0$, the conveyor belt model yields higher throughput. This suggests that when limited to short-range communication radios like Bluetooth, the chain-relay model should not be used.

The number of ferries such that $T = \frac{R}{2}$, which is the maximum data rate for the chain-relay model, is less for the conveyor belt model if $v > \frac{Rr}{b}$.

Delay

For $R \to \infty$ the chain model yields lower delay than the conveyor belt. When dealing with fast planes, i.e., $v \to \infty$, the conveyor belt model yields lower delay.

In summary, the conveyor belt model is superior with fast ferries, whereas the chain-relay model should be deployed with high data rate or long-range radios.

3 Algorithm Design

The previous section analyzed two models of possible ferry movement between one task node communicating to one data sink. Now we evaluate an algorithm which assigns ferry time to nodes in the case of multiple task nodes transmitting to one data sink.

3.1 Full Knowledge

The following scenario exemplifies the use of a single ferry to serve as communication enabler in a centralized, client-server star network. k stationary task nodes send traffic at their individual, but time-invariant, rates to a central server outside their coverage area. The ferry connects clients and server in a round-robin fashion, i.e., nodes are served one after the other. The ferry moves between client and server according to the conveyor-belt model. We assume full knowledge of the network parameters, including traffic flow rate f_i for each node i, distances d_i from node to server, maximum wireless data rate R at each node, buffer size b and constant speed v of the ferry. The ferry's radio can match any node's data rate, and the time for one buffer exchange is $\frac{b}{R}$. There is always a full buffer transmitted.

[3] $n_{opt} = 2$ in this case.

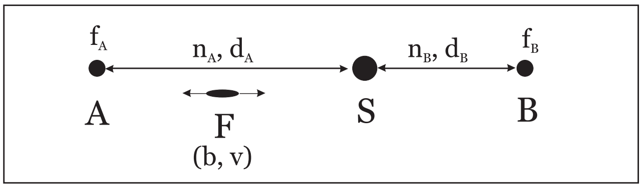


Figure 4: Problem notation: One ferry serving two nodes

The goal is to minimize the average packet delay in the network. The parameter to optimize for is the fraction of the total cycle time n_i that the ferry cycles between a given node i and the server before going on to serve the next node. A ferry can, for instance, spend more time serving nodes with higher flow rate requirement, than other nodes. Figure 4 shows the network setup with $k = 2$ nodes.

The objective function for packet delay in the network can be derived analogous to the conveyor-belt model by averaging over best case and worst case packet delay for each node. To consider all packets in the network, we further need to sum the delays over all nodes. Let each node's serving time, i.e., the total time it takes the ferry to cycle between node and server, be $n_i(2\frac{d_i}{v}+2\frac{b}{R}) = n_i\tau_i$. The total cycle time is $T_{cycle} = \sum_i n_i\tau_i$.

A packet experiences the shortest delay when it gets loaded onto the ferry last, i.e., $\frac{d_j}{v} + \frac{b}{R_j}$. If a packet arrives just after a ferry has departed, it incurs the longest delay, i.e., the cycle time to serve all other nodes, plus its own serving time. This delay is given by $3\frac{d_j}{v} + 2\frac{b}{R_j} + \sum_{i,i\neq j} n_i\tau_i$.

Averaging over both these delays yields the objective function:

$$d = \sum_{j=1}^{k}\left(2\frac{d_j}{v} + \frac{3}{2}\frac{b}{R_j} + \frac{1}{2}\sum_{i=1;\ i\neq j}^{k} n_i\tau_i\right) \qquad (7)$$

The first two terms of (7), which do not depend on n_i, are positive and constant by definition, thus they would only add to the total delay. For simplicity, these terms can be neglected without affecting the solution, and (7) simplifies to

$$d = \frac{k-1}{2}T_{cycle}.$$

Note that this model weighs the delays equally by nodes. An alternative would weigh each node's delay by its traffic volume (i.e., its flow rate f_i). It turns out that this does not affect the results and so will not be used for simplicity.

To meet the traffic demand for each node, the following flow constraints need to hold.

$$f_i \cdot T_{cycle} \leq n_i \cdot b\ \forall\ i \in \{1,..,k\}, \qquad (8)$$

$$n_i \geq 0. \qquad (9)$$

Note that if the constraints are satisfied for any $\mathbf{n} = (n_1, n_2, \ldots, n_k)$ then they can be satisfied for any scaled version of $\mathbf{n}$. Similarly the delay can be minimized by scaling $\mathbf{n}$ so that $n_i \to 0$. Therefore for concreteness we let

$$\sum_{i=1}^{k} n_i\tau_i = T_{cycle} = 1. \qquad (10)$$

Substituting (10) in (8) leads to

$$n_i \geq \frac{f_i}{b}. \qquad (11)$$

Substituting (11) in (10) gives

$$\sum_{i=1}^{k}\left(\frac{f_i d_i}{v\ b} + \frac{f_i}{R_i}\right) \leq \frac{1}{2}. \qquad (12)$$

Now even if $v \to \infty$, buffers need to be transmitted twice and (12) reduces to $\sum_i f_i \leq \frac{R}{2}$. If the data rate approaches ∞ the ferry still needs to fly back and forth twice to deliver data and (12) reduces to a constraint on flying time.

The fraction of time spent serving each node just depends on this node's flow requirement. The separation distance of node and server is not directly relevant. If only a few packets are transmitted from the node, the ferry visits this node less frequently, but instead delivers more packets on links with higher traffic flow.

Thus we have the following algorithm. First, the constraint in (12) is evaluated. If it is satisfied then the ferry can carry all flows successfully. Next the n_i are normalized to $n'_i = \frac{n_i}{\sum n_i}$ so that n'_i represents the fraction of time each node is visited. A simple stochastic algorithm would choose the next node to visit according to the probabilities $\{n'_i\}$. A deterministic algorithm would use a fractional visit counter c_i for each node to track visits.

1. Set $c_i = 0\ \forall\ i$.
2. While $\max\{c_i\} < 1$ set $c_i = c_i + n'_i\ \forall\ i$.
3. Let $j = \arg\max_i\{c_i\}$.
4. Visit node j and set $c_j = c_j - 1$.

5. When the ferry visit is finished, go to 2.

This specifies a fair algorithm in the sense that if all n_i are equal then each node will be visited once per cycle. If traffic was in the opposite direction from the central hub to the task nodes, then the scheduling would be trivial. The ferry would choose the task node with the most data waiting. This strategy is equivalent to the above deterministic algorithm.

3.2 Variable Data Rate Communication

All aforementioned approaches to determining the best location or route for a ferry were based on a simple communication model. In this model two nodes are able to communicate if the separation distance between them does not exceed a certain threshold. Once this distance threshold is exceeded, direct communication between the nodes is assumed impossible.

The model in reality is richer than this. Many digital wireless interfaces such as IEEE 802.11 have multiple data rates. When communicators are close the communication rate is high (e.g. 54 Mbps in 802.11g). When communicators are far the communication rate is lower (802.11g has 14 different rates between 54 and 1 Mbps). The rate decision is based on signal power to noise power ratio (SNR) and packet success measurements made over the channel. This phenomenon is an example of a general principle. The well known Hartley-Shannon Law states that the channel capacity C, which is the theoretical maximum rate at which information passes error free over a channel, is equal to

$$C = W \log_2(1 + \mathrm{SNR}), \tag{13}$$

where W denotes the channel bandwidth. Depending on channel coding, actual data rates can be pushed very close to this theoretical maximum. The assumption of a wireless propagation model which links distance and SNR in the fashion $SNR = \frac{k}{d^\epsilon}$ leads to a relationship between data rate and distance

$$C = W \log_2(1 + \frac{k}{d^\epsilon}), \tag{14}$$

where k is a constant, and ϵ the path-loss exponent, usually between 2 and 4. The above equation turns the discrete data rate assumption in earlier works into a continuous function. Hereby, there exists an obvious trade-off between data rate and distance, which can be leveraged for providing improved connectivity of multiple nodes through ferry placement. The idea is that the ferry can exploit communication while moving between multiple nodes in order to more quickly transfer data.

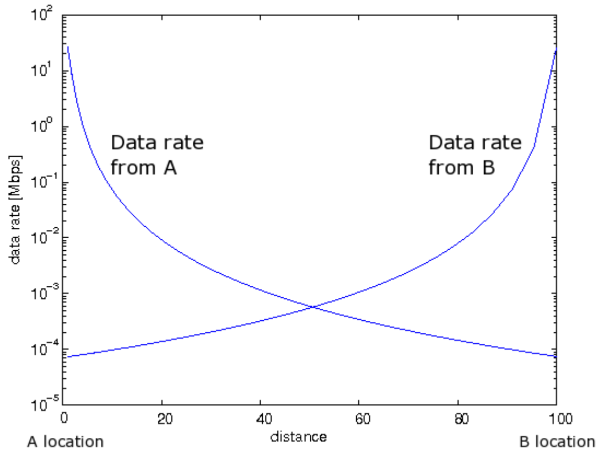


Figure 5: The possible data rates between ferry and nodes A and B as a function of ferry separation distance from A.

Figure 5 shows the achievable data rate between a ferry and node as a function of their separation distance for two nodes A and B.

In this paper we are concerned with the region when the ferries and nodes are relatively widely spaced. In this regime, SNR $\ll 1$ and we can approximate the Shannon capacity as:

$$R(d) = 1.44W\frac{k}{d^\epsilon} = R_0\frac{d_0^\epsilon}{d^\epsilon} \tag{15}$$

where R_0 is the bandwidth at a reference location d_0.

Scenario 1: One Single Ferry

Consider a node sending data to a sink. The nodes are separated by distance l. The optimal strategy for a single ferry is for the ferry to start at the midpoint, fly toward the node until its buffer is half full, then fly away from the node. Upon reaching the midpoint, its buffer will be full. Continuing to fly toward the sink, it will start to unload its buffer, until it has unloaded half of its buffer, then fly away from the sink. Upon reaching the midpoint, the rest of the buffer will have been unloaded and the transfer will be complete. In this way, the plane only moves as close to a node as is necessary to complete a transfer.

Consider two nodes separated by distance l. When a ferry is at distance y from a node and flying at velocity v, as it moves over a distance dy, it will relay $\frac{R(y)dy}{v}$ bits. So moving from the midpoint to within x of the node it will load:

$$B(x) = \int_x^{l/2} \frac{R_0 d_0^\epsilon}{v y^\epsilon} dy = \frac{R_0 d_0^\epsilon}{v(\epsilon-1)}\left(\frac{1}{x^{\epsilon-1}} - \frac{1}{(l/2)^{\epsilon-1}}\right)$$
$$= \frac{R_0 d_0^\epsilon}{v(\epsilon-1)x^{\epsilon-1}}\left(1-\left(\frac{2x}{l}\right)^{\epsilon-1}\right)$$

Setting $B(x) = b/2$, setting $d_0 = l/2$ and solving for x yields;

$$x = \frac{l}{2}\left(\frac{1}{1+\frac{b(\epsilon-1)v}{R_0 l}}\right)^{\frac{1}{\epsilon-1}}$$

The cycle time is then

$$T_{cycle} = 4(l/2 - x)/v$$
$$= \frac{2l}{v}\left(1-\left(\frac{1}{1+\frac{b(\epsilon-1)v}{R_0 l}}\right)^{\frac{1}{\epsilon-1}}\right)$$

The time $2l/v$ is the total time for the vehicle to fly from one node to the other and back. The quantity in the brackets is the fractional savings in time because of the transferred data enroute. To better see what is going on, we compute the throughput for the special case of free-space path loss, $\epsilon = 2$:

$$T = \frac{b}{T_{cycle}} = \frac{R_0}{2} + \frac{v \cdot b}{2l} \quad (16)$$

In this case we see that if v or b are large or R_0 or l are small then the throughput is one buffer per round-trip flight between nodes, $\frac{v \cdot b}{2l}$. If v or b are small or R_0 or l are large then the throughput is half the rate at the midpoint, $\frac{R_0}{2}$.

It would be instructive to compare this to the fixed radius communication model. In this model the data rate is R at radius r. Substituting this into the rate model yields $R_0 = Rr^\epsilon/(l/2)^\epsilon$ and:

$$T = \frac{2Rr^2}{l^2} + \frac{v \cdot b}{2l} \quad (17)$$

Let T_{vr} be the above throughput with variable rates, and T_{fr} be the fixed radius throughput in (6). We note that $T_{fr} \le R/2$, while T_{vr} is unbounded as v or b increases. In the best case, when $l = 2r$, $T_{fr} = R/2$ which is less than $T_{vr} = R/2 + v \cdot b/(4r)$. While T_{vr} is generally superior, for large l they both have similar throughput of about $v \cdot b/(2l)$.

Scenario 2: Multiple Ferries

The superiority of the variable rate model is more pronounced in the case of multiple ferries. In the conveyor

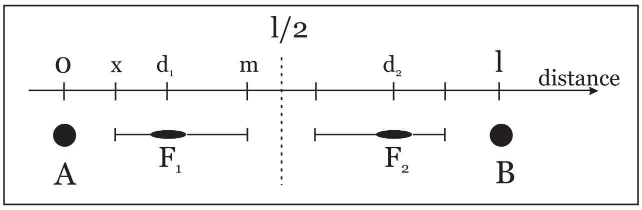


Figure 6: Multiple ferry scenario with $n = 2$ ferries.

belt, more ferries can always be added by passing closer to the source and destination nodes where the rates are higher. In the chain relay, the effective l is shortened with each additional ferry and the communication is more into the regime where the variable rate method dominates. The conveyor belt model is trivial, so we will look at multiple ferries in the chain relay model.

We assume n ferries which serve evenly spaced division points between communication points A and B. Figure 6 shows the two ferry scenario. The ferry operation dictates that the gap between two adjacent ferries is being closed at twice the speed as the gap between a fixed endpoint and a ferry. However, throughput goes down. Looking at $n = 2$ ferries, we see that $d_2 - d_1 = 2d_1 = 2(l - d_2)$. Substituting $2l$ for l, $2v$ for v, and $R_0 = \frac{Rr^\epsilon}{(l/2)^\epsilon}$ in (16) this fact becomes obvious. Now, with n ferries the approximate throughput between adjacent ferries is given by substituting $v = 2v$ and $l = \frac{2l}{n}$ in (17), resulting in

$$T_{vr} = n^2 \frac{Rr^2}{2l^2} + n\frac{v \cdot b}{2l}.$$

So we see that throughput scales quadratic with the number of ferries serving the endpoints[4].

4 Related Work

Li and Rus [4] are one of the first to address deliberate trajectory changes to make message sending between two communicating nodes possible. Two algorithms are presented - one for full knowledge of node motions the other under partial knowledge. Each algorithm results in the local update of a node's trajectory. Errors in the estimation of neighbor node locations are studied, and the algorithms are simulated. Special attention is paid to the influence of the node's message relaying duty on the primary task of each node. They use an idealized propagation model, and have no consideration of multiple flows at a time.

[4]Recall this is the case where $\frac{l}{n}$ is large and (15) applies.

The first notion of message ferrying was developed by Zhao [7] who looks at routing of messages in partitioned wireless ad-hoc networks, where it is infeasible to route messages along pre-established routes. Instead, messages are *carried* by so-called "ferries" until they reach their respective destinations. Known traffic patterns are exploited to form efficient ferry routes that minimize delays and support bandwidth requirements. While at first looking at stationary ground nodes, in [8] the authors develop two algorithms for ferry route design in the mobile scenario. One ferry services all nodes. Nodes or the ferry can initiate the messaging scheme. Either the ferry location/trajectory is assumed to be known or the nodes are equipped with long range radios to notify the ferry of their locations. This way, the authors circumvent the problem of finding nodes and ferries. In [9] they extends the route design problem to include multiple ferries. Buffer constraints are not considered. Simulations are done in the network simulator ns-2. All path planning algorithms build upon the Traveling Salesman Approach.

Unlike our approach, the TSP solution will always find an optimal cycle path among all nodes. Thus every node is visited once per cycle. In the simple case of one outlying node with a low flowrate this will limit the network capacity and increase average packet delay in the network. The node will be visited in every cycle regardless of its location or flow requirement.

Shah et al. [6] deal with the problem of sensor data collection in sparse sensor networks. So-called MULEs pick up the data when in close range to the data-gathering nodes, buffer it for some time, and drop it off at wired access points. The main advantage is the power savings in the nodes, since only short-range transmission of data is necessary. An analytic model of the system is introduced, and scaling with regards to number of mules, sensors and access points is examined. Results are meant to provide guidelines for deployment of similar systems.

Merugu et al. [5] look at identifying recurring 'paths over time' in highly partitioned networks with explicit node movements. In these *store, carry, and forward* networks it sometimes is more efficient to carry data instead of forwarding it to the closest neighbor. Their work is based on an assumed periodicity of movement in most mobile nodes. The result is a space-time routing framework which relies on space-time routing tables where the next hop is selected from current as well as future neighbors. An algorithm for creating the routing tables as to minimize delay is introduced and empirically evaluated using simulations.

Chatzigiannakis et al. [3] introduce the notion of the *support* of the network, which refers to a subset of mobile nodes dedicated to message relaying. The movement of the support is controlled by a subprotocol. Two possible motion models are proposed: (a) the *snake* protocol, where support nodes stay pairwise adjacent at all times, and (b) the *runners* protocol, where each support node performs an independent random walk on a motion graph. It is shown that basic message passing can be improved especially in highly dynamic environments. The drawback of the approach is a limitation of the support movement along an abstract graph, which does not relate to real-world coordinates. Further, nodes in the runners protocol perform random movements which can not guarantee expedited delivery of high-priority messages.

5 Conclusion and Future Work

This paper analyzes two novel ways of designing ferry routes in a delay-tolerant network, namely the chain-relay model and the conveyor belt model. It is shown that each of the models has its merits, depending on the environment it is to be deployed in. A simple client-server star network, where one ferry serves multiple nodes according to the conveyor belt model, is studied and a delay-optimal scheduling algorithm is derived. Further, a new problem framework for data ferrying in delay-tolerant networks is described. It is argued that viewing the communication data rate as a continuous function of node separation distance can greatly enhance network performance. The design of mobility algorithms as well as transmission protocols within this framework will be the focus of future work.

References

[1] T. X. Brown, B. Argrow, C. Dixon, S. Doshi, R.-G. Thekkekunnel, and D. Henkel. Ad hoc uav-ground network (augnet). In *AIAA 3rd Unmanned Unlimited Technical Conference*, Chicago, IL, September 2004.

[2] T. X. Brown, S. Doshi, S. Jadhav, D. Henkel, and R.-G. Thekkekunnel. A full scale wireless ad hoc network test bed. In *Proceedings of ISART'05*, NTIA Special Publications SP-05-418, pages 51–60, March 2005.

[3] I. Chatzigiannakis, S. Nikoletseas, and P. Spirakis. Experimental evaluation of basic communication for ad-hoc mobile networks. In *Proc. of the 5th Annual Workshop on Algorithmic Engineering - WAE'01*, volume 2141 of *Springer Lecture Notes in Computer Science*, pages 159–171, 2001.

[4] Q. Li and D. Rus. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *Mobile Computing and Networking (Mobicom)*, pages 44–55, 2000.

[5] S. Merugu, M. Ammar, and E. Zegura. Routing in space and time in networks with predictable mobility. Technical Report GIT-CC-04-07, Georgia Institute of Technology, March 2004.

[6] R. C. Shah, S. Roy, S. Jain, and W. Brunette. Data mules: Modelling and analysis of a three-tier architecture for sparse sensor networks. *Elsevier Ad Hoc Networks Journal*, (1):215–233, Sept. 2003.

[7] W. Zhao and M. Ammar. Message ferrying: Proactive routing in highly-partitioned wireless ad hoc networks. In *the 9th IEEE International Workshop on Future Trends of Distributed Computing Systems*, May 2003.

[8] W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *MobiHoc'04*, May 24-26 2004.

[9] W. Zhao, M. Ammar, and E. Zegura. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. In *Proc. of InfoCom 2005*, 2005.

Development of Performance Testing Methods for Dynamic Frequency Selection (DFS) 5-GHz Wireless Access Systems (WAS)

Frank Sanders, Jeffery Wepman, and Steven Engelking
NTIA Institute for Telecommunication Sciences, Boulder, CO USA

Voice: +1-303-497-7600/3165/4430; fax: +1-303-497-3680
fsanders@its.bldrdoc.gov, jwepman@its.bldrdoc.gov, sengelking@its.bldrdoc.gov

Abstract. *Dynamic frequency selection (DFS) is an agile radio technology designed to allow wireless access systems (WAS) to operate in 5-GHz spectrum bands that are allocated on a primary basis to radiolocation systems (radars) without causing interference to radar operations. DFS is designed to accomplish this feat by detecting co-channel radar emissions and then avoiding or vacating any locally occupied radar frequencies. DFS technology thus promises to provide more radio spectrum for applications including multimedia transmission without denying use of that spectrum to existing users. Because the successful deployment of 5-GHz DFS technology in commercially available WAS devices depends critically on the ability of testing labs to verify that such products can detect co-channel radar emissions and vacate those channels, the development of adequate performance verification methods has been critical to the development of DFS technology as a whole. This paper summarizes the history of DFS spectrum allocation by the International Telecommunication Union (ITU), the specification of DFS performance parameters in a seminal ITU Recommendation, and the development in the United States of DFS performance verification test methods that can be applied to commercially produced DFS WAS products. That development has been carried out primarily by the U.S. Department of Commerce National Telecommunications and Information Administration Office of Spectrum Management and the NTIA Institute for Telecommunication Sciences, working in close coordination with other Federal agencies (including the Federal Communications Commission) and U.S. industry.*

1. Introduction

As a limited resource, radio spectrum should be used as efficiently as possible. To that end, a variety of schemes have been suggested in recent years to share spectrum between services that have historically been deemed to be incompatible. Two such services have been wireless communications (both fixed and mobile) and the radiolocation (radar) service. Radar receivers, being noise-limited in their performance, typically experience interference effects (especially manifested as lost targets) from high duty cycle signals (such as wireless communications) at interference-to-noise ratios in the receiver IF sections at levels ranging between -12 dB to -6 dB [1]. This effect is shown, for example, for a long-range, air-search radar in Figure 1, where targets that have a 90% probability of detection in the absence of interference begin to be lost at an I/N level of -10 dB. Prevention of such interference has been considered so challenging that these services have historically been allocated in separate spectrum bands to ensure that electromagnetic compatibility problems do not occur between them.

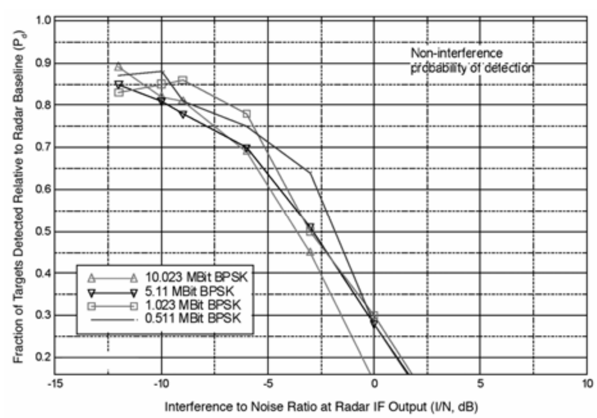


Figure 1. Example of the effects of interference on the probability of detection of targets by a long-range, air-search radar. Interference effects begin to occur at an I/N level of -10 dB.

This band-separation approach to spectrum allocation has been challenged somewhat in recent years. Another approach is now being attempted in the 5-GHz part of the spectrum, in which some operations of wireless communication devices are to share spectrum bands with radars. The primary solution is a technical approach called dynamic frequency selection (DFS). If successful, DFS should allow wireless access systems (WAS)[1] to operate in radar bands through the detection and avoidance of frequencies that are used locally by radar transmitters.

This paper begins with the historic development of this new approach to spectrum sharing from a regulatory and technical standpoint. It then describes the development of techniques and protocols for testing and verifying the ability of commercially manufactured 5-GHz DFS WAS devices to detect radar emissions and avoid operations on those radar frequencies.

2. Example of the Use of Separate Spectrum Bands for WAS Versus Other Services

Separate spectrum allocations for different (and technically incompatible) services have often been successful. In the United States, for example, WAS that are compliant with the IEEE 802.11b and 802.11g standards are operated on an unlicensed basis in the 2400-2483.5 MHz band. This represents the lower part of the 2400-2500 MHz band (2.4-GHz band) that is allocated for industrial, scientific, and medical (ISM) uses; the other major occupier of the 2.4-GHz ISM spectrum is microwave oven emissions [2, 3].

The cut-off frequencies of 2400 MHz and 2483.5 MHz were empirically derived from measurements of microwave oven emissions at the Institute for Telecommunication Sciences (ITS)[2] as *limiting* frequencies between which microwave oven emissions might cause interference to reception of signals in the direct broadcasting satellite (sound) (BSS) service. Consequently, the BSS service was allocated on a primary basis to frequencies between 2310-2360 MHz and 2500-2690 MHz[3] [4, 5]. Just above the (former) upper BSS band, aeronautical search radars, such as are used at airports, have long been allocated on a primary basis between 2700-2900 MHz. The approach of designating separate band allocations for these three services (ie., lower BSS; ISM-WAS-microwave ovens; upper BSS; and aeronautical radars), with so-called guard band separations between them of 40 MHz, 16.5 MHz, and 10 MHz, respectively, have worked to ensure electromagnetic compatibility between their operations.

3. The Development of the DFS Concept for Sharing Spectrum Between the Radar Service and WAS

In the 5-GHz part of the spectrum, regulatory requirements and band allocations have varied widely around the world for WAS 802.11-compatible devices. This variation has tended to make it difficult to manufacture IEEE 802.11a-compatible equipment for the 5-GHz bands [6]. Moreover, radar services for the purposes of radiolocation and radionavigation are allocated in the US on a primary basis in a set of bands running contiguously between 5250-5850 MHz. The use of this band by radars is substantial, as documented in an NTIA Report [7]. But unlike the example of the 2.4-GHz band described above, which does not overlap a radar band, the 5725-5875 MHz ISM band partly overlaps radar band allocations.

Internationally, the 5-GHz radar band allocations are somewhat more complicated, but they likewise overlap the 5-GHz spectrum that is perceived as desirable for use by WAS. These internationally recognized 5-GHz primary allocations for radar bands are as follows [8]: 5250-5350 MHz is allocated to the radiolocation service on a primary basis; 5470-5650 MHz is allocated to the maritime radionavigation service on a primary basis; ground-based meteorological radars operate between 5600-5650 MHz on a basis of equality with stations in the maritime radionavigation service; and 5650-5725 MHz is allocated to the radiolocation service on a primary basis.

It is desirable to make 802.11a WAS devices available on a worldwide basis, if possible, in 5-GHz spectrum without an undue burden on manufacturers to produce many different varieties of 5-GHz WAS to meet varying 5-GHz spectrum requirements from one governmental administration to the next. (Such harmonization would also make it easier for international travelers to use 5-GHz WAS devices as they move across boundaries between administrations.) The overlap of primary radar spectrum allocations with the 5-GHz ISM band has made it desirable to find technological methods to allow 5-GHz WAS (which have historically tended to use ISM spectrum for their operations) to share spectrum with radars. An ancillary goal has been to ensure a spread of WAS loading across the 5-GHz spectrum, in order to reduce the aggregate WAS emission levels at the frequencies of the satellites of the fixed satellite service and earth exploration satellite service (EESS active).

[1] Usually more specifically described as wireless local area networks (WLANs), radio local area networks (RLANs), or Wi-Fi systems.

[2] ITS, located in Boulder, CO, is part of the U.S. Department of Commerce, National Telecommunications and Information Administration (NTIA).

[3] The 2500-2690 MHz allocation for BSS was recently deleted by the FCC, but is mentioned here because it illustrates the principle of separation of services between bands.

The needs for 5-GHz spectrum harmonization and development of technical methods to allow sharing of 5-GHz spectrum between radars and WAS were addressed by the International Telecommunication Union Radiocommunication Sector (ITU-R) at the World Radio Conference of June-July 2003 (WRC-03). At that conference, the ITU administrations considered the dual needs to harmonize frequencies in the bands 5150-5350 MHz and 5470-5725 MHz for the mobile service to facilitate the introduction of wireless access systems (WAS), while still protecting the radars that operate in the bands 5250-5350 and 5470-5725 MHz. The document that was produced on this subject at WRC-03 was ITU-R Recommendation M.1652 [8]. It recommends technical mitigation techniques for government administrations to use in facilitating spectrum sharing between radars and WAS devices in the referenced 5-GHz radar bands; DFS is the technique that is primarily recommended and described in M.1652 as the solution to the problem.

4. DFS as Described in ITU-R M.1652

ITU-R M.1652 was developed through lengthy and difficult technical work by experts from many administrations in the ITU-R over the course of several years, notably in the Joint Task Group 8A/9B (JTG 8A/9B) but also to some extent in Working Party 8B (WP 8B). Thus, the content of ITU-R M.1652 actually represents the culmination of work that began many years before 2003 in various ITU-R working groups.

The Recommendation states that harmonized frequencies in the bands 5150-5350 MHz and 5470-5725 MHz for the mobile service would facilitate the introduction of WAS, including RLANs, and goes on to call for protection of radar operations in the bands 5250-5350 and 5470-5725 MHz through the implementation of technical methods for mitigation of interference effects in radar receivers. It contains detailed technical annexes that describe the characteristics of 5-GHz radars (as submitted over a number of years from several ITU-R member administrations including the US). The DFS method for protection of these radars from interference by WAS occupies the bulk of the remaining document annexes.

The goal of DFS, as described in M.1652, is for WAS-type communication systems to automatically sense the presence of radars on locally occupied frequencies and to vacate those frequencies in a timely fashion. A fundamental assumption behind DFS is that most 5-GHz radars scan their beams through space in some fashion, and that such radar emissions will be most detectable at the moments in which they scan their transmitted beams across WAS locations.

The criteria for detection of radar emissions by a DFS-equipped WAS are described in M.1652 as follows: at the location of the WAS, the minimum detection threshold for radar signals is set at -62 dBm for WAS devices with a maximum effective isotropic radiated power (EIRP) of less than 200 mW and -64 dBm for devices with a maximum EIRP of 200 mW to 1 W averaged over 1 μs.

The detection thresholds are to apply to power levels within the WAS receiver circuitry that are normalized to the output of a receiver antenna with a gain of 0 dBi, in the bandwidth of the WAS device. For example, suppose that a WAS device operates at an emission level of 100 mW, with an antenna gain of 2.5 dBi and internal circuit gain of 6 dB between the antenna and the detector circuitry. Then a radar signal will be registered in the detector circuitry at a level that is (2.5+6)=8.5 dB higher than if it were monitored with an isotropic antenna through a receiver with zero gain. Consequently, the internal detection threshold setting for such a device would not be -62 dBm, but rather 8.5 dB higher, at -53.5 dBm.

Recommendation M.1652 only calls for the radar detection functionality (hardware and software) to be built into the WAS controller, but in such a way that *every* transmitter in the WAS will be guaranteed to cease transmissions when the controller is alerted to the presence of a radar signal. In practice, for example, a WAS consisting of an access point (AP) and multiple client units might only perform radar detection through a monitor incorporated into the AP and its computer. When a radar emission is detected by the AP, that unit would in turn alert all of the clients to stop transmitting on the existing frequency and move rapidly to another frequency.

M.1652 recommends that the detection of radar signals should occur under two different conditions: both when the WAS is initially activated, and also subsequently while the WAS is operating. When the WAS is initially turned on, it is supposed to monitor a selected radio channel for radar emissions for 60 seconds. If this channel availability check detects any radar signal according to the predetermined criteria (as given above), then the WAS controller must not allow the system to operate on that channel for at least the next 30 minutes, and instead must attempt to begin operations on a different channel (necessitating yet another channel availability check on the new channel, of course).

The channel availability check lasts for 60 seconds because some radars, especially some meteorological radars, only make a complete 360-degree scan about once per minute, and the availability check must last at least long enough to be able to intercept such emissions. Most other types of 5-GHz radars perform 360-degree beam scans in less than 60 seconds, and will be detectable by a 60-second interval for channel availability monitoring.

As an additional protection for some radars, any channel between 5600-5650 MHz that has been previously flagged as containing a radar signal may not be used without 10 minutes of additional, continuous monitoring. (Some administrations, including the US government, simply may not allow any operations in that band at all.)

After the WAS begins to operate on a given frequency, the system must continue to monitor for radar signals while it is transmitting and receiving traffic. This so-called in-service monitoring for radar signals is supposed to be performed during quiet intervals between data bursts. It is crucial for two reasons. First, many 5-GHz radars are mobile, and may move into an area, using a WAS channel, after a WAS has already begun to operate. Second, even fixed 5-GHz radars may not operate continuously; such a radar could be inactive when the WAS is turned on, but could then become active on the WAS channel at some point in time after the WAS has begun operations.

When a channel must be vacated due to detection of a radar signal, M.1652 recommends that all transmissions on that channel should cease within 10 seconds. But normal traffic on the channel is supposed to cease within 200 ms, and not more than a total 100 ms of that interval should actually be occupied by data traffic. Subsequently, the remainder of the 10-second shut-down interval is supposed to be occupied by not more than a total of 20 ms of intermittent communications devoted to managing the vacating of the channel (and presumably the coordination of the transfer of data operations to another radio channel) by all of the units in the WAS. The WAS message announcing the need to cease transmissions on the radar-occupied channel is supposed to be broadcast repeatedly to ensure that the channel in question is totally vacated by all units.

Subsequent to detection of a radar signal on a WAS channel, M.1652 requires the system to vacate that frequency for at least 30 minutes (called the non-occupancy period) before attempting any more operations on that channel. Instead, the system must identify and use another channel. But M.1652 states that no frequency should be utilized by the WAS until the 60-second availability check has been performed. To fulfill this recommendation while continuing its transmissions in a seamless manner, a WAS would have to either jump to a non-DFS channel (that is, one that is not in a radar band) upon radar detection, or else would have to always monitor at least one back-up channel for at least 60 seconds at a time while simultaneously transmitting traffic on the active WAS data channel.

5. Tradeoffs in DFS Designs

From the description of DFS functionality in M.1652, the question naturally arises as to whether a power detection threshold alone is adequate (in practice) to discriminate between radar emissions versus other energy on a channel, especially during in-service monitoring. Given that some noise may occur due to non-radar emissions, and that unwanted 5-GHz emissions might even originate from a neighboring (but uncoordinated) WAS, it is likely that WAS manufacturers may use more parameters than mere power detection to determine whether or not radar emissions are occurring on their systems' channels. For example, some sort of pattern detection algorithm might be invoked to search for particular intervals between a non-WAS series of incident pulses during in-service monitoring. This would help to avoid false alarms that could be caused by triggering on random noise pulses. But that approach would be complicated by at least three facts: radar pulse repetition intervals (PRI) vary widely from one radar model (and operational mode) to another; administrations will not usually release details of radar emission parameters such as PRI for particular radars; and in any event the intervals between radar pulses are not always fixed, but rather may vary (that is, may be jittered or staggered) in time.

Another possible method for verifying that radar pulses are present and avoiding false alarms might be to discriminate on the basis of radar pulse width versus WAS data packet lengths. But this solution would be complicated by problems similar to those of discriminating on PRI: pulse widths emitted by radars will vary widely depending upon the model type and operational mode; administrations will not usually release detailed information about radar pulse widths; and finally, some radar pulse widths are comparable in length to some WAS data packets.

In reality, it is likely that WAS DFS algorithms in deployed systems may typically use some sort of composite information about received amplitudes, detected pulse widths, intervals between incident pulses, and the number of pulses observed sequentially in some predetermined interval of time to adequately discriminate between WAS traffic and radar signals. Whatever method is used, the key to designing a successful DFS algorithm will be to implement a methodology that strikes a balance between avoiding the generation of an excessive number of false alarms (which would then cause an unnecessarily large number of channel hops by the WAS), while at the same time being sufficiently sensitive to genuine radar pulses as to pass product testing requirements.

It should be emphasized that 5-GHz radars operated by various administrations vary greatly in their operational characteristics, and that administrations will not ordinarily make the details of these characteristics publicly available. Even if they did, the sheer variety of PRIs, pulse widths, beam scanning characteristics, frequency hopping capabilities, and pulse coding features that can occur in 5-GHz radars would make such an accounting nearly useless for purposes of detecting radar signals. Moreover, the future development of new radar modes in existing systems, plus the development of entirely new radar systems, would tend to render moot the desirability of obtaining details of radar emissions for the purpose of designing DFS radar-detection algorithms.[4]

Whatever sensing method is used, it is understood at present that manufacturers of WAS devices that will be available in the US will not be requested by NTIA to disclose the algorithms by which they detect the presence of radar signals. Instead, they will only be required to successfully pass performance testing with specified types of pulses at the incident power levels of -64 dBm or -62 dBm, in accordance with the ITU requirements described above.

Another inherent tradeoff in DFS algorithms is between the amount of time occupied by data traffic versus the amount of time that can be devoted to monitoring for radar signals. Radar signals can only be observed when quiet intervals occur between WAS data packets. The higher the traffic level, the less time there is to monitor for radar pulses. To understand the problem of detecting a burst of radar pulses as a radar beam sweeps across a WAS location, consider the following parameters that would be representative of a typical radar emission: the pulse width is 1 μs; the pulse repetition interval is 1 ms, the beamwidth lasts about 20 ms between the 3-dB points; and the beam-scanning interval has a periodicity of 5 sec. To obtain a 100 percent probability of detecting such a signal on the first scan of the radar beam across the WAS location, the detection algorithm would have to sample for radar pulses at least once every 20 ms (so as to never miss the radar beam-scan event that will inevitably occur once every 5 sec), and each of those monitoring events would have to last at least 3 ms (so as to extend through the interval of several radar pulses). Monitoring for at least 3 ms every 20 ms, the monitoring duty cycle would be (3/20) = 15 percent; this percentage of time would be unavailable for data traffic. As significant as this duty cycle would be, the need to perform such monitoring at least every 20 ms would represent a severe restriction on the maximum allowable length of data packets (which would of course never be able to exceed 20 ms in this example).

This example serves to demonstrate that the probability of detecting radar pulses during any given radar scan across a WAS location will realistically be less than 100 percent per radar beam-scan event; the actual probability of radar detection (on a per-scan-of-the-radar-beam basis) will be proportional to the ratio of the duration and frequency of the monitoring periods to the duration of radar beamwidth (such beamwidths typically being on the order of 10 ms to perhaps 100 ms) and the frequency of the radar beam-scanning behavior. Indeed, it is partly because this in-service monitoring detection probability cannot be expected to achieve 100 percent that Recommendation M.1652 specifies that DFS-equipped WAS devices should monitor *continuously* on the selected frequency for at least one full minute on start-up, prior to the beginning of the first data transmission.

6. Testing the Performance of DFS-Equipped WAS Devices

There are inherent uncertainties in understanding the performance of DFS-equipped WAS devices on a purely analytical basis, especially since the algorithms used by particular systems do not have to be disclosed by manufacturers and thus are not available for analysis. Therefore, a critical imperative for the approval of these devices has been to develop a method for testing and verifying their ability to sense (and respond to) radar signals on their operational frequencies.

This work is being moved forward in three stages. The first stage has been to develop and demonstrate a laboratory test bed for verifying the performance of prototype DFS-equipped WAS devices; the second has been to demonstrate the performance of some prototype devices under actual field conditions near a deployed 5-GHz radar; and the third will be to set up permanent testing systems in laboratories where actual certification

[4] In the US, rules have been developed to prevent any end user from accessing the WAS device algorithms or extracting *any* information about detected radar signals.

and acceptance tests will be performed on a long-term basis. The first two stages of work have been accomplished by groups within NTIA's OSM and ITS organizations, working in close coordination with the Federal Communication Commission (FCC), other Federal agencies, and industry personnel. The third stage of this effort will be implemented by the FCC.

It was determined about three years ago that the best approach for testing would be to expose DFS-equipped WAS devices to a variety of pulses of electromagnetic energy that would be representative of 5-GHz radar signals, consistent with the pulse width and pulse repetition parameters that are described in one of the Annexes of Recommendation M.1652.[5] It was decided that the parameters of those testing pulses would be divulged to manufacturers prior to testing, but it was further specified that under no circumstances were manufacturers to pre-program their prototype devices to search specifically for those pulse parameters. Table 1 contains pulse parameters that have been used for some of the testing at the ITS laboratory.

Table 1. Parameters of some of the radar pulses used for DFS laboratory testing in the US.

Radar test signal	Pulse repetition frequency (pps)	Pulse width (μs)	Burst length (ms)/no. of pulses	Burst period (sec)	Frequency hopping rate
Fixed frequency radar signal 1	700	1	26/18	10	n/a
Fixed frequency radar signal 2	1800	1	5/10	2	n/a
Frequency hopping radar signal	3000	1	100/300	10	1 kHz

During 2003, NTIA personnel at OSM and ITS developed a testing system that would be used to transmit these test radar pulses at representative, prototype DFS devices at frequencies near 5-GHz. As a result of considerable discussion about the desirability of coupling the test radar pulses into the DFS receivers via hardline connections versus through radiative coupling, ultimately a decision was made to perform radiatively coupled testing. This decision resulted partly from a desire to test the DFS devices in as close to a true operating mode as possible, and partly because of the practical consideration that WAS devices do not generally incorporate removable antennas, and thus it is not ordinarily possible to disconnect the antennas of such devices (as would be necessary to accomplish hardline-coupled testing) without inflicting some sort of damage to the devices.

As finally completed, the ITS 'radar' transmitter system is built around a vector signal generator (VSG) and a fast-switching programmable microwave synthesizer that are programmed with sets of pulse widths, pulse repetition intervals, operating frequencies, and pseudo-random frequency-hopping parameters that had been previously specified and agreed through inter-agency and industry coordination.[6] The system is capable of generating any of the specified radar pulses at RF frequencies near 5-GHz. The system has been operated inside a laboratory using a broadband horn antenna that is vertically polarized. At a distance of 3 m, a platform is set up for the DFS devices that are to be tested.

At the beginning of a test series, a calibrated, vertically polarized horn antenna is placed on the receiver platform in lieu of a DFS device, and the horn is connected to a spectrum analyzer. The spectrum analyzer is adjusted to 1-MHz Gaussian bandwidth (specified at the 3-dB points) and the analyzer's detection mode is set to positive peak. For the purposes of prototype-demonstration testing, the output power of the 'radar' is adjusted to produce a received power level somewhere between about -58 dBm to -60 dBm in the spectrum analyzer, as adjusted for the gain of the horn (typically about 10 dB relative to isotropic) and the loss in the cable between the horn and the analyzer.[7]

When the incident radar power level has been calibrated at the receiver platform, a prototype DFS-equipped AP unit is placed on the platform and is connected to its controller computer. One WAS client (with its own computer) is placed at a remote location in the laboratory, and data traffic is initiated between the AP and the client. (Ad hoc (client-to-client) networks have not been tested to date.) The traffic has been public-domain MPEG-2 video and wav-type audio files. The

[5] It is emphasized, however, that none of the sets of pulse parameter characteristics used in DFS testing in the US have replicated the parameters of any particular 5-GHz radars, nor will they.

[6] The radar waveform parameters are contained in the FCC 5-GHz Report and Order (FCC docket 03-122 at http://gullfoss2.fcc.gov/prod/ecfs/comsrch_v2.cgi).

[7] This power value is of course slightly higher than the -62 dBm to -64 dBm levels specified in M.1652. This is because the NTIA work has been intended to demonstrate the feasibility of the testing hardware and software; final acceptance tests performed by the FCC will presumably use power levels that will match the specified detection thresholds.

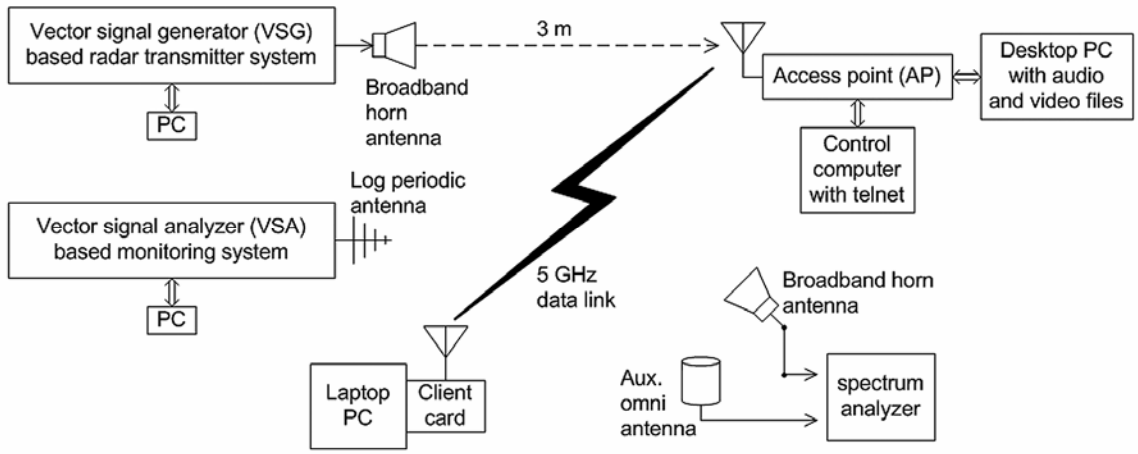


Figure 2. Overall setup of the DFS testing system that has been developed and used by NTIA. The VSA and spectrum analyzer monitoring antennas are located off-axis from the radar pulses and the 5-GHz data link.

video files have achieved an average of roughly 50 percent channel loading, and the wav files somewhat less; the video files have been used in most tests because they produce the highest loading. The overall test setup is shown in Figure 2.

Both the start-up and the in-service monitoring modes of the DFS device are tested with this set-up. In the start-up mode, a monitoring antenna connected to a spectrum analyzer is used to first of all verify that the WAS does in fact remain quiet for the first minute of its operation, with data traffic commencing at the conclusion of the 1-minute interval.

Subsequently, the AP-client start-up is repeated, but with a single burst of about 15 radar pulses transmitted at the calibrated power level once during the 1-minute start-up monitoring period. In these tests, the radar burst is either transmitted 6 sec after initial start-up or else 6 sec before the end of the 1-minute interval. For these tests, the spectrum analyzer is used to verify that the DFS does *not* initiate any operations on that frequency at the end of the 1-minute start-up interval.

With start-up behavior verified, the tests move on to verification of DFS performance during in-service monitoring. For that testing, the 1-minute start-up monitoring is suspended to speed the testing rate.

For in-service monitoring tests, the AP-client traffic is established and then a burst of about 15 radar pulses is fired on the WAS frequency if the radar mode is non-frequency hopping. (In pseudo-random frequency-hopping radar modes, a much longer series of pulses must be transmitted. This is necessary because a long interval may elapse before any of the radar pulses occur on the WAS frequency.) Every time a set of radar pulses is transmitted, three actions are taken: First, it is noted whether the AP did or did not react to the radar burst. Second, a spectrum analyzer connected to a monitoring antenna and operating in a zero-hertz span mode (that is, operating in the time domain) is used to provide a coarse verification of both the presence of the radar pulses and the shut-down (if any) of the AP-client traffic on the radar frequency. Third, a vector signal analyzer (VSA) connected to a second monitoring antenna is used to record the shut-down behavior of the AP-client traffic in enough detail to meet the specifications of Recommendation M.1652.

Furthermore, at least one in-service monitoring test is performed in which the WAS is exposed to a burst of radar pulses, and in which the WAS RF output is then monitored for the next 30 minutes to ensure that it does not attempt to use that frequency again during this so-called non-occupancy interval.

The spectrum analyzer provides low-resolution (millisecond time resolution) data traces virtually instantaneously with each radar burst, over a time interval that can last a few seconds or longer. In contrast, the VSA can provide data with microsecond time resolution (which is required to verify DFS conformance with testing requirements), but needs many minutes to provide these data, and cannot take such data for such a long period as the spectrum analyzer. Because of the long time required to generate its data outputs, the VSA is not used to record shut-down behavior of the WAS for every radar burst; instead, the VSA is used to record representative samples for a few radar bursts in each radar mode. The spectrum analyzer, in contrast, is used to verify DFS shut-down behavior (or lack thereof) for every radar burst. Together, the combination of the spectrum analyzer and the VSA have been found to provide an adequate set of DFS test-performance verification data. Dedicated computers are used to control each of these instruments and to record information from them during testing.

A statistically significant number of radar bursts are transmitted in each radar mode, ultimately providing the probability of detection of radar emissions by each prototype DFS device. It has been observed that DFS-equipped WAS devices usually detect single-frequency radar signals more successfully than frequency-hopped radar signals.

It is emphasized that the tests performed with this system in the spring of 2004 and the late summer of 2005 have not been performed to certify or accept any particular DFS devices for sale in the US. Rather, their purpose was to demonstrate the feasibility and utility of this system for such testing. All but one of the tested devices have been based on 802.11 architecture; the remaining system has used a frame-based architecture in which the talk/listen ratio is user-controlled.

A testing system identical to the one developed at ITS might be used for actual acceptance tests, but other parties have been invited to propose (and demonstrate) the utility and feasibility of alternative approaches.[8] The system described here is not the only possible approach, but to date it is the only system that has been publicly demonstrated and reported.

In the second testing stage, some prototype DFS devices were tested in proximity to an operational 5-GHz radar at a field location. For these tests, the monitoring and verification equipment were again a combination of spectrum analyzer and VSA. But for these tests, the manufacturers were *not* informed of the radar pulse parameters that were transmitted by the radar.

In addition, the performance of that radar has been tested in the presence of an aggregate emission from DFS-equipped 5-GHz WAS devices. For those tests, aggregate WAS emissions were recorded with a VSA at ITS. At the field location, those recordings were radiated from a VSG through an antenna and into the radar receiver; the effects were noted as a function of the power level of the aggregate DFS emissions.

7. Preliminary DFS Testing Results

Figure 3, from the first round of testing in Boulder in 2004[9], contains time-domain data that show a burst of radar pulses and the cessation of operation of a DFS-equipped WAS before the burst has finished. This figure is typical of the spectrum analyzer data obtained in those tests. The VSA data show qualitatively similar information at much higher time resolution.

The results of the initial bench tests in 2004 showed that the 5-GHz devices failed to achieve a good rate of radar signal detection. Overall, between all the manufacturers the radar detection capabilities of the devices tested were moderate at best and the radar detection was highly dependent upon the amount of loading of the data channel. That is, detection occurred at a higher rate when the audio file was being streamed than when the video file was being used.

A key finding of the first round of testing was that the devices were not able to detect radar pulses that were comparable in length to typical 802.11 data packets. The devices apparently tended to misidentify long radar pulses as corrupted 802.11 data packets.

The characteristics of radars that use longer pulse widths are contained in ITU-R M.1652 and these radars must be protected and detected in a timely manner. It is understood by the authors that WAS-against-radar DFS tests performed by other government administrations have drawn similar results and conclusions. Those tests have used radar parameters similar to those that were used in the NTIA bench tests (personal communications[10]).

The second round of DFS testing in the US in 2005 indicated that significant improvements had been achieved in the performance of the DFS radar-sensing capabilities against the types of radar parameters listed in ITU-R M.1652, although frequency-hopping radar signals, which were difficult to detect in the first round of tests, were still difficult to identify in the second round.

[8] Indeed, it is possible that DFS acceptance testing may ultimately be performed with hardline-coupled radar pulses rather than radiatively-coupled pulses.

[9] This first round of testing at ITS was done in two phases, in January and May 2004.

[10] ITU-R Working Party 8B meetings, Geneva, in March and September 2005.

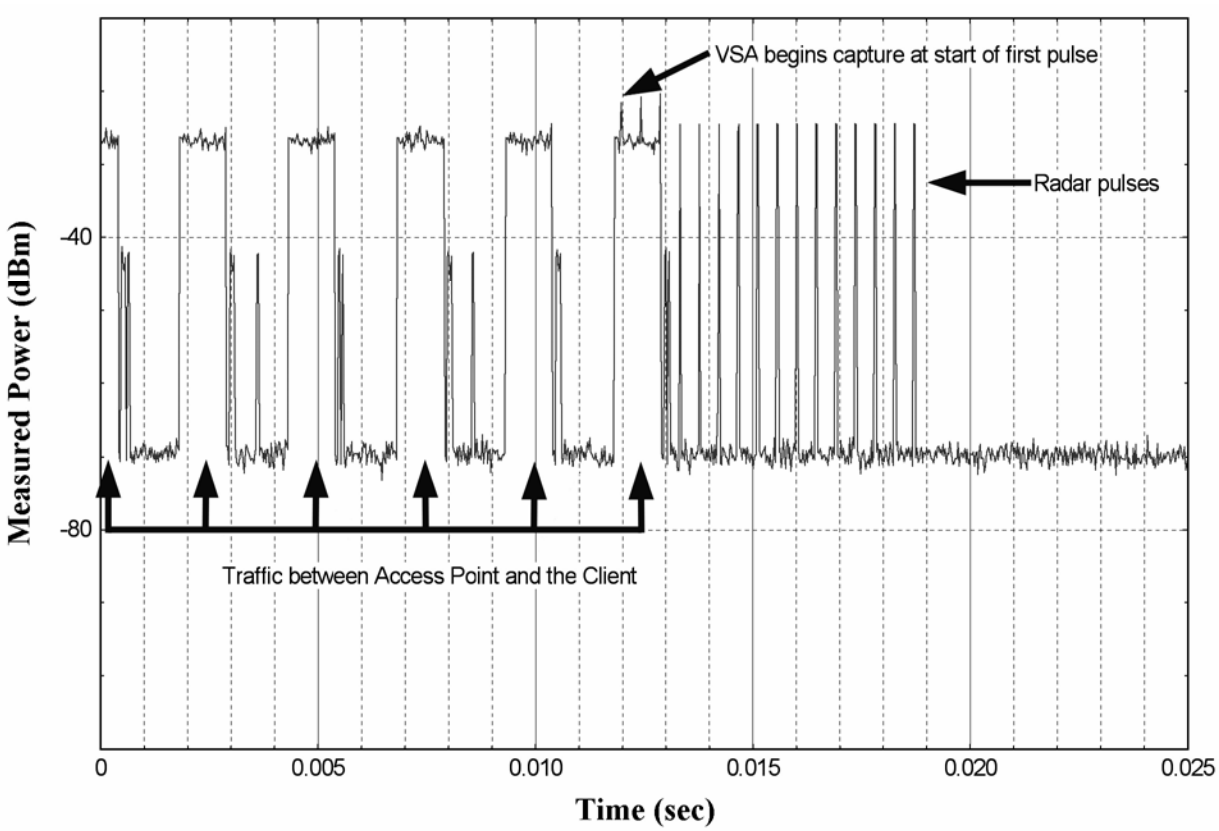


Figure 3. Time domain data from DFS testing, showing the cessation of data traffic on a DFS-equipped WAS channel when a burst of radar pulses occurred.

8. Next Steps for Implementation of DFS in the US

With the first and second stages of DFS testing (that is, laboratory demonstrations with replicated radar pulses and field tests against an operational radar) having been completed on prototype DFS-equipped WAS devices, the next steps in the US will be for the FCC to complete the rules for DFS performance criteria and testing protocols, and to implement systems for actual compliance testing. NTIA is providing all results of its work to the FCC to assist in this process.

9. References

[1] F.H. Sanders, “Technical challenges to spectrum sharing between radar and non-radar (communication) systems,” In: “Proceedings of the International Symposium on Advanced Radio Technologies, March 1-3, 2005,” NTIA Special Publication SP-05-418, Mar. 2005, pp. 21-29.

[2] P.E. Gawthrop, F.H. Sanders, K.B. Nebbia, and J.J. Sell, “Radio spectrum measurements of individual microwave ovens,” NTIA Report in two volumes: 94-303-1 and 94-303-2, Mar. 1994.

[3] M.G. Biggs, F.H. Sanders, B.J. Ramsey, “Measurements to characterize aggregate signal emissions in the 2400-2500 MHz frequency range,” NTIA Report 95-323, Aug. 1995.

[4] C.A. Filippi, R.L. Hinkle, K.B. Nebbia, B.J. Ramsey, and F.H. Sanders, “Accommodation of broadcast satellite (sound) and mobile satellite services in the 2300-2450 MHz band,” NTIA Technical Memorandum TM-92-154, Jan. 1992.

[5] *Manual of regulations and procedures for federal radio frequency management*, National Telecommuni-

cations and Information Administration, U.S. Department of Commerce, May 2003 edition, revised Sep. 2005.

[6] J. Stancavage, "Wi-Fi popularity extends to 5-GHz," *Wireless Systems Design*, Penton Media Inc., Apr. 2004. Available electronically at URL: http://www.wsdmag.com/Articles/ArticleID/7915/7915.html.

[7] F.H. Sanders, "Measured occupancy of 5850-5925 MHz and adjacent 5-GHz spectrum in the United States," NTIA Report 00-373, Dec. 1999.

[8] International Telecommunication Union Radio-communication Sector, "Dynamic frequency selection (DFS) in wireless access systems including radio local area networks for the purpose of protecting the radiodetermination service in the 5 GHz band," ITU-R Recommendation M.1652, June 2003.

Acknowledgments

The authors wish to thank Robert Sole of NTIA's Office of Spectrum Management for his important participation in and support of this work.

Impulse Radio Transmitter using Time Hopping and Direct Sequence Spread Spectrum Codes for UWB Communications

Antonio Mollfulleda
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Tel.0034 93 6452920
Fax 0034 93 6452901
antonio.mollfulleda@cttc.es

Joan A. Leyva
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Tel.0034 93 6452917
Fax 0034 93 6452901
joananton.leyva@cttc.es

Lluís Berenguer
ITOWA, Investigation Total Ware
Tel. 0034 93 6452917
Fax 0034 93 6452901
lluis.berenguer@cttc.es

Abstract— This work presents an UWB transmitter for reduced data rates. Time Hopping (TH) and Direct Sequence (DS) spread spectrum codes are used to improve the spectral efficiency below the FCC mask. The pulse generator is composed of a Class-S digital pulse amplifier and two Step Recover Diode pulse generator. An Inverter/Non-Inverter is introduced to implement the DS-SS code. In the last stage, a Wilkinson Power Combiner is presented to feed a single antenna. Theoretical and measurement results show that the PSD reaches a maximum of -45dBm/MHz and present a peak-to-average ratio of about 10dB.

Index Terms—Ultra-Wideband, Step Recovery diode, Class-S amplifier, Inverter / Non-Inverter

I. INTRODUCTION

The activities on Impulse Radio (IR) communication systems started on 1973 with the Ross Patent. In this system the carrier of information signal is a pulse train. Since then, communication systems based on IR have been used on radar, imaging and particularly on military systems as a low probability detection system. One characteristic of all this systems is that they are based on short duration pulses, associating IR system to wideband or Ultra-Wideband (UWB) systems. In 1998 the Federal Communications Commission (FCC) defined Ultra-Wideband systems as any communication based on a signal with a fractional bandwidth greater than 20% of the carrier frequency or with a bandwidth greater than 500 MHz. This definition includes both, carrier-based systems as well as impulse radio systems. In 2002 the FCC published a regulatory report allowing UWB signal transmissions with radiated Power Spectral Density (PSD) below –42dBm/MHz in the 3.1GHz-10.6 GHz frequency range. Following this report, a number of emerging commercial applications have revived significant research and development. Considering UWB systems there are two different possibilities: high data rate for short-range distances, as Bluetooth alternative; or, low data rate for high-range distances (up to 300 meters). Second chance is the one selected for the presented transmitter. In any case techniques for generating narrow pulses and UWB signals presenting efficient usage of the available power below the FCC mask is a hot research topic.

The transmission quality of any communication system depends on the received Signal to Noise Ratio (SNR), which in FCC compliant UWB systems, is strongly related to the efficient usage of the spectrum. However, the PSD of a generic IR system presents high power spectral lines that force to reduce the transmitted power in order to keep the radiated signal well bellow the mask limitations. The spectral lines are a consequence of cyclo-stationary property [1]. Using spreading codes in which more than one pulse per symbol is transmitted the level spectral lines can be reduced. Two types of codes can be used in traditional IR systems. In Time-Hopping Spread Spectrum (TH-SS) codes the pulse position is randomized inside some time interval known as a time hopping interval. The main purpose of the TH codes is to reduce the effect of the cyclo-stationary property. In Direct Sequence Spread Spectrum (DS-SS) codes the pulse sign is randomized. The DS codes are useful because from a statistical point of view it cancels the expected value of the pulse amplitude.

In this paper we present the architecture of an UWB transmitter. In section II IR signal is analyzed. Following, the design of an IR transmitter using TH-SS and DS-SS codes to reduce the effects of the spectral lines is presented in section II. Finally, the measurement results achieved are shown and the conclusion of this work is exposed.

This work is partially supported by the Spanish Ministry of Industry Tourism and Trade under contracts FIT-330200-2004-281 (PLANETS MEDEA+) and FIT-330201-2004-11 (QUETZAL), by the Spanish Ministry of Science and Education TIC2002-04594-C02 and jointly financed by FEDER and by the Catalan Government SGR 2005-00996 and SGR2005-00690.

II. IMPULSE RADIO SIGNAL

In contrast to traditional narrowband systems where the carrier of information is a sinusoidal signal, in a communication link based on IR the carrier of information is a pulse train [2]. This section describes the main modulation and coding techniques for a generalized IR system. The exposition is organized in three parts. First, the signal model is introduced including different modulation schemes as well as the signal construction using spreading codes. Second, the expression of the PSD is provided. It is shown that the PSD of an IR modulated signal is composed of a continuous part that depends on the pulse spectrum and a discrete part consisting on high power spectral lines. These spectral lines are an undesirable effect that force the reduction of the transmission power to keep the radiated spectrum below the FCC mask. Finally, it is shown that the use of spreading codes, Time Hopping (TH) and Direct Sequence (DS) codes, reduce the power of the spectral lines, which results in a more efficient usage of the available power below the mask.

A. Impulse Radio Signal Construction

The IR signal is based on the transmission of a pulse train and can be described as follows,

$$s(t) = \sum_{i=\infty}^{-\infty} a_i p(t - T_i) \qquad (1)$$

where a_i is the pulse amplitude, T_i determines the pulse position and *p(t)* is the pulse waveform. With *M-ary* modulation, *log2(M)* information bits are grouped to form one symbol. In *M-ary* Pulse Amplitude Modulation (M-PAM), a_i is chosen between the set {0, 1, …, M-1}, each representing one symbol value. In *M-ary* Pulse Position Modulation (M-PPM), the symbol pulse can be transmitted with *M* different delays, each representing one symbol value. In this case, the set of values for T_i is $\{p(t - m \cdot T_\Delta)\}$ where T_D is the so-called modulation index and $m \in \{0, \ 1, \ \cdots \ M-1\}$. Note that T_D must greater than the pulse width, which is denoted as T_P. Monocycle pulses are preferred because the radiation efficiency at the antenna is better [3]. For modern UWB-IR systems, the Gaussian monocycle is selected as a good candidate because it can be easily fitted below the FCC mask [3][4]. The mathematical expression in time and frequency domain of the Gaussian monocycle is given by,

$$p(t) = A \cdot \sqrt{2e}\,\frac{t}{t} \cdot e^{-\left(\frac{t}{t}\right)^2} \qquad (2)$$

$$P(f) = -jAp\sqrt{2pe} \cdot f \cdot t^2 e^{-(p f t)^2} \qquad (3)$$

where A is the pulse amplitude and t is the parameter that controls the pulse width, T_P. From (3), it follows that the central frequency f_c and the 3-dB bandwidth $BW_{3\text{-}dB}$ of the Gaussian monocycle are given by,

$$f_c = \frac{1}{tp\sqrt{2}} \qquad (4)$$

$$BW_{3-dB} = \frac{1.16}{tp\sqrt{2}} \qquad (5)$$

Let's define the central frequency of the pulse as the middle point between the limits of the FCC mask (3.1-10.6GHz). It follows that f_c = *6.85GHz* and thus, t = *38.5 ps*. Figure 1 depicts the time domain response of the Gaussian monocycle, resulting a pulse width of about T_P » *180 ps*. Figure 2 show the frequency domain of the same pulse. Note that filter requirements to meet the FCC regulations can be easily achieved.

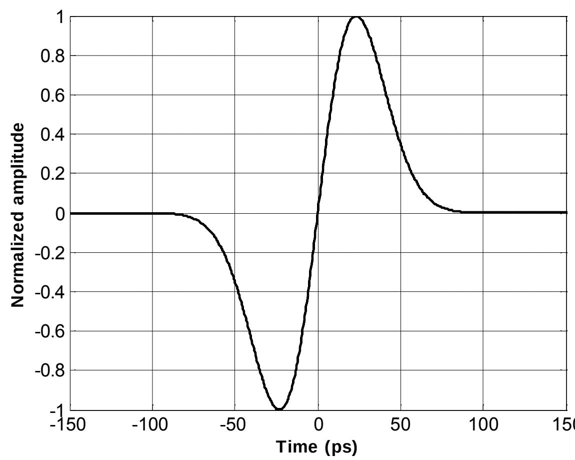


Fig. 1. Gaussian monocycle with t = *38.5ps*.

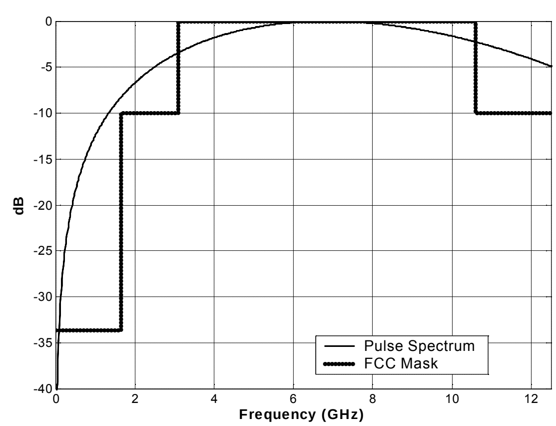


Fig. 2. Frequency response with t = *38.5ps* .

IR signal are commonly based on spreading codes for several purposes. These codes can be considered to make the communication more robust against interferences, or as a medium access mechanism [5]. The spreading codes can also be used to reduce the spectral lines in the PSD of the radiated signal, which is the main concern in an FCC compliant transmitter. When a spreading code is applied, more than one pulse is transmitted for each symbol.

Let's consider an IR communication system having one bit per symbol and working at a rate of $R_b = 1/T_b$ *bits/s*, where T_b is the bit period. In order to introduce the spreading codes, each bit interval is divided into *Ns* equally sized intervals of length *Ts*, named frame intervals. Note that $Ns \cdot Ts$ must be smaller than or equal to the bit period *Tb*. In the same way the frame interval is divided into N_h equally spaced time intervals known as chip intervals. In each frame interval only one modulated pulse is transmitted. The pulse sign or the position of the pulse inside the frame interval depends on the spreading code applied to the IR signal. Different spreading codes can be defined for generic IR signal:

·No Spreading Code. Only one modulated pulse is transmitted per bit. In this case $Ns = 1$, $N_h = 1$ and $Ts = Tb$.

·Repetition Code. The same modulated pulse is transmitted at the beginning of each frame interval. Note that both, pulse amplitude and pulse position are the same for all frame intervals of one bit. In this case $Ns > 1$, $N_h = 1$ and $Ns \cdot Ts$ £ *Tb*.

·Direct Sequence Spread Spectrum Code (DS-SS). The modulated pulse is transmitted at the beginning of each frame interval. The sign of the pulse in each frame interval is modified according a direct sequence spreading code. In conventional DS-SS systems the code sequence is of length *Ns* and the same sequence is used for all bits. In this case $Ns>1$, $N_h = 1$ and $Ns \cdot Ts$ £ *Tb*.

·Time Hopping Spread Spectrum Code (TH-SS). In this case, for each frame interval only one pulse is transmitted on one of the chip intervals. The TH code determines the chip interval in which the modulated pulse will be located into each frame of each bit. N conventional TH-SS coding $Ns>1$, $N_h >1$ and $Ns \cdot Ts$ £ *Tb*.

From a practical point of view the TH code is generated by a pseudo-random number generator, which inevitably results that the TH code will repeat itself. Let's denote T_{TH} as the repetition period and N_b the number of bits transmitted in this period. The signal transmitted in the k^{-th} period when TH and DS spread spectrum codes are used can be written as,

$$s_k(t) = \sum_{l=0}^{N_b} \sum_{h=0}^{N_s-1} c_h^{DS,l} \cdot a_l^k \cdot p\left(t - lT_b - b_l^k T_\Delta - c_h^{TH,l} T_c - hT_s\right) \quad (6)$$

where $s_k(t)$ is the transmitted signal conveying the k^{-th} sequence N_b consecutive bits; a_{lk} is the amplitude of the *l-th* bit of the *k-th* sequence and it's used for M-PAM modulation. When no PAM modulation is used this parameter is set to $a_{lk} = 1$. b_{lk} is the time position of the *l-th* symbol of the *k-th* sequence. It carries the information corresponding to M-PPM modulation. When the PPM modulation is not used, this parameter is set to $b_{lk}=0$. Inside each bit interval *h* denotes the frame number. $c_h^{DS,l}$ is the DS code chip amplitude for the *h-th* frame of the *l-th* symbol. When no DS code is used this parameter is set to $c_h^{DS,l} = 1$. $c_h^{TH,l}$ selects the chip interval of the *h-th* frame of the *l-th* symbol according with the TH code. When no TH code is used this parameter is set to $c_h^{TH,l} = 0$. T_s is the duration of the frame interval. Finally, T_c is the duration of the chip interval, T_b denotes the duration of the symbol interval and T_D represents the Modulation index for PPM.

As a matter of example consider an IR signal with the following parameters,

$$\begin{aligned} &N_b = 2, \quad N_s = 2, \quad N_h = 2, \quad 2-PPM \\ &bit \quad sequence \ \{b_0 = 1, b_1 = 0\} \\ &TH \quad Code \quad sequence \ \{0, \quad 1, \quad 1, \quad 0\} \\ &DS \quad Code \quad sequence \ \{1, \quad -1, \quad 1, \quad -1\} \end{aligned} \quad (7)$$

Fig. 3 shows the constructed signal corresponding to one period of the TH code. Note that the pulse sign is inverted according to the DS code whereas the pulse position depends on both, the TH code and the input bits.

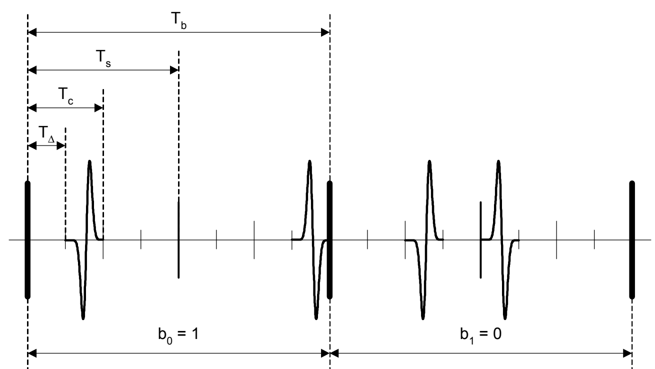


Fig. 3 Example of IR signal using TH and DS codes

B. Power Spectral Density

In general, the signal in (6) is a cyclo-stationary random process. As a consequence the PSD $S_y(f)$ consists of a continuous component $S_y^C(f)$ and a discrete component $S_y^D(f)$ [6]. The continuous part is mainly dependent on the pulse spectrum. The discrete part is a set of spectral lines, whose separation in frequency is proportional to the periodicity of the cyclo-stationary process. In some IR systems these spectral lines could be exploited for timing synchronization at the receiver [7]. However, in an FCC compliant transmitter they can be problematic because they cause interference to narrow-band systems. A possible solution to the latter problem could be to reduce the transmitted power to assure that all spectral lines are below the FCC mask, but in this case the total transmitted power would be too small for a reliable link. Therefore, we can define the spectrum efficiency as follows,

$$h = \frac{Peak\ Power\ of\ Discrete\ Part}{Maximum\ Power\ of\ Continuous\ Part} \quad (8)$$

Analytical expressions of the PSD for a generalized UWB-IR signal in (6) can be found in [8] and [1]. In [8] the PSD is calculated taking into account the effect of the timing jitter and assuming a completely random TH code. A PSD expression is given in [1] for both, short and long TH spreading codes. Following the notation in [1], the PSD of a generalized signal described in (6) can be written as:

$$\begin{aligned} P_y(w) = \frac{1}{T_{TH}} E\left\{ \left| S_p(w) \right|^2 - E\left\{ S_p(w) S_q^*(w) \right\} \right\} + \\ \frac{1}{T_{TH}^2} E\left\{ S_p(w) S_q^*(w) \right\} \sum_k d\left(w - \frac{2p\,k}{T_{TH}} \right) \end{aligned} \quad (9)$$

where p and q are used to identify two different waveforms signals conveying N_b symbols, T_{TH} is the duration of signal $s_p(t)$ and $S_p(w)$ if the Fourier Transform (FT) of $s_p(t)$. The continuous and discrete part of the PSD can be clearly distinguished as the first and second term respectively of (9).

Let's consider the case when 2-PPM using TH-SS and DS-SS codes is used. In this case the period of the TH code is a number N_b of consecutive symbols. Formulating as in [1] the PSD can be written as,

$$\begin{aligned} S_y(f) = \frac{\left| P(f) \right|^2}{T_{TH}} \left\{ \left\{ R_0^a - R_1^a R_1^b(w) \right\} \sum_l T_l(w) T_l^*(w) \right\} + \\ \frac{\left| P(f) \right|^2}{T_{TH}{}^2} \left\{ R_1^a R_1^b(w) \sum_{l,n} T_l(w) T_n^*(w) \right\} \Pi_{TH}(w) \end{aligned} \quad (10)$$

where $\left| P(f) \right|^2$ is the spectrum of the pulse,

$R_0^a = E\left\{ a_l^p a_n^p \right\}$ for l=n,

$R_1^a = E\left\{ a_l^p a_n^p \right\}$ for l ¹ n,

$R_0^b(w) = 1$ for l=n,

$R_1^b(w) = E\left\{ e^{-jw\left(b_l^p - b_n^p\right)T_\Delta} \right\}$ for l ¹ n,

$T_l(w) = \sum_h e^{-jwc_{l,h}T_c} e^{-jwhT_f} e^{-jwlT_b}$,

$\Pi_T(w) = \frac{1}{T} \sum_k d\left(w - \frac{2p\,k}{T} \right)$.

Note that with the term $T_l(w)$, which depends on the TH code, and the term $R_1{}^a$, which depends on the DS code, the discrete part of the PSD can be reduced, improving the spectrum efficiency.

C. Time Hopping Codes Generation

There are a number of code generators that might be used to generate the TH codes for UWB systems [2] [9]. The main purpose for using TH and DS codes in this IR transmitter is to reduce the spectral lines of the radiated PSD. However, it can be shown that there will be spikes at the spectrum that cannot be removed by using spread spectrum codes [2]. Maximum Length (ML) codes present a good compromise between complexity of implementation and the spectrum efficiency.

The TH code is a sequence of integers inside the set $\{0, \ldots N_S\text{-}1\}$. The ML codes for time hopping can be generated as shown in Fig. 4. It is a pseudo-random bit generator based on a primitive polynomial. We have chosen the following primitive polynomial to generate the TH codes,

$$P(x) = 1 + x^2 + x^{10} \quad (11)$$

The Symbols of the TH codes can be obtained by multiplying some bits of the register counter by powers of two as expressed as follows,

$$\begin{aligned} S &= \begin{bmatrix} s_0 & s_1 & \cdots & s_M \end{bmatrix}^T \\ TH_S &= \begin{bmatrix} 2^0 & 2^1 & 2^2 & \ldots & 2^n \end{bmatrix} \bullet b \end{aligned} \quad (12)$$

where $M = log_2(N_S)$ and TH_S is a vector containing the sequence of the TH code.

As a matter of example let's consider an IR transmitter with the following parameters:

- Pulse waveform: Gaussian Monocycle
- Number of frames per symbol $N_S = 8$.
- Number of chip Intervals per frame, $N_h = 16$ $(M = 4)$. A 4-bit TH code is generated
- Number of symbols in a period of the TH code, $N_b = 32$
- $T_c = 25ns$
- $T_D = 1ns$
- $T_{TH} = 102.4\ ms$
- Taking one bit of the scrambler of Figure 5 is formed the DS code.

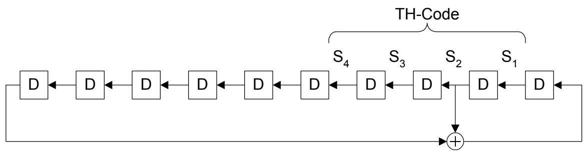


Fig. 4. Generation of ML Spreading Codes.

Fig. 5 shows the resulting PSD. The ratio between discrete part of the spectrum and the continuous part is reduced to approximately 10 dB.

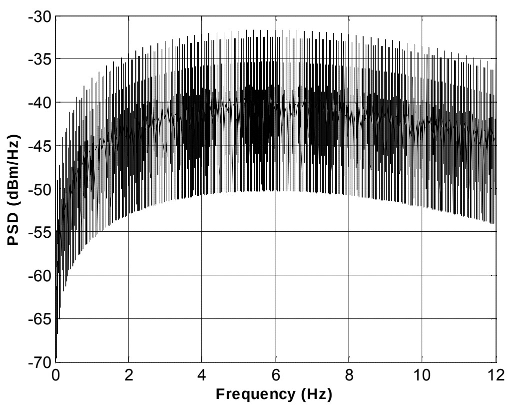


Fig. 5. PSD using DS and TH Spreading codes.

III. Transmitter Architecture

The IR transmitter is composed of a digital unit, in which the baseband processing is executed, a Digital Pulse Amplifier (DPA), a dual channel pulse generator, an Inverter/Non-Inverter block, which inverts the sign of only one of the pulse channels and a power combiner to radiate both channels on the same antenna. Fig. 6 shows the block diagram of the transmitter.

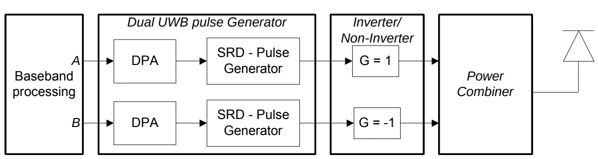


Fig. 6. Transmitter Block Diagram

The digital unit constructs the IR signal using TH codes as described in the previous section and delivers the sequence of digital pulses. The DS code is implemented by generating those pulses with DS code equal to 1 through channel A and those pulses with DS code equal to –1 through channel B. The two DPA drive the digital pulses to the dual channel pulse generator. The pulse generator is based on Step Recovery Diodes (SRD). As a result, the transmitter is able to generate two Gaussian monocycle pulses with opposite polarity for DS-SS codes.

A. *Digital Pulse Amplifier*

The digital pulse amplifier consists of a Class-S amplifier, which drives a medium power bipolar transistor to work as a switch [10]. The bipolar transistor is the responsible of delivering enough current to the SRD pulse generator. Pulse amplifier circuit is depicted in Fig. 7. Transistors Q1 and Q2 (Infineon BFT92P and BFR92 respectively) perform the two-position switch. The complementary pair Q1 and Q2 ensures rapid switching and adequate voltage to maintain the ON/OFF state of the output transistor Q3 (Infineon BFG235). The zener diode (Z_1) and base resistance (R1 and R2) are used for accommodating the voltage levels in DC mode. In order to raise the switching speed base capacitors C1 and C2 are introduced. The voltage variation through those capacitors generates additional current, which contributes to the rapid removal of stored charge in the PN-junction of the transistors. The low pass filter formed by the RC circuit limits the maximum Pulse Repetition Frequency (PRF). The constant time of this RC circuit must be $RC < 1/10 f_p$, where f_p is the maximum pulse repetition frequency. It must be noticed that the SRD pulse generator requires 50 Ohms output impedance of the pulse amplifier.

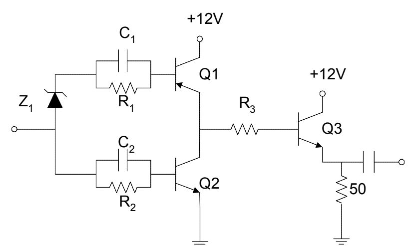


Fig. 7. Digital Pulse Amplifier

B. *SRD Pulse Generator*

SRD diodes have been used during years for the generation of short duration pulses [11]. The PRF of conventional SRD is often about 10 MHz. On the other hand, in UWB systems the PRF is needed to be around 60 MHz, combined with 3 V pulses, to obtain the maximum PSD allowed level. In this section we present a transmitter capable to operate up to 40 MHz thanks to a high-speed digital pulse amplifier able to drive enough current for the SRD circuit.

The main characteristic of a SRD diode is the very fast switching from forward to reverse modes. This property is the one that permits to generate falling or raising edge signal of tenths of picoseconds. While the diode is forwarded- polarized electric charge is stored in the P-N junction of the diode. The total stored charge depends on the average current and the recombination time of the diode according to

$$Q_F = I_F \cdot t \tag{13}$$

where Q_F is the stored charge in the forward mode I_F is the average current flowing through the PN-junction and t is the recombination time of the diode.

Fig. 8 shows the designed Gaussian pulse generator circuit. It is based on typical circuit configurations for Gaussian pulse sharpening using SRD diodes [11] and monocycle Gaussian circuits [12] [13]. At the beginning of operation, the voltage source V_{in} is set to ground level. In this mode the SRD-1 operates in forward polarization through the bias circuit. The total stored charge in the PN-junction is given by (13). When the voltage source V_{in} rises (it can take a few nanoseconds), reverse current flows through the diode removing the stored charge. During this discharge period the diode keeps the low impedance state and consequently its voltage remains close to ground level. When the stored charge is completely removed, the SRD-1 switches abruptly to a high-impedance state. In this moment the input voltage is transmitted to the load. It generates a rising edge which duration is equal to the switching time of the diode, that can be lower than 100ps.

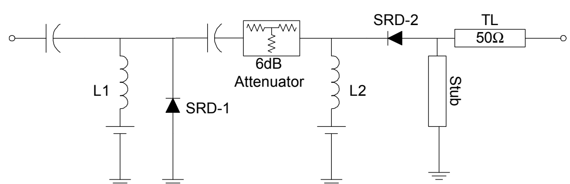


Fig. 8. SRD Gaussian pulse generation circuit

In the same way that SRD-1 is used to configure the rising edge of the pulse, SRD-2 is used to generate the falling edge of the Gaussian pulse. Whereas SRD-1 is operating on low impedance state, the output of the first state does not change from low level. During this time SRD-2 also operates in forward polarization through the bias circuit. Again the stored charge of SRD-2 is given by (13). In the moment in which SRD-1 switches to high impedance state, the rising edge is propagated to SRD-2. During the rise time, reverse current flows through SRD-2 removing the stored charge. When the PN-Junction of SRD-2 is completely empty of charge, it switches to high impedance state forcing the falling edge of the Gaussian pulse.

Following the Gaussian pulse generator, a passive shaping- network is used to generate a Gaussian monocycle pulse. Different shaping-network topologies may be used to perform this operation [11][12]. Short-circuited stub is a simple an easy network that works properly to create the Gaussian monocycle pulse. The stub generates a reflection of the pulse inverting the phase and introducing a delay t_P that depends on its length. This reflection, added to the former pulse, produces the desired pulse. To obtain the optimum Gaussian monocycle, t_P should be the half of the original pulse generated.

In order to increase the amplitude of the pulse and reduce the ripple and/or reflections generated by the circuit, some modifications have been done to the typical circuit. The 50-Ohm load used in [11] after SRD-2 has been eliminated and SRD-2 is then connected directly to ground using the short-circuited stub of the shaping network. Then, a reflection still appears at the output because the pulse goes through de SRD-2 backwards to SRD-1 generating again another pulse. To solve this problem a 6dB-attenuator has been placed between the diodes, so that the reflection is then considered negligible. Although this solution reduces the amplitude, the pulse generated has improved ripple as depicted in Fig. 9.

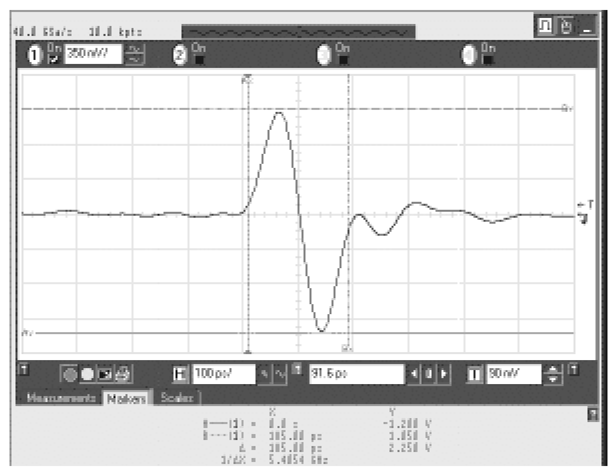


Fig. 9. Gaussian Monocycle Pulse (185 ps @ 2.2 V)

The Gaussian Monocycle pulse is not completely symmetric (87,5%). Inductance at SRD pins is the main fact that contributes to the asymmetry of the pulse [11]. Other characteristic that also affects is that the fall and rise of the pulse (before the shaping network) are produced in different ways. Whereas the rise is produced by a slow edge done by the digital amplifier towards SRD-1, the fall is consequence of the fast transition of the SRD-1 diode going through SRD-2.

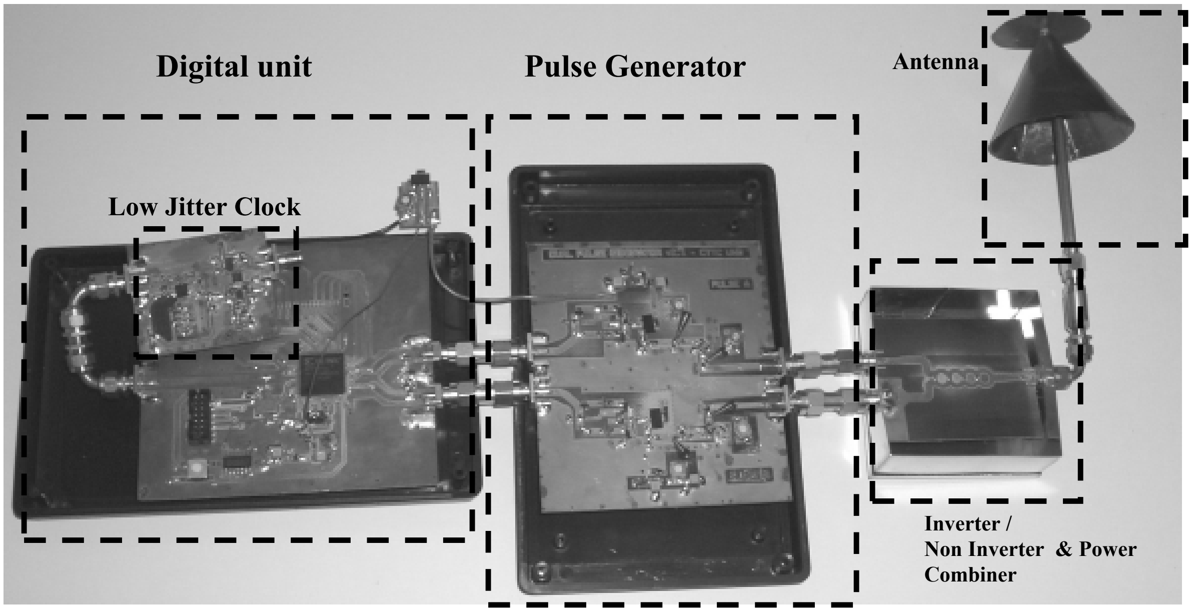


Fig. 10 Impulse Radio Transmitter

C. Inverter / Non- Inverter Circuit

The inverter key point lies in the transitions from unbalanced to balanced line. Changing the output ground plane side causes the variation of the electric potential reference, and the consequent pulse inversion.

The use of active devices to perform the inverter is refused due to their limited bandwidth. An alternative passive inverter circuit using slotlines is studied in [14]. This technique is also discarded due to the narrower slots needed to achieve low impedance slotlines. Passive pulse inverting technique based on Double-Sided Parallel-Strip Lines combined with microstrip lines has been chosen.

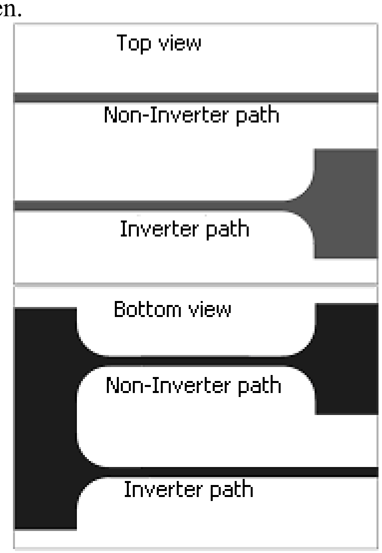


Fig. 11. Inverter / Non Inerter Scheme

A 50Ohms Double-sided parallel-strip line connects the input and output ports of the inverter. Transitions between the Double-sided parallel-strip lines (balanced lines) and the microstrip lines (unbalanced) are needed. Different types of these transitions (called baluns) are analyzed in [15]. Circularly tapered transitions are used in the presented design.

The Inverter/Non-Inverter circuit has two different paths: one performs the pulse inversion whereas the other does not change the pulse sign. In Fig. 11 is depicted the final layout of this stage including the top and bottom layers. The Non-Inverter branch line has the same length as the Inverter branch. This is necessary to maintain both ways synchronized and same strip loses.

D. Power Combiner

A Wilkinson Power Divider is used to combine the inverted and the non-inverted pulses. Wilkinson structure has been chosen due to good matching in all ports and good isolation between the combined ports. To increase the useful bandwidth a Four-Stage Wilkinson Power Combiner is designed [16].

A complete circuit composed of the “Inverter/Non-Inverter” and the “Wilkinson Combiner” has been built. In order to provide a direct connection to the foreseen Double-Sided strip antenna, the Wilkinson Combiner is also based on Double-sided parallel-strip lines.

The characteristic impedance of a double-sided parallel-strip line with dielectric separation h is twice the characteristic impedance of a microstrip line with dielectric thickness $h/2$ (with the same strip width) [15]. Resistance values used in double-sided parallel-strip

Wilkinson combiner also changed with reference to former microstrip configuration. The used structure uses 50 Ohms instead of 100 Ohms placed in microstrip.

E. UWB Antenna

In order to make measurements we have implemented a discone antenna due to its omni-directional radiation property and broadband characteristic. Fig. 12 shows the main the mechanical diagram of the developed antenna.

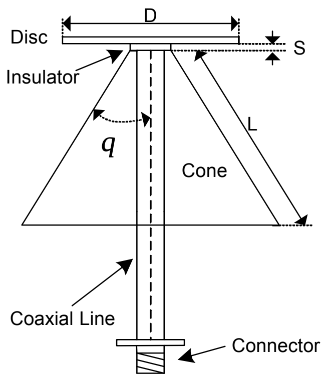


Fig. 12 Discone Antenna Diagram

To calculate the dimensions we used [17][18],

$$L(inch) = \frac{2958}{F} \quad (13) \qquad q = 25^{\circ} - 40^{\circ} \tag{14}$$

$$D(inch) = \frac{2008}{F} \tag{15}$$

where F is the starting frequency in *MHz*. The cone length L and disc diameter D fix the starting frequency of the antenna. To achieve the maximum frequency range the separation between the disc and the cone S should be as short as possible.

In Fig. 13 is shown the input return loss measured at the network analyzer. The parameter S11 is kept below *10 dB* from *1.5GHz* to frequencies above *11GHz*.

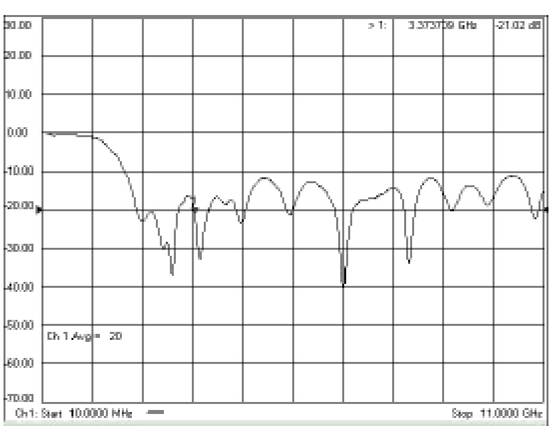


Fig. 13. Antenna Input Return Loss (dB)

F. Reference Oscillator

Considering that UWB pulses are about 200 ps width, the reference oscillator clock that controls the transmission or the reception must have very low jitter or, in other terms, very low phase-noise. If this fact is not achieved synchronization between the transmitter and the receiver will be highly degraded.

A low jitter clock has been implemented using Maxim MAX2620 low phase noise oscillator (-110dBc/Hz) and Hittite HMC394LP4 frequency divider. The oscillator can work at frequencies up to *1 GHz*. The divider consists on a 5 bit programmable digital counter, which provides lower frequency clock signals. Fig. 14 shows jitter measurement of the developed reference oscillator. The result gives a rms-jitter of *1.7 ps* and a peak-to-peak jitter of *10 ps*.

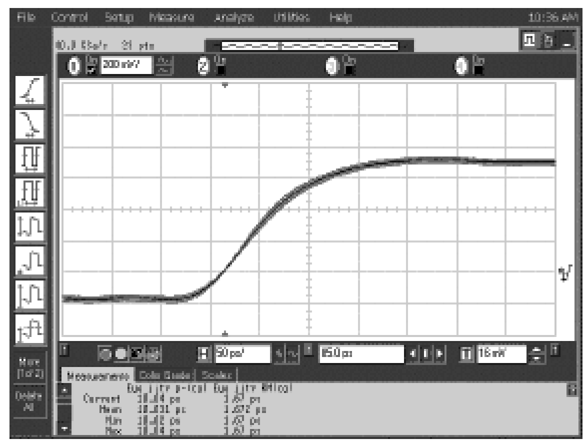


Fig. 14. Jitter measurement

IV. MEASUREMENT RESULTS

The whole transmitter, as shown in Fig. 10, has been fabricated in microstrip technology using a substrate TACONIC RF30-0300, having *0.76mm* thickness and a relative permittivity $e_r = 3$. The SRD diodes by Micrometrics (MSD700-19-1) have a switching time of about *70ps*.

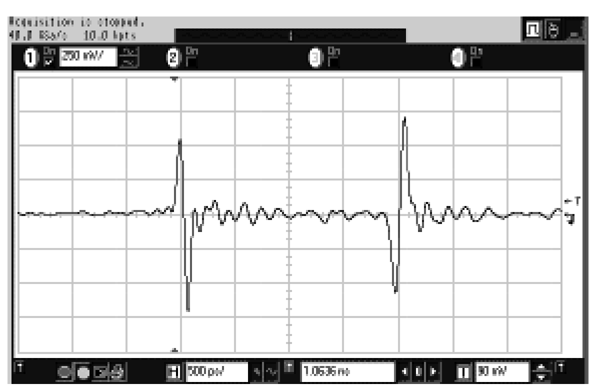


Fig. 15. Combined pulses at the same port..

The implemented transmitter generates two Gaussian monocycle pulses. The CPLD generates pulses through two output ports that are connected to their respective SRD pulse generator. Once the two pulses are generated, the Inverter/ Non-Inverter block produce a positive and a negative pulse needed when using DS-SS codes. Finally, the combiner permits to transmit the pulses using the same antenna. Fig. 15 measurement result when both, the inverted and the non-inverted pulses are generated through the common output port.

Using TH-SS codes allows to reduce the peak-to-average ratio in the PSD, or the spectrum efficiency as defined in (8). Fig. 16 shows the measurement result when TH and DS spreading codes are used. The spectrum efficiency has been reduced to *10 dB* as the theoretical result of Fig. 5 shows. The maximum output level is about *–45dBm/Mhz*, which is close to the permitted limit by FCC mask (*–42dBm/Mhz*). Finally note that total 10dB bandwidth of the pulse goes from 2 to 9 GHz. This corresponds to a Gaussian monocycle pulse having $t = 50ps$ and a pulse width of about *185ps*.

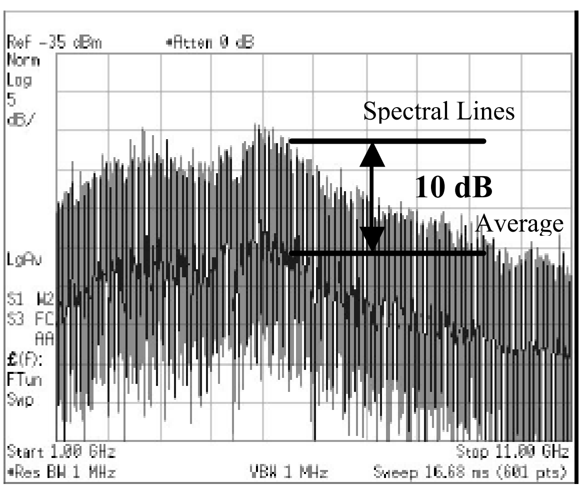


Fig. 16. Generated PSD using TH – DS Codes

V. CONCLUSIONS

This paper presents an UWB transmitter based on impulse radio, which is a hot research topic since the FCC delivered the regulations regarding UWB in 2002. The main concern in systems compliant with the FCC mask is the efficient usage of the available power below the mask limitations. To this end, the use time hopping and direct sequence spreading codes are used to reduce the level of the spectral lines in the radiated PSD. Maximum length sequences for TH codes are presented as a good compromise between complexity and the achieved spectral efficiency.

The developed transmitter uses an SRD-based circuit and a passive shaping network to generate a Gaussian monocycle. The measured pulse width is about 185ps. A Class-S digital pulse amplifier using conventional RF transistors has been designed to drive SRD pulse generator that can work up to 40MHz. To implement the DS spreading codes, a pulse inversion block has been developed using UWB transitions between parallel doubled sided strip lines and microstrip lines, shown an ideal performance for this application. The MAXIM 2620 shows very good Jitter performance for UWB systems, which presents an rms-jitter of *1.7ps*. With respect to the antenna, a discone antenna is developed to perform field measurements. This antenna is operative from 1.5GHz up to 11GHz.

Measurements on the generated PSD show a maximum level of –45dBm/MHz, which is kept below the mask limitation at –42dBm/MHz. In addition the measured PSD using TH and DS spreading codes corresponds to theory predictions giving a ratio between the peak power of the spectral lines and the pulse spectrum of about 10dB.

REFERENCES

[1] J. Romme, L. Piazzo, "On the Power Spectral Density of Time-Hopping Impulse Radio", Ultra-Wideband Systems and Technologies, 2002, Digest of papers 2002, IEEE Conf. on, May. 2002.

[2] C. Müller, S. Zeisberg, H. Seidel, A. Finger, "Spreading Properties of Time Hopping Codes in Ultra Wideband Systems", IEEE 7º Symp. On Spread-Spectrum Tech. & Appl., Prague, Czech Republic, Sept. 2002.

[3] X. Chen, S. Kiaei, "Monocycle shapes for Ultra Wideband System" Proceedings IEEE 2002.

[4] Sheng, P. Orlik, A.M. Haimovich, L. J. Cimini, Jr. J. Zhang, "On the spectral and Power Requirements for Ultra-wideband Transmission", Proceedings IEEE 2003

[5] M.Z. Win, R.A. Scholtz, "Impulse Radio: How It Works", IEEE Commun. Letters, Vol. 2, No. 2, Feb. 1998

[6] S. Benedetto, E. Biglieri, *Principles of Digital Transmission With Wireless Applications*, Kluwer Academic / Plenum Publishers, New York 1999.

[7] O. Albert, C.F. Mecklenbräuker, "Low-Power Ultra-Wideband Radio Testbed for Short Range Data Transmission", Int. Workshop on Ultra-Wideband systems (IWUWBS 2003), June 2003.

[8] M.Z. Win, "A unified Analysis of Generalized Time-Hopping Spread-Spectrum Signals in the Presence of Timing Jitter", IEEE Journal on Selected Areas in Communications, Vol. 20, No. 9, Dec. 2002.

[9] S.V. Maric, E.L. Titlebaum, "A Class of Frequency Hop Codes with Nearly Ideal Characteristics for Use in Multiple-Access Spread Spectrum

Communications and Radar and Sonar systems", IEEE Trans. On Commun., Vol. 40, No. 9, Sept. 1992.
[10] H. L. Krauss, Ch. W. Bostian, F. H. Raab, "Solid State Engineering", John Wiley & sons, Canada 1980.
[11] "Pulse and Waveform Generation with Step Recovery Diodes", Hewlett Packard AN-918.
[12] J.S. Lee, C. Nguyen, T. Scullion, "New Uniplanar Monocycle Pulse Generator and Transformer for Time-domain Microwave Applications", IEEE Trans. On Microwave Theory and Techniques, Vol. 49, No. 6, June 2001.
[13] A. Mollfulleda, P. Miskovsky, J. Mateu, "Rubost Shaping Network for Impulse Radio and UWB Signal Generator", 2005 IEEE MTT-S International Microwave symposium, 12-17 June 2005.
[14] K. C. Gupta, R. Garg, I.J. Bahl, and P. Bhartia, *Microstrip Lines and Slotlines*, 2nd ed.Artech House, Inc., 1996.
[15] S.G, Kim, K. Chang, "Ultrawide-Band Transitions and New Microwave Components Using Double-Sided Parallel-Strip Lines", IEEE Trans. On Microwave Theory and Tech., vol. 52, No. 9, September 2004.
[16] D.M. Pozar, "Microwave Engineerin, 2nd Edition", 1998, John Wiley & Sons, Inc.
[17] Matthew R. Wills , A Different Antenna for the Mode-Stirred Chamber , http://www.ieee.org/organizations/pubs/newsletters/emcs/fall01/wills.html
[18] A UHF Discone Antenna for scanners, http://www.northcountryradio.com/Articles/discone.htm

Signal Processing and Spectrum Use

Brett Glass
Advanced Communication Concepts
Phone: (307) 761-2895
Email: brett@lariat.org

Paper unavailable at time of printing.

Tri-Band RF Transceivers for Dynamic Spectrum Access

Nishant Kumar, Yu-Dong Yao, Terrance Bernard, and Amitoj Singh

Wireless Information Systems Engineering Laboratory (WISELAB)
Department of Electrical and Computer Engineering
Stevens Institute of Technology, Hoboken, NJ 07030
Email: {nkumar, yyao}@stevens.edu

Abstract: *In order support dynamic spectrum access research and conduct DSA related experimentations, we have developed an RF transceiver test bed which operates over three industrial, scientific and medical (ISM) bands (902-928 MHz, 2.4-2.4835 GHz, and 5.725-5.850 GHz). The paper describes an approach where a baseband signal is up-converted at the transmitter to the ISM band and then down-converted to the base band at the receiver. The paper shows the basic architecture of the design and concludes with some results of an experiment done using this test bed.*

1. Introduction

In a recent spectrum reform policy document [1], four basic spectrum usage models were considered, (a) government planning and close supervision of spectrum, (b) spectrum as private property, (c) spectrum access markets, and (d) spectrum commons. In order to improve spectrum utilization, the concept of dynamic spectrum access (DSA) [1, 2] was proposed by FCC Spectrum Policy Task Force. Key functions in implementing DSA include sensing the radio frequency (RF) environment, identifying available channel resources, and adapting transmissions through the available channels. We focus our research on the DSA RF environments, specifically, the interference and noise characteristics, and developing access and resource allocation algorithms for DSA systems. In order to support our ongoing research and conduct DSA related experimentations, we have developed an RF transceiver test bed which operates over three industrial, scientific and medical (ISM) bands (902-928 MHz, 2.4-2.4835 GHz, and 5.725-5.850 GHz).

The outcrop of the recent wireless devices can be attributed to ISM bands which were initially reserved internationally for the non-commercial and unlicensed use of RF electromagnetic fields for industrial, scientific and medical purposes. They are unlicensed but are governed by rules made with respect to the transmission technology and power. These rules have helped in the congenial growth of short range devices and made helped in the development of technologies which are not only power efficient but has also increased the density of transmitting unit to an unprecedented number. The FCC regulation not only allowed non-governmental uses of these frequencies, but also the unquestioned use of radio spectrum giving us the freedom to experiment new technologies and ideas.

There are three commonly used ISM bands, 902 – 928 MHz, 2.4 – 2.4835 GHz and 5.725 – 5.850 GHz. FCC specifies that all the transmission in these bands should be limited to frequency hopping and direct sequence spread spectrum and the peak power output of the transmitter shall not exceed 1 Watt. If transmitting antennas of directional gain greater than 6 dBi are used, the power shall be reduced by the amount in dB that the directional gain of the antenna exceeds 6 dBi. Frequency hopping systems shall have hopping channel carrier frequencies separated by a minimum of 25 kHz or the 20 dB bandwidth of the hopping channel, whichever is greater.

This paper introduces a test-bed targeted to exploit the three ISM bands. With the use of off-the-shelf components and simple design, transmission and reception can be done at any desired frequency in the ISM band. Easy control and simple interface allow it to be integrated with computers and support in the development of communication protocols, especially, for dynamic spectrum access systems. This setup can also be used for measurement of channel usage and noise level. A fast sweep through a target band of frequency and subsequently storing the data for further analysis can be easily archived.

2. Architecture and Descriptions

Figure 1 shows the block diagram for the first band (902 – 928 MHz) transmission and reception.

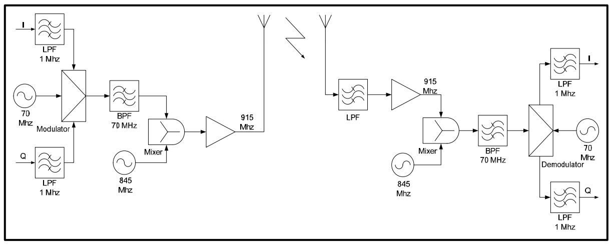


Fig. 1. Block diagram for 902 – 928 MHz band operation.

In the following, we describe the architecture and components in implementing the 902 – 928 MHz transceiver.

Transmitter: The basic components of the transmitter of the transceiver are the modulator, band-pass filter, mixer, local oscillator (LO) source and amplifier. The design also uses low pass filters at the IQ end to remove noise from the input base-band signal. The signal is quadrature modulated at 70 MHz. The 70 MHz needed for the modulation is provided by a frequency synthesizer. The modulated signal is passed through a band pass filter with centre frequency of 70 MHz. This removes noise from the modulated signal. The filtered signal is then fed into a mixer, which accepts the signal at the IF end and a frequency of 845 MHz by the frequency source at the local oscillator input and produces a carrier signal at 915 MHz. This signal is amplified and transmitted.

Receiver: The receiver end of the board reverses the operation of the transmitter. The received signal is first sent through a low-pass filter to remove most of the noise/interference. This signal is then fed into an amplifier before being sent into a mixer which receives a frequency of 845 MHz from the local oscillator and subtracts this from the 913 MHz signal input to have a signal of 70 MHz generated at the output. A band pass filter is used to filter this signal out and send the signal to the demodulator. The demodulator takes a 70 MHz input from the local oscillator to demodulate the signal which recovers the original signal.

Similar design is used for the other two ISM bands (2.4 – 2.4835 GHz and 5.725 – 5.850 GHz) with switches to switch over different frequency bands. Signal can be tapped at any point in the circuit and can be analyzed using an oscilloscope or a spectrum analyzer. Figure 2 shows a high-level block diagram for the tri-band operation.

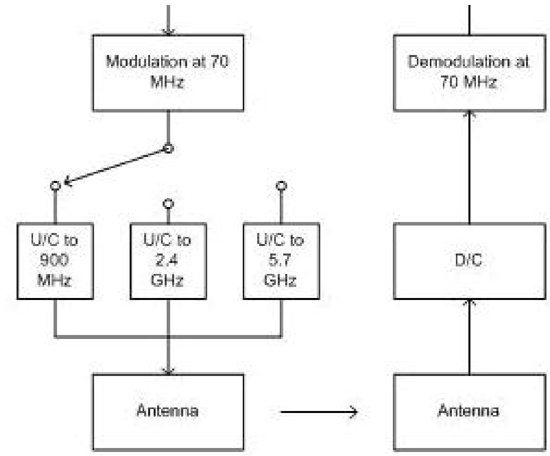


Fig. 2. A high-level block diagram for tri-band operation.

Figure 3 shows a developed prototype of the tri-band transceivers.

Fig. 3. A tri-band prototype.

3. Experiment

We have been using the developed tri-band for dynamic spectrum access investigation. The test bed is used to measure the channel and interference. The test bed facilitates in targeting a band of channel and sweeping through to measure the strength of signal. This can be done at baseband frequency or at the intermediate frequency.

Figure 4 (a) through (f) shows photos taken during a transmission/reception experiment.

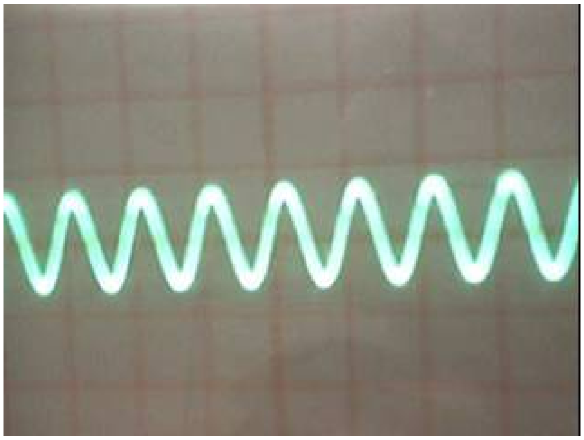


Fig. 4 (a). Input signal.

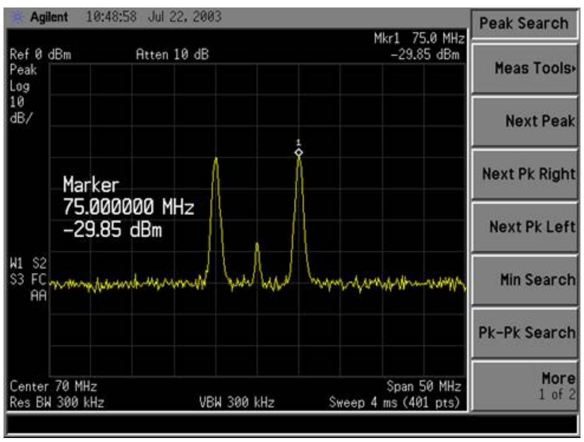


Fig. 4 (b). Signal modulated to 70 MHz.

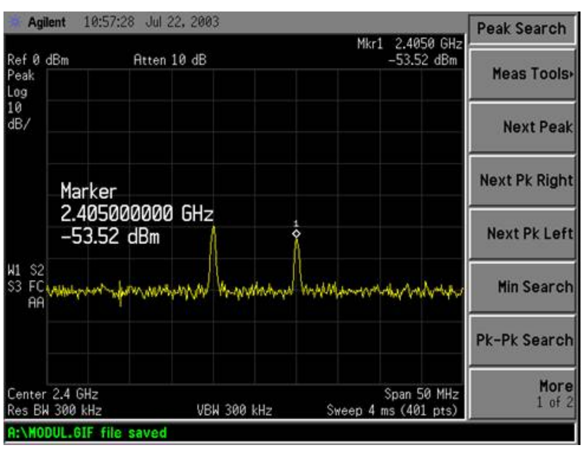


Fig. 4 (c). Signal up-converted to 2.4 GHz.

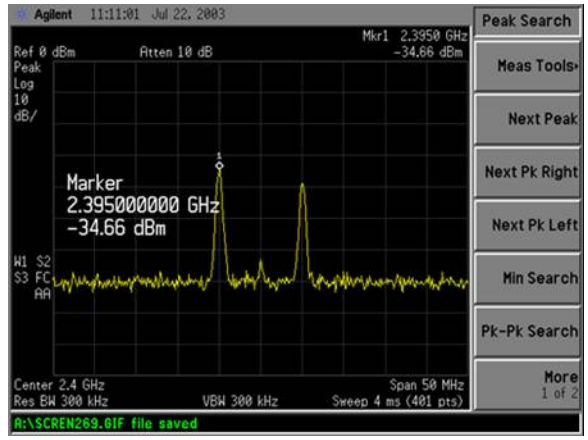


Fig. 4 (d). Signal transmitted.

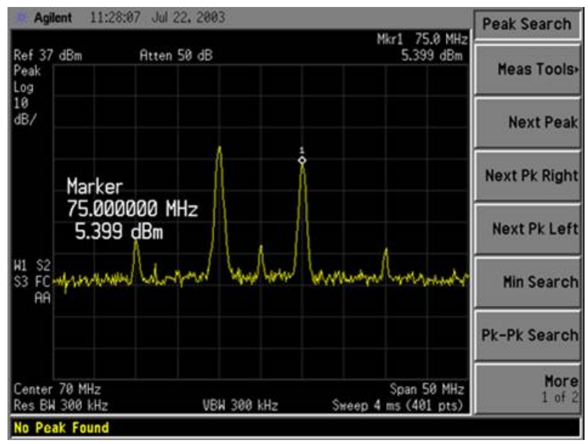


Fig. 4 (e). Signal down-converted to 70 MHz.

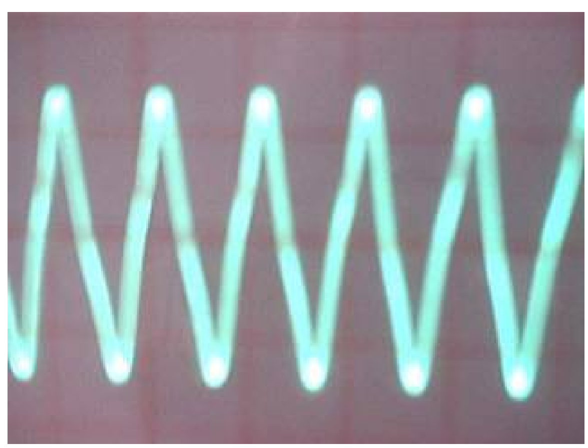


Fig. 4 (f). Output signal.

The above photos show the basic transceiver flow in a communication system where it starts with a baseband input signal, Fig. 4 (a), which is modulated by an intermediate frequency, Fig. 4 (b). This signal is then up-converted to 2.4 GHz frequency, Fig. 4 (c), which is amplified and transmitted, Fig. 4 (d). At the receiver side, the signal is filtered and then down-converted to the intermediate frequency, Fig. 4 (e), and finally, demodulated to baseband, Fig. 4 (f).

4. Conclusions

This paper first discusses the use of a tri-band RF test bed for dynamic spectrum access research. It describes the architecture of the test bed and its elements. The transmitter and receiver block diagrams and circuits are shown and the designs are discussed. An experimentation using the test bed using the test bed is also presented.

Some features of the developed test bed include flexibility to perform testing at every frequency stage and the components/designs can be changed according to the requirements of testing scenarios. The minimum detectable signal (MDS) is -62 dBm as compared to -52 dBm for typical off-the-shelf transceiver boards. The local oscillator stability is in the range between +/- 1 ppm and +/- 2.5 ppm as compared to +/- 75 ppm for typical off-the-shelf transceiver boards.

References:

[1] R. M. Entman, *Balancing Policy Options in a Turbulent Telecommunications Market: A Report of the Seventeenth Annual Aspen Institute Conference on Telecommunications Policy*, The Aspen Institute, 2002.

[2] P. Kolodzy, *Next Generation Spectrum Policy*, Spring Meeting of the Aspen Institute Conference on Telecommunications Policy, April 2002.

[3] *Frequency Spectrum Rules and Regulations,* FCC Rules: CFR Title 47 Part 18. (http://www.airlinx.com).

[4] B. Razavi, *RF Microelectronics*, Prentice Hall PTR, November 1997.

[5] H. Taub and D. L. Schilling, *Principles of Communications Systems*, McGraw-Hill Companies, 2nd edition, April 1986.

[6] *Mini-Circuits Component Datasheets*, (http://www.minicircuits.com).

[7] D. Grandblaise, D. Bourse, K. Moessner, and P. Leaves, “Dynamic spectrum allocation in composite reconfigurable wireless networks,” *IEEE Communications Magazine*, vol. 42, no. 5, pp. 72-81, May 2004.

[8] V. Brik, E. Rozner, S. Banerjee, and P. Bahl, “DSAP: A Protocol for Coordinated Spectrum Access,” *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 611 – 614, November 2005.

Multimedia Quality of Service and Net Neutrality on Wireless Networks

Tom Lookabaugh and Douglas C. Sicker

Interdisciplinary Telecommunications Program and Department of Computer Science
University of Colorado at Boulder
[Tom.Lookabaugh, Douglas.Sicker]@colorado.edu
(303) 492-0429
Fax (303) 492-1112

Multimedia traffic differs in its demand on networks from traditional internet access applications, in particular requiring sustained sessions and high rates. A shift towards multimedia traffic consequently has a direct impact on the amount of oversubscription – or sharing of capacity among users - that is acceptable on a given network. Oversubscription is disproportionally important to the economics of wireless internet access providers relative to other access technologies such as cable and DSL. This means that wireless providers can be expected to feel more strongly the economic rationales for discrimination among content based on traffic type. But such discrimination flies in the face of the principle of net neutrality – a social policy goal that attempts to minimize discrimination in order to encourage innovation and openness on the internet. The consequence is that we can expect wireless internet access to be at the center of the debate about reasonable policy approaches to network neutrality that attempt to separate more and less socially desirable economic rationales for discrimination.

1. Introduction

Net neutrality – the principle that internet access providers should not discriminate among different instances of internet traffic – is at the core of an important policy debate about how the internet will evolve. Proponents of net neutrality claim it is indispensable to the innovation and openness of the internet that society has come to expect and cherish. Skeptics argue that service providers either have every right to discriminate among traffic types since they own the networks (an argument that then runs into serious questions of what society wishes to permit in tradeoffs between property rights and social policy goals) or that there are simple and compelling economic reasons for discrimination among traffic types based on their consumption of network resources.

The latter argument – that of compelling economic rationales for discrimination – is sharp in the case of a shift from traditional internet access applications such as web browsing towards an increasing fraction of multimedia traffic such as voice, music, and video. Moreover, we argue here that wireless internet access providers are disproportionally affected by this shift relative to other service providers such as DSL or cable modem. The consequence is that wireless internet access providers are likely to find themselves among the first attempting to defend deviations from net neutrality for economic reasons and so constitute a key test case and audience for any policy proposals that seek to balance, encourage, or enforce the reasonable application of net neutrality.

2. Multimedia Traffic and Wireless

Multimedia traffic has different characteristics than traditional internet traffic such as web browsing. Web browsing is characterized by multiple short bursts of data (as successive transactions between a web browser and a server generate and fulfill requests for data). Performance is largely dictated by achieving low latency and high burst data rates while guaranteeing that data is delivered error free. Since high sustained data rates are not necessary, it is feasible to share a channel with high capacity among multiple users. File transfer is another common internet traffic type; the emphasis here is on high average rate and error free transmission, but with allowable significant variation in instantaneous rate.

Real-time multimedia traffic makes two fundamentally different tradeoffs: a need for sustained data rates (at a rate that depends on the application, such as voice, video, music, etc.) and some tolerance for errors [1]. Latency is important if the multimedia is part of an interactive communication and less relevant if not (for example, one way streaming of a television program). Non real-time multimedia traffic has similar needs to those of large file transfers, albeit likely with some higher tolerance for errors than, say, a file of financial information. Of course, error correction can be handled at higher levels of the protocol stack. Quality of

Service (QoS) is a term used to connote the set of requirements of an application or the state of a network in terms of meeting those requirements. As we've seen, for multimedia these tend to focus on attributes of data rate (or "bandwidth"), rate variation, latency, and error rate.

Different physical layer networks face different challenges in meeting particular QoS expectations. Wireless networks have two particular characteristics of interest: a shared physical channel and a propensity for variations in error performance (caused, for example, by fading during relative motion of transmitter and receiver) [2].

Traditional cellular telephone networks solve the QoS problem for interactive voice by emulating the traditional circuit switched infrastructure of the public switched telephone network. Calls are blocked unless capacity can be dedicated to the call and the channel can be expected to have a low enough error rate to sustain intelligibility; what errors do occur are mitigated through error concealment.

Recent wireless networks, though, are much more likely to select a packet switched infrastructure rather than circuit switched, with the goal of efficiently handling non media applications. Indeed, this trend is evident in almost all networks. QoS for voice calls – and for other multimedia traffic – then becomes a configuration and management choice for the network rather than an intrinsic characteristic of the network design. These modern wireless networks are frequently designed first via natural extensions of computer local area network principles rather than as extensions of PSTN principles; the QoS needs of multimedia are either presumed solved adequately by abundant over provisioning or through explicit QoS semantic overlays on networks. In either case, though, we tend to exacerbate a fundamental characteristic of the economics of wireless internet access: the expectation and reliance on shared use.

3. Wireless Internet Access Economics and Oversubscription

Internet access services typically advertise and compete on the basis of data rate. For example, a DSL offering current as of the date of this paper advertised "up to 1.5 Mbps download/896 kbps upload" [3] while a competing cable company offer advertised "up to an unbelievable 6 Megs" but later caveats this with "Actual speeds may vary and are not guaranteed. Many factors affect speed." [4] One of the most important factors that affect speed is oversubscription.

Oversubscription, also known as statistical multiplexing, is really a statement of the relationship between a marketing or advertising contention about the rate that a customer could enjoy (e.g., 1.5 Mbps) and the amount of capacity available on the network if all users are actively attempting to secure the advertised rate. For example, if a network has a total capacity of 15 Mbps and is shared among 100 users all of who are told that they are purchasing a 1.5 Mbps service, the oversubscription ratio is 10 – the ratio between the 150 Mbps that would be required to serve all users simultaneously and the 15 Mbps that is actually available. Oversubscription ratios are a design choice to be made in any part of the network that is shared. In wireless and cable modems the last mile infrastructure is shared, though not in DSL. However, all networks share capacity in internet switching, routing, backhaul, and transport.

At first glance, oversubscription sounds like service provider fraud, but it is not for both legal and practical reasons. Legally, service providers use the terminology "up to" in describing speeds as well as other qualifications as we saw earlier in this section. Practically, as we indicated in the previous section, web surfing performance is largely determined by being able to support short high speed bursts. Providing request for bursts are not often coincident, it may well be true that a user on a 10x (or more) oversubscribed network sees the advertised rate (or even a better rate) almost anytime that it is important to the user experience.

Oversubscription ratio choices are usually considered proprietary information by service providers. This leaves end users to discipline service providers through a qualitative assessment of service quality and then expressing their preferences through purchase decisions. Of course, when there are only one or two broadband options in the market, expressing a preference can be somewhat limited. Because of the lock-in effects around service decisions as well as variability and noisiness in qualitative assessment, we can expect this effect to be present but slow and fickle. In spite of a lack of public information, industry trade publications suggest that service providers select oversubscription ratios ranging from 10 to 100 [5,6,7]. Wireless service providers, with both shared physical last mile links and shared transport and backhaul, engage in oversubscription.

The extent to which a particular physical network type is cost sensitive to oversubscription depends on the relative share of network cost that is devoted to shared infrastructure versus dedicated (per customer) infrastructure. The component of the network that is devoted to transport and backhaul of internet traffic tends both to be more common among different physical network types and a smaller fraction of total cost than the portion dedicated to last mile link access.

Consequently, most variation in sensitivity to oversubscription will arise from the last mile link.

DSL links do not share the last mile link; a dedicated copper twisted pair is used to carry traffic to each end user, although there is sharing in the DSL access multiplexer (DSLAM) at the telephone company's central office. Both cable systems and wireless internet access systems are more thoroughly shared. Cable systems share the bandwidth dedicated to cable modem use among a number of end users over a portion of the electromagnetic bandwidth on coax cables that reach all the end users. Similarly, wireless systems share capacity in the electromagnetic spectrum – either licensed or unlicensed – among users. Consequently, we can expect oversubscription to be a more important economic driver for cable and wireless networks than for DSL networks.

Further, there are subtle but important differences between cable and wireless networks. Both cable and wireless networks increase per user bandwidth (with increasing demand or increasing numbers of subscribes in a location) by geographic splitting: the subdivision of a shared area into smaller units that are independently shared. In cable plant this is called node splitting; in wireless plant it is called cell splitting. Cable node splitting involves adding additional electronics to a node or placing aggregation nodes deeper in the cable plant (closer to end users); importantly, though, these nodes are still located within the plant that the cable company already owns and maintains. Wireless cell splitting may involve the more expensive task of finding, renting or purchasing, and equipping suitable additional cell sites in real property not owned (or net yet owned) by the operator (new or existing towers, new building placements) as well as adding backhaul capacity to these locations. Moreover, a key piece of shared infrastructure, the wireless base station, is the subject of intentional cost shifting using advanced radio technologies – adding cost in the RF and signal processing portions of the base station in return for cheaper subscriber terminals. This is an eminently sensible choice in terms of overall system economics, but has the side effect of increasing the sensitivity of wireless system economics to oversubscription choices relative to those of other technologies.

The overall impact of these physical differences is that wireless network economics are likely to be more sensitive to oversubscription choices than either DSL or cable networks.

4. Oversubscription and Multimedia QoS

If wireless internet access network economics tend to be more sensitive to oversubscription ratios than other network types, it follows that they will be disproportionably impacted by changes in usage patterns that affect the relationship between oversubscription ratios and user perceptions of service.

In Section 2, we noted that multimedia traffic has differing QoS requirements. In particular, real-time multimedia traffic is (usually) not bursty but involves a sustained and steady demand for capacity. Non real-time multimedia traffic is similar to file transfers in that it benefits from high average capacity. However, multimedia files tend to be large relative to many other file types, so we can consider non-real time multimedia to represent the special case of large file transfers.

Unfortunately, both the case of a steady demand for capacity or the case of multiple large file transfers work directly against the acceptability of high oversubscription ratios for end users. In our 10x oversubscription example from section 3, if any ten of the 100 users attempt to access their "up to 1.5 Mbps" by streaming 1.5 Mbps video streams – rather than doing bursty web surfing – the remaining 90 customers are out of luck (or, more typically, all subscribers will see tangible service degradation). Similarly, if many users decide to simultaneously download as files whatever audio-video content they find compelling on their hard drive based computers or portable devices, all will quickly see much lower than the "up to" rates that were advertised a the network capacity saturates with sustained continuous data transfer sessions.

Hence, a shift towards increasing multimedia traffic as a fraction of total traffic on internet access networks will increase pressure on service providers to lower oversubscription ratios, and wireless network providers will be disproportionally affected relative to cable or DSL providers. To some extent the impact can be managed by explicit QoS configuration choices, but the overall impact is unavoidable.

5. Net Neutrality

In simple terms, network neutrality advocates contend that consumers should be able to access the content and applications of their choice without network carriers discriminating among these choices[1]. For example, under network neutrality a consumer would be able to use her Vonage phone service without the carrier blocking or otherwise degrading that service. However, there are a number of situations where a carrier might justify discriminating against certain content and applications. The obvious one would be whether the content was legal. Another would be based on network

[1] It should be pointed out that the concept of network neutrality is not new. It has its origins in various prior regulatory concepts of common carriage.

management requirements; in other words, the carrier claims that it is necessary to limit traffic flowing across its network in order to maintain certain performance expectations. Still another would be based on the need to limit access for network security reasons. This might include the need to prevent malicious reconnaissance such as vulnerability scanning or block the delivery of packets that might contain malware (such as computer viruses and worms). An arguably more contentious justification could be based on the claim that the content or application provider should compensate the carrier for the use of its network. Indeed, AT&T CEO Ed Whitacre was recently quoted as saying, "What [Google, Vonage, and others] would like to do is to use my pipes free. But I ain't going to let them do that." [8] Similarly, the CEO of Verizon, Ivan Seidenberg, stated that broadband application and content providers should "share the cost" of operating broadband networks. [9] In terms of wireless network neutrality, in November 2005 House Committee meeting, a Verizon Wireless executive stated that network neutrality should not apply to wireless carriers and indicated that carriers should be allowed to block traffic as they thought appropriate. [11]

In the policy statement Federal Communications Commission (FCC) 05-151,[2] the Federal Communications Commission stated four principles directed at ensuring that "broadband networks are widely deployed, open, affordable, and accessible to all consumers". [10] They include:

- To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to access the lawful Internet content of their choice.
- To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to run applications and use services of their choice, subject to the needs of law enforcement.
- To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to connect their choice of legal devices that do not harm the network.
- To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to competition among network providers, application and service providers, and content providers.

They also stated that they have "jurisdiction necessary to ensure that providers of telecommunications for Internet access or Internet Protocol-enabled (IP-enabled) services are operated in a neutral manner". [10] In other words, if they want to enforce network neutrality they have the authority. However, a footnote in the statement indicates that the principles are subject to "reasonable network management". This of course raises the question as to what constitutes reasonable network management; the concern here being that network management becomes a justification for selective content blocking and other similar measures.

In March of 2005, the FCC fined Madison River Communications (a rural phone company) for blocking connections to Internet phone providers (such as Vonage). As a result of FCC pressure, Madison River also agreed not to block such calls in the future. In this case, while several of the aforementioned justifications for blocking content were provided, the FCC found none of these as acceptable. [8] This action together with the policy statement suggests that the FCC does view network neutrality as a socially desirable goal.

This situation of network neutrality creates a number of interesting tensions. At the heart of this issue is the reasonable desire of the carriers to pass on the cost of network, but this is countered with the concern that the carriers will exhibit anticompetitive behavior and thereby erode much of what has made the internet successful. Telecommunications regulators have decided to move away from the heavily regulated model to embrace a model based on market discipline. However, with (at best) only two broadband players in most markets, it is difficult to belief that competition will produce the desired effect. We described earlier that the Internet is shifting away from the best-effort model where all bits are essentially treated equally to a model where certain bits can be given priority on the network; in other words, be provided differentiated service. As network carriers move to this differentiated service model, they will have the ability to decide whose bits are provided with priority treatment. This decision will likely be based on favoring their own bits or bits from carriers who pay for priority treatment. The concern is that those who don't pay could be relegated to a service level that doesn't support their application demands. What makes this a big concern is that the innovation of the internet has occurred not by the carriers but by the application providers.

In thinking about network neutrality within the wireless context, there exist an interesting history that might influence the thinking of both policy makers and consumers. Our common view of wireless is in terms

[2] It should be mentioned that these principles echo the earlier "Four Freedoms" described by Michael Powell.

of cellular telephony and WiFi (and likely more the former than the later). The cellular systems in the U.S. have a long history of using proprietary equipment to maintain control of equipment used on their networks. They also have a long history of embracing closed architectural choices and implementing differentiated pricing models for service selection (for example, paying extra to access the Internet on your cellular phone). In this sense, it should not be surprising to see these carriers taking steps that would violate some aspects of network neutrality. However, users have recently come to think of wireless to include WiFi and most implementations of WiFi (keeping with the traditional Internet model) do not violate network neutrality. As more handsets begin supporting both cellular and WiFi services, it will be interesting to see how wireless carriers respond to users that might make use of their handsets in ways that undermine potential profits for the wireless providers. It is not hard to imagine that a carrier would attempt to block handover to a "free" network or restrict the handover to a WiFi service that partners (i.e., pays) the carrier.

The issue of wireless network neutrality has arisen in a separate policy debate, namely municipal WiFi.[3] In a Request For Proposal (RFP) concerning the development of a citywide WiFi network put out by the city of San Francisco, reply comments contained language directly targeting issues of network neutrality. Quoting from the list of "demands", replies included (but were not limited to the following): [12]

- Open Access and Network Neutrality
- Assurances that free services don't lag substantially behind premium services, possibly requiring free services have at least one third the bandwidth of the premium service.
- Free service that is robust enough for general web use, email and messaging, and VOIP (in the initial rollout)

What makes this interesting is the explicit public call for network neutrality within a wireless network. Also of interest is the explicit demands for measurable bandwidth characteristics and support for realtime services such as VoIP. This shifts the network neutrality debate into defined and measurable expectations.

6. The Interaction of Net Neutrality, Wireless, Multimedia, and QoS

At the end of Section 4, we noted that wireless service providers building new access networks around traditional shared access models that are effective for internet access like browsing will find themselves disproportionally affected relative to their DSL and cable competitors by a shift towards more multimedia traffic on such networks. Such a shift would be a consequence of the use of Voice over IP (VoIP) as an increasing or even a sole method of providing telephony applications and increasing streaming and downloading of music and video content on wireless networks. We also saw in Section 5 that a primary caveat to the principle of net neutrality is the need for "reasonable network management."

Service providers can make a plausible case that avoiding material degradation in the end user experience as the traffic mix shifts is an eminently practical form of "reasonable network management" and that either the service provider needs to manage the actual mix (by, for instance, restricting the amount of multimedia traffic allowed on the network or how it is treated through explicit QoS mechanisms) or by charging end users, application providers, or both at different rates depending on some measure of usage. At the same time, there will be an abiding concern that service providers might consider other reasons to stray from network neutrality, such as discriminating in favor of forms of traffic in which the service provider has a commercial interest. It will be challenging to separate these different rationales for deviation from network neutrality and an important subject of technical, economic, and regulatory analysis over the coming years. And, as we have seen, wireless internet access service providers are likely to be at the sharp point of this debate as they will feel most quickly and strongly the effects of shifts towards multimedia traffic.

7. Conclusion

We have argued here that a shift towards more multimedia traffic (such as VoIP, audio, and video) on internet access networks tends to have disproportionate impact on the economics of wireless network internet access providers compared to cable or DSL providers. This means that wireless providers are likely to feel most strongly the need to deviate from network neutrality – a policy that the FCC and others have argued is socially desirable - for reasons of "reasonable network management." The implication is that the

[3] The debate over municipal WiFi concerns whether or not a municipality should be allowed to build and/or offer a WiFi service. The carriers have opposed this concept and have contended (among other things) that municipalities are not capable of maintaining such networks. Their efforts to stop the build out of municipal WiFi have mostly focused on the state legislatures and they have been successful at getting a number of states to pass such legislation. The matter is now being considered at the Federal level.

coming debate about how to separate reasonable exceptions to network neutrality from ones that are deemed anti-competitive or not in the public interest will be particularly sharp in the case of wireless providers.

As the policy debate and analysis around network neutrality continues, we can expect to use wireless access as an important test case for the impact of different policy approaches. Moreover, we will find net neutrality and other policy goals – such as the viability of wireless internet access as the so-called "third broadband pipe" in creating a competitive environment for broadband access – to become intermingled. Hence we see the evaluation of wireless network economics as a critical and ongoing component of this important policy debate.

8. References

[1] Wang, Zheng. *Internet QoS: Architectures and Mechanisms for Quality of Service*. Morgan Kauffman, 2001.

[2] Rappaport, Theodore. *Wireless Communications: Principles and Practice*, 2nd ed. Prentice-Hall, 2001.

[3] Qwest, "Residential: Internet/DSL," at https://iot.qwest.com/iot/internet/DslPickService.do?npa=303&nxx=413&sfx=1121&isp=&offerId=&prodDescCode=, visited 1/24/06.

[4] Comcast, "Special Offer" http://www.comcast.com/Benefits/CHSIDetails/Slot3PageOne.asp?LinkID=366, visited 1/24/06.

[5] Mitchell, Tom, "Avoiding the Pitfalls of Oversubscription in DSL Networks," *xchange*, at http://www.x-changemag.com/articles/041feat1.html, visited 1/24/06.

[6] Newman, Stagg. "Broadband Access Platforms," McKinsey and Company, April 14, 2005, at http://www.fcc.gov/oet/tac/april26-02-docs/BB-Access-Tech.pdf. visited 1/24/06.

[7] Earthlink Enterprise T1/T3 – Frequently Accessed Questions, 2005, at http://www.earthlink.net/biz/highspeed/enterprise/pdfs/Enterprise_FAQ.pdf, visited 1/24/06.

[8] At Stake: The Net As We Know It, Business Week Online, http://www.businessweek.com/technology/content/dec2005/tc20051215_141991.htm, posted December 15, 2005, visited January 24, 2006.

[9] Verizon Says Google, Microsoft Should Pay For Internet Apps, InformationWeek, http://www.informationweek.com/story/showArticle.jhtml?articleID=175801854, Posted Jan. 5, 2006, Visited January 24, 2006.

[10] FCC 05-151, Policy Statement, Federal Communications Commission, September 23, 2005.

[11] Verizon Wireless: Scrap Network Neutrality, http://techdirt.com/articles/20051110/1313223_F.shtml, Posted November 10th, 2005, Visited January 24, 2006.

[12] Webpage: The Tech of TechConnect, http://leftinsf.com/blog/index.php/archives/525, posted, December 21, 2005, visited January 24, 2006.

Spectrum Sharing Using Cognitive Radio Technology

Paul Greenis
Adapt4
Phone: (321) 259-5009
Email: paul@adapt4.com

Abstract--*Cognitive Radios (CRs) are a classification of software-defined radios (SDRs) that include complex, embedded software that 'harvests' wireless bandwidth. This means that the CR monitors a defined band of spectrum and determines which portions are in use at any instant in time. Portions that are not used are candidates for use by the CR. It is the method used for the harvesting that is of particular interest to regulators and spectrum owners. CRs must take every precaution to conduct harvesting on a 'non-interfering' basis. The presentation will discuss the critical aspects of sharing this unused spectrum while minimizing the probability of interfering with the primary user or licensee. Several concerns must be addressed in the design of the CR to ensure all precautions are taken. This includes such things as transmit power levels, waveform selection, frequency hopping and use detection sensitivity.*
In addition to optimizing the CR spectrum access techniques, several usage detection challenges must be overcome. There are several conditions which may exist that could cause 'undetected' interference with the primary users. These include 'hidden transmitters' and 'silent receivers' The presentation will address what methods can be taken to overcome these challenges. Additional capabilities of the SDR will be discussed which allow it to also address interoperability within a wide range of users, such as public safety organizations. It offers a powerful platform to support waveforms and access techniques necessary to communicate with existing, legacy radios as well as networking features which support the dynamic creation of user-defined communities. These features also make it a 'good fit' for military and Homeland Security applications.

Public Safety
Environmental Noise Challenges
for Land Mobile Radio Vocoders

D.J. Atkinson
Institute for Telecommunication Sciences
Phone – 303.497.5281
FAX – 303.497.5969
E-Mail – dj@its.bldrdoc.gov

Abstract

This paper investigates the effect on vocoders of background noise that exists in environments where public safety officials and first responders need to communicate. Four different environments were examined: police cruiser, fire engine, rescue boat, and rescue helicopter. During the examination, noise levels were measured, as well as speech levels of actual practitioners working in that environment. Based on those results a controlled laboratory experiment was conducted to determine the effectiveness of vocoders used in communications equipment subjected to those environments. The experimental conclusion is that while digital vocoders do not perform as well as analog transmission in the presence of high levels of background noise, moderate amounts of noise cancellation prior to the signal injection into the vocoder can somewhat mitigate those detrimental effects.

1. Introduction

Public Safety agencies continue to have insufficient spectrum to adequately meet their communications requirements, particularly in population centers. One facet to solving this problem is to increase the number of channels in the currently allocated spectrum. Currently in process is a migration from 25 kHz voice channels to 12.5 kHz voice channels across public safety agencies. One result of using traditional analog modulation in a narrowband channel is that the analog voice quality suffers significantly. Therefore digital voice coder/decoders (vocoders) are necessary to effectively accomplish the transition.

The Project 25 effort, a partnership between the Association of Public-Safety Communications Officials (APCO), the National Association of State Telecommunications Directors (NASTD), the Federal Partnership for Interoperable Communications (FPIC), and the Telecommunications Industry Association (TIA), has produced a suite of standards for a digital radio communication system that utilizes a low-bit-rate digital vocoder to implement these narrowband voice channels. When the vocoder was initially adopted in the early 1990s, there was a significant amount of testing done to ensure that the public safety community was getting the best audio quality available. However, one aspect that was overlooked in the investigation was how the vocoder would behave in the extremely high noise level environments in which public safety practitioners must operate. Consequently, there have been many practitioner complaints about voice quality levels on digital communication systems, particularly in environments with high background noise levels.

Currently, Project 25 is considering the effects of further narrowbanding to 6.25 kHz-equivalent channels. As part of their investigation into the continued channel squeeze, the Project 25 Steering Committee, in conjunction with the FPIC initiated an investigation that would provide scientific results showing the behavior of vocoders in the presence of high levels of background noise. Included in this study were the original Project 25 vocoder (7200 bits per second), a compatible vocoder at the same bit rate that includes enhancements developed in the years since the initial adoption, and a half-rate vocoder (3600 bits per second) in the same technology family.

This paper provides an overview of the investigation, including the noise measurement and sampling, and the design and conduct of the subjective test. Finally, a summary of the subjective testing results is presented.

2. Acknowledgements

The author of this paper wishes to express thanks to the following, whose assistance was essential to the success of the experiment:

- The City of Mesa, Arizona, and the US Coast Guard for providing opportunities to record high-quality public safety environmental noise.
- The Federal Partnership for Interoperable Communications (FPIC), sponsored by the DHS Office of the CIO.
- DHS Office of Interoperability and Compatibility's SAFECOM program, which provided travel arrangements for the listening subjects.
- The following communities/agencies for adjusting schedules and allowing their public safety personnel to travel to Boulder to participate in the experiment: City of Mesa, Arizona; City of Phoenix, Arizona; County of Los Angeles, California; City of Roseville, California; Yolo County, California; California Highway Patrol; City of Gurnee, Illinois; North Carolina State Highway Patrol; City of Alexandria, Virginia; Four Mile Rural Volunteer Fire Department, Colorado; US Postal Inspection Service.

3. Environmental Noise Sampling

To adequately determine signal-to-noise ratios (SNRs) at the microphone of a public safety practitioner's radio it was necessary to record and take sound pressure level (SPL) measurements of noise and speech+noise in several common environmental situations.

3.1 Noise Sampling

For each noise condition, a 48 kHz[1] sampling 16 bit-per-sample recording was made with two Beyerdynamic MCD 100 digital microphones[2] and a portable digital audio tape (DAT) recorder. In addition, SPL measurements were made.

For vehicle interior sampling, the microphones were separated by six inches and placed at ear level in a position equivalent to where a center passenger would sit in the front of the vehicle. The Beyerdynamic MCD 100 digital microphones have a cardioid pattern. The microphones were positioned such that the cardioid pattern was 60 degrees offset from straight ahead and crossing the center plane of the vehicle (i.e., the left microphone records sounds from the right half of the vehicle and vice versa). Figure 1 illustrates the dual cardioid pickup pattern achieved by this microphone placement, which provided excellent stereo separation while still providing good coverage from front-center, where a significant amount of noise emanates. The back-center is assumed to be the lowest noise point, and therefore it is reasonable to minimize sensitivity at this point in the recording field.

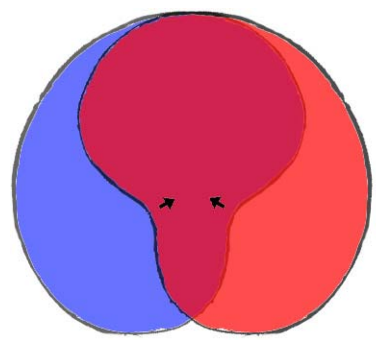


Figure 1. Dual cardioid pickup pattern.

For vehicle exterior sampling, the microphones were placed at 63 inches above ground (average ear height of a person of average height) and 5 feet from the passenger door of the vehicle in the same configuration as just described for interior sampling and shown in Figure 1, with the change that the pickup pattern was focused toward the vehicle.

SPL measurements were conducted using a Galaxy Audio CM-150 Sound Level Meter in both A-weighted and C-weighted[3] modes. For each sampling condition, measurements were made in the "forward" directions and at 60-degree increments throughout the horizontal plane. In the forward position, average, peak and minimum SPL were

[1] Sampling at 48 kHz provides audio frequency response equivalent to (or better than) that of a compact disc and provides an integral divisor when downsampling to 8 kHz, the sampling rate used at the input to the voice coder.

[2] Certain commercial equipment, materials, and/or programs are identified in this report to specify adequately the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the program or equipment identified is necessarily the best available for this application.

[3] A-weighted SPL measurements correlate with loudness for softer signals, while the broader band C-weighted SPL measurements correlate with loudness for louder signals.

measured. Only average SPL was measured in the other positions. Recording conditions and measured average SPL are summarized in Table 1. SPLs presented in the table are average values for the forward case. Accuracy of the instrumentation is ±1.5 dB.

Table 1. Summary of Noise Conditions and Levels.

Condition	Average Sound Pressure Level (SPL)	
	A wt.	C wt.
Fire – Idle – Interior	70 dB	86 dB
Fire – Fast Idle – Interior	84 dB	94 dB
Fire – Idle – Exterior	79 dB	88 dB
Fire – Fast Idle – Exterior	87 dB	98 dB
Fire – 30 MPH – No Siren	82 dB	94 dB
Fire – 30 MPH – Siren	88 dB	96 dB
Fire – 60 MPH – No Siren	86 dB	110 dB
Fire – 60 MPH – Siren	87 dB	110 dB
Police – 30 MPH – No Siren	62 dB	86 dB
Police – 30 MPH – Siren	75 dB	86 dB
Police – 60 MPH – No Siren	67 dB	91 dB
Police – 60 MPH – Siren	70 dB	88 dB
Helicopter – Exterior	105 dB	107 dB
Helicopter – Interior Idle	95 dB	107 dB
Helicopter – Interior Flight	96 dB	107 dB
Boat – 4.8 knots	69 dB	89 dB
Boat – 30 knots	88 dB	92 dB

Because a subjective test evaluating all of these noise environments would be unreasonably large, a subset of these environments was chosen for the subjective experiment. The environments chosen were:

- Fire Truck – 30 MPH – Siren
- Fire Truck – 60 MPH – No Siren
- Boat – 30 knots
- Helicopter – Interior Flight
- Police – 60 MPH – Siren

Figure 2 provides the power spectral density (PSD) plots for these noise samples, and shows that for most public safety noise environments, the majority of the noise power is very low frequency. The exception to this is the noise on an outboard boat, which has a peak power at approximately 200 Hz.

3.2 Speech+Noise Sampling

In addition to the noise recordings, a number of speech+noise recordings were made at 48 kHz sampling, 16 bits per sample. These recordings were made with three different microphones: a Beyerdynamic MCD 100 digital microphone, a Motorola HMN1061A ("Spectra") microphone, and a David Clark M-1/DC amplified dynamic noise canceling microphone. In addition, SPL measurements were made.

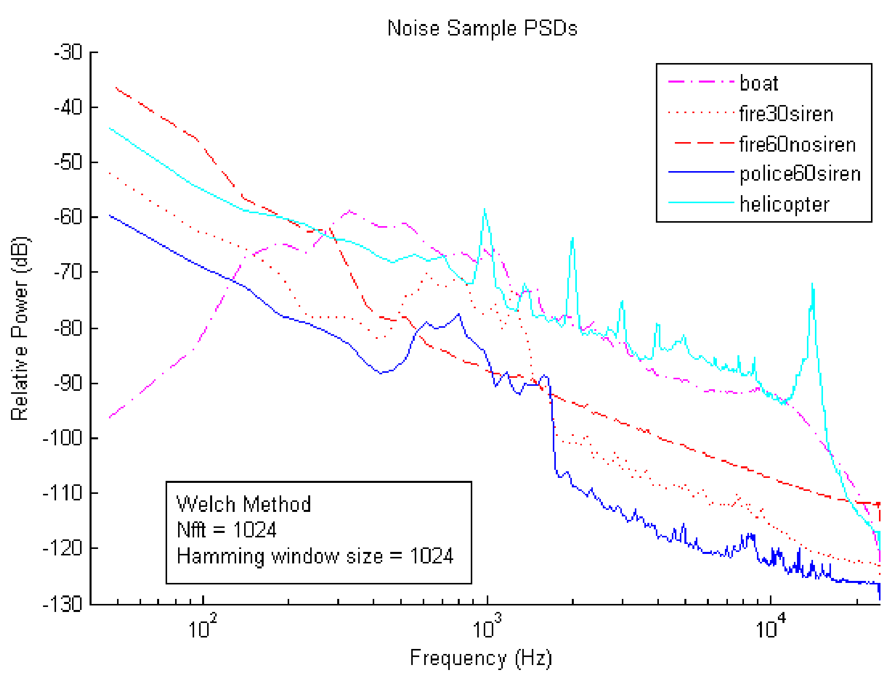


Figure 2. Power spectral density plots for noise conditions used in subjective vocoder experiment.

For speech+noise recording and measurements, the microphone was hand-held at the mouth reference point.[4] For interior samples, the talker was positioned in the front passenger/navigator/copilot seat facing the front of the vehicle. For exterior samples, the talker was positioned five feet from the passenger door of the vehicle and facing toward the vehicle. The recordings and measurements were used to identify SNRs in environments where public safety radio equipment is used. Table 2 summarizes the C-weighted SNRs observed with the CM-150 for the conditions used in the subjective tests.

Table 2. Summary of SNRs Observed in the Field.

Condition	SNR
Fire – 30 MPH – Siren	3 dB
Fire – 60 MPH – No Siren	-11 dB
Police – 60 MPH – Siren	9 dB
Helicopter – Interior Flight	-5 dB
Boat – 30 knots	0 dB

Because three of the conditions were considered extreme by the TIA TR8 Vocoder Task Group, they requested that additional conditions be added with a 6 dB gain in SNR, to mimic the use of a noise-canceling microphone. The added conditions were: Fire – 60 MPH – No Siren @ -5 dB SNR, Helicopter – Interior Flight @ 1 dB SNR, and Boat – 30 knots @ 6 dB SNR.

4. Subjective Experiment

4.1 Experiment Design

Design of a subjective speech quality assessment experiment is an exacting process. Several Recommendations from the International Telecommunication Union (ITU) address factors in the design of subjective experiments. The two primary recommendations that served as the guide for this experiment were ITU-T Recommendation P.800 [3] and ITU-T Recommendation P.830 [4].

Based on those Recommendations, a subjective voice quality experiment typically proceeds as follows. A test subject, or listener, is seated in a sound-attenuated chamber. Inside the chamber, the subject uses headphones to first set the volume at his/her preferred level. The subject then listens to several 5- to 7-second long pairs of sentences, rating the quality of each pair of sentences in turn. The subject is asked to provide their opinion of the overall quality of the speech that they hear on a rating scale that uses the terms “Excellent”, “Good”, “Fair”, “Poor”, and “Bad”. Quality ratings are converted to numbers on a five-point scale where 1 represents a “Bad” rating and 5 represents an “Excellent” rating. These scores are then averaged across listeners, talkers, and sentence pairs to compute an MOS.

The listening material was developed based on the following factors:

- Talkers: 4 male, 4 female
- Sentence pairs: 5 per talker
- Input level: -28 dB_{OV}[5] [5] [6]
- Vocoders: P25 Full Rate (FR)[7], P25 proposed Enhanced Full Rate (EFR), P25 proposed Enhanced Half Rate (EHR), and 16-bit linear PCM as a reference
- Background noise conditions: Clean (no background noise), Fire Truck (60 MPH, No Siren – 2 noise levels), Fire Truck (30 MPH, Siren), Police Cruiser (60 MPH, Siren), Helicopter (Cruising – 2 noise levels), Coast Guard Boat (30 Kts – 2 noise levels)
- Control conditions: modulated noise reference unit (MNRU) [8] at -5, 5, 15, and 25 dBQ.[6]

In addition to the use of realistic background noise, another novel factor in this experiment is the use of actual public safety practitioners as listeners. This experiment used 27 listeners that represented law enforcement, emergency medical services, fire, dispatch, and federal users.

4.2 Experiment Development Issues

During the development of the experiment, two issues arose that required some additional consideration to produce acceptable results. These were proper mixing of the speech and noise, and signal levels output from the MNRU tool at low signal-to-distortion ratios.

4.2.1 Mixing Speech and Noise

Because the noise levels used in this experiment were so high, it was critical that the mixing of speech and noise be done as accurately as possible to achieve the same levels that were observed in the field in such a way that the speech portion of the signal was always at -28 dB_{OV} when it was input to the vocoder or MNRU. Factors addressed included 1) compensation for the weighting factors used in the field

[4] The mouth reference point is 25 mm in front of the lip plane as defined by ITU-T Recommendation P.51[2].

[5] dB relative to the overload point of a digital system. In this case, the digital system is 16 bits per sample.
[6] dB of signal to distortion ratio

measurements, and 2) downsampling from the 48 kHz sampled recordings to the 8 kHz sampled files that were input into the vocoder software and played to the listeners. Production of a scaled, mixed, 8 kHz sampled speech file is accomplished in three stages: 1) computation of speech gain factor G_S, 2) computation of noise gain factor G_N, and 3) creating the file to be presented to the coder. The process is shown in Figure 3.

G_S is a gain correction factor required to present the 8 kHz sampled speech to the vocoder at -28 dB_{OV}. G_S is computed by taking 48 kHz sampled speech, downsampling to 8 kHz [6], and computing the active signal level [5][6] (ASL1). Once ASL1 is known, G_S is computed according to:

$$G_S = -28 - \text{ASL1}$$

G_S is computed for each sentence pair to be passed through the codec. G_N is the gain factor required to achieve the noise level such that mixing (summing) the noise and the scaled speech signal will result in the SNR specified in this experiment. Those SNRs were developed using measurements taken using a C-weighted scale, which is the appropriate weighting for environments above 85 dB_{SPL}[9]. Therefore, the gain correction must take into account the C-weighting to ensure that the SNR presented to the speech coder represents the conditions measured in the field.

First, the active signal level of the C-weighted speech must be computed. This is done by applying G_S to the 48 kHz sampled speech, filtering the speech through a C-weighted filter[7] (Wc)[10], and computing the active signal level (ASL2) [5][6].

Next, the active signal level of the C-weighted noise must be computed. This is done by mixing (summing) the left and right channels of the stereo noise recording, filtering the noise through a C-weighted filter (Wc), and computing the active signal level (ASL3) [5][6].

Using the SNR specified for the particular condition, G_N can then be computed according to:

$$G_N = (\text{ASL2} - \text{ASL3}) - \text{SNR}$$

G_N is computed for each speech file/noise file combination to be mixed.

[7] The frequency domain amplitude transfer function for the C-weighted filter is given by:

$$Rc(f) = \frac{12200^2 * f^2}{(f^2 + 20.6^2)(f^2 + 12200^2)}$$

where f is frequency in Hz.

Prior to mixing (summing) with the noise signal, the speech is scaled according to G_S. This ensures that the speech will be at -28 dB_{OV} after the speech signal is downsampled to 8 kHz.

Prior to mixing (summing) with the speech signal, the noise signal is scaled according to G_N. This ensures that the SNR of the mixed file will be correct according to the C-weighted measurements recorded in the field and specified in this experiment.

The scaled speech and scaled noise are mixed (summed), and downsampled to 8 kHz[6].

The final step in the process is the addition of a high-pass filter (HPF) prior to injection of the signal into the codec. This provides protection of the codec against the extremely high levels of low-frequency noise encountered in some of the environments being simulated. In addition, the filter provides a response that is representative of that provided by a microphone that might be used with a radio implementing the codec.

The HPF implemented here is a 2nd order IIR filter with F_c = 20.6 Hz and F_s = 8000 Hz. The Z-domain representation of this filter is given by:

$$Y(z) = 0.984 \cdot \frac{1 - 2z^{-1} + z^{-2}}{1 - 1.9679z^{-1} + 0.96816z^{-2}} X(z)$$

This filter was derived from the high-pass portion of the C-weighted filter used to compute the SNR.

4.2.2 MNRU Signal Level

During the development of the experiment, there was some concern that the high noise levels, relative to the -28 dB_{OV} speech signal, might cause clipping to occur in some speech files. During the process of ensuring that none of the mixed files contained clipped samples, clipping was observed in the output of the MNRU software tool at the -5 dBQ level. When noise is added to speech, and the noise level is similar to, or higher than, the speech level, the resulting combined signal will have a level that can be significantly higher than the original speech signal. The curve representing this effective gain factor is shown in Figure 4. At levels higher than 10 dBQ, the gain factor is negligible, less than 0.5 dB. At levels lower than 10 dBQ (5 and -5 dBQ for this experiment) the gain factor had to be compensated for by adjusting the level of the speech file prior to input into the software tool. The gain correction factors applied to the 5 and -5 dBQ conditions were -1.2 dB and -6.1 dB, respectively.

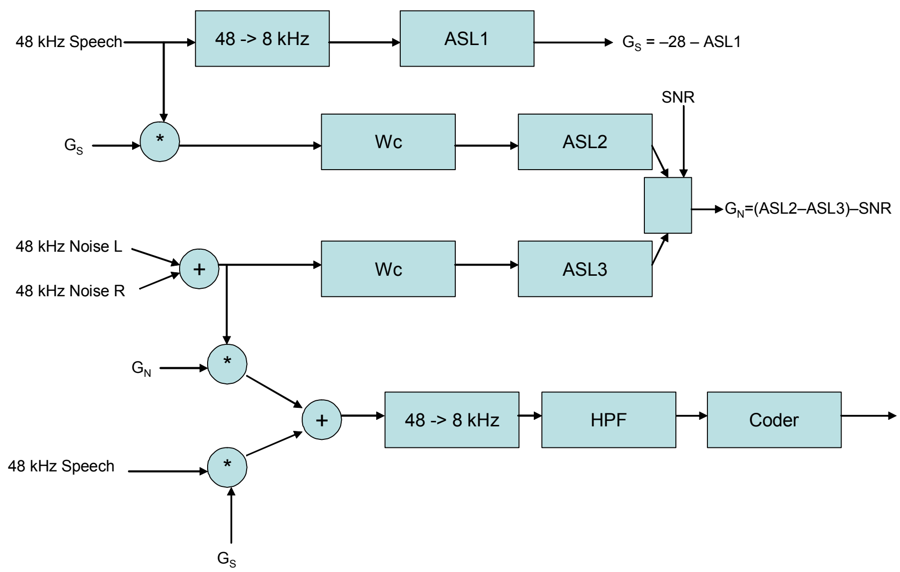


Figure 3. Process for producing a mixed speech plus noise file.

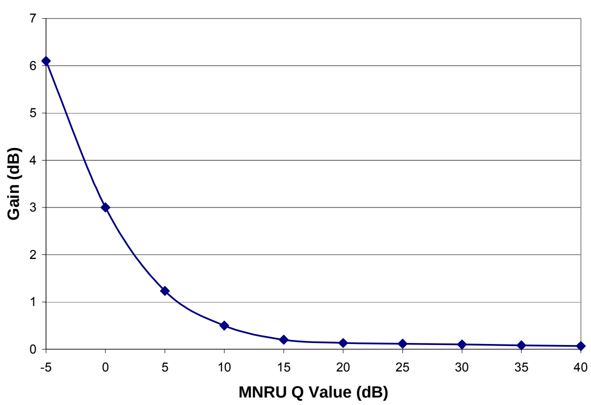


Figure 4. MNRU Gain by Q Value (dB)

5. Experiment Results

Figure 5 contains a bar chart of Mean Opinion Scores (MOSs) for the three vocoders under test and the reference conditions. The pattern in each bar repesents the background noise condition. Visual examination of the figure shows that there are differences between the reference conditions and the three vocoders, and that the two enhanced vocoders are better than the original full-rate vocoder.

Table 3 contains the results for all 40 conditions included in this experiment. MOS values vary from

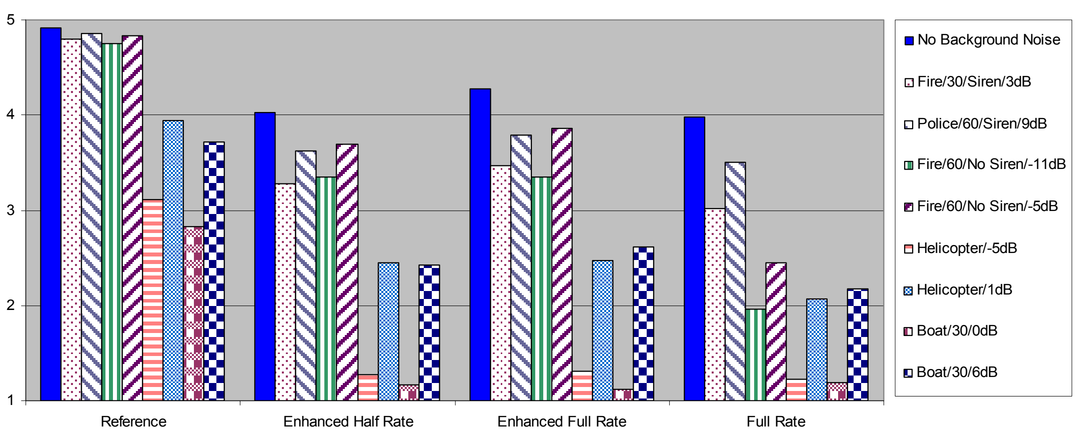


Figure 5. Selected Mean Opinion Scores.

Table 3. Mean Opinion Scores and Standard Deviations for All Conditions

Vocoder	Background Noise	MOS	Sdev
Linear PCM		4.913	0.082
25 dBQ MNRU		4.221	0.177
15 dBQ MNRU		3.279	0.162
5 dBQ MNRU		2.236	0.139
-5 dBQ MNRU		1.096	0.092
Enhanced Half Rate	Fire 30 Siren (3 dB)	3.284	0.221
Enhanced Full Rate	Fire 30 Siren (3 dB)	3.471	0.302
Full Rate	Fire 30 Siren (3 dB)	3.014	0.203
Reference	Fire 30 Siren (3 dB)	4.798	0.084
Enhanced Half Rate	Police 60 Siren (9 dB)	3.620	0.255
Enhanced Full Rate	Police 60 Siren (9 dB)	3.788	0.220
Full Rate	Police 60 Siren (9 dB)	3.500	0.127
Reference	Police 60 Siren (9 dB)	4.856	0.094
Enhanced Half Rate	Fire 60 No Siren (-11 dB)	3.346	0.276
Enhanced Full Rate	Fire 60 No Siren (-11 dB)	3.351	0.254
Full Rate	Fire 60 No Siren (-11 dB)	1.957	0.357
Reference	Fire 60 No Siren (-11 dB)	4.750	0.074
Enhanced Half Rate	Fire 60 No Siren (-5 dB)	3.697	0.291
Enhanced Full Rate	Fire 60 No Siren (-5 dB)	3.865	0.240
Full Rate	Fire 60 No Siren (-5 dB)	2.452	0.426
Reference	Fire 60 No Siren (-5 dB)	4.832	0.041
Enhanced Half Rate	Helicopter (-5 dB)	1.274	0.222
Enhanced Full Rate	Helicopter (-5 dB)	1.303	0.242
Full Rate	Helicopter (-5 dB)	1.231	0.190
Reference	Helicopter (-5 dB)	3.115	0.340
Enhanced Half Rate	Helicopter (1 dB)	2.452	0.256
Enhanced Full Rate	Helicopter (1 dB)	2.466	0.375
Full Rate	Helicopter (1 dB)	2.067	0.376
Reference	Helicopter (1 dB)	3.938	0.225
Enhanced Half Rate	Boat 30 knots (0 dB)	1.163	0.096
Enhanced Full Rate	Boat 30 knots (0 dB)	1.115	0.139
Full Rate	Boat 30 knots (0 dB)	1.188	0.108
Reference	Boat 30 knots (0 dB)	2.822	0.288
Enhanced Half Rate	Boat 30 knots (6 dB)	2.428	0.329
Enhanced Full Rate	Boat 30 knots (6 dB)	2.615	0.307
Full Rate	Boat 30 knots (6 dB)	2.178	0.354
Reference	Boat 30 knots (6 dB)	3.716	0.291
Enhanced Half Rate	No Background Noise	4.029	0.367
Enhanced Full Rate	No Background Noise	4.274	0.183
Full Rate	No Background Noise	3.976	0.279

low of 1.096 to a high of 4.913 (5.0 is the maximum possible score). Standard deviations vary from 0.082, which is typical for the extreme ends of the quality range (i.e., 1 and 5) to 0.376, which is slightly larger than desired for a subjective experiment. The latter might have been mitigated through the use of a larger subject pool.

5.1 Observations

Because the listeners were public safety practitioners, they were somewhat different from typical "naïve" listeners. Most listeners will take a break when it is offered – in this case after each of the four sessions. However, many of the practitioners wanted to complete the whole experiment without taking the offered break between sessions. Some of the responses included "I have to listen to this type of material for 12 hours a day, this is nothing," and "This is easier than work because there's only one person talking at a time."

6. Summary of Results

Table 4 summarizes the statistically significant differences among the vocoders and between the vocoders and the reference condition by showing which vocoder was better if there was a difference or "Equiv" if they were equivalent. It also indicates whether any of the vocoders produced a speech signal that would be at all useable, where useable is defined as a MOS greater than 2.0.

Significant observations in the table are:

- There were two conditions (helicopter @ -5 dB SNR and boat at 0 dB SNR) for which none of the vocoders produced a speech signal of acceptable quality. These conditions are the only ones where all three vocoders were statistically equivalent.
- For those two conditions mentioned in the previous bullet, an increase of 6 dB SNR (as might be achieved from a noise-canceling microphone) made a significant difference in the quality of speech produced by the two enhanced vocoders (at least 1 full MOS point increase).

Table 4. Summary of Statistically Significant Differences Among Vocoders Tested.

	Statistically Significant Difference From			Any vocoder
	FR to EFR	FR to EHR	EHR to EFR	>2.0 MOS?
No Background Noise	EFR	Equiv	EFR	Yes
Fire 30 Siren (3 dB)	EFR	EHR	Equiv	Yes
Police 60 Siren (9 dB)	EFR	Equiv	Equiv	Yes
Fire 60 No Siren (-11 dB)	EFR	EHR	Equiv	Yes
Fire 60 No Siren (-5 dB)	EFR	EHR	Equiv	Yes
Helicopter (-5 dB)	Equiv	Equiv	Equiv	No
Helicopter (1 dB)	EFR	EHR	Equiv	Yes
Boat 30 knots (0 dB)	Equiv	Equiv	Equiv	No
Boat 30 knots (6 dB)	EFR	EHR	Equiv	Yes

- For all conditions except those where no vocoder had a MOS > 2.0, the Enhanced Full Rate vocoder was significantly better than the Full Rate vocoder.
- For five of the seven conditions for which one or more vocoders produced acceptable quality, the Enhanced Half Rate vocoder was significantly better than the Full Rate vocoder.
- The Enhanced Full Rate vocoder was significantly better than the Enhanced Half Rate vocoder for only one condition: the reference condition (no background noise).

7. References

[1] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies."

[2] ITU-T Recommendation P.51, "Artificial Mouth."

[3] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality."

[4] ITU-T Recommendation P.830, "Subjective performance assessment of telephone band and wide band digital codecs."

[5] ITU-T Recommendation P.56, "Objective measurement of active speech level."

[6] ITU-T Recommendation G.191, "Software tools for speech and audio coding standardization."

[7] ANSI/TIA/EIA 102.BABA, APCO Project 25 Vocoder Description, May 1998.

[8] ITU-T Recommendation P.810, "Modulated noise reference unit (MNRU)."

[9] ANSI S1.4, "Specification for Sound Level Meters," 2001.

[10] ANSI S1.42, "Design Response of Weighting Networks for Acoustical Measurements," 2001.

Radio Communications for Emergency Responders in Large Public Buildings: Comparing Analog and Digital Modulation*

Kate A. Remley [1], Marc Rütschlin, Dylan F. Williams,
Robert T. Johnk, Galen Koepke, Christopher L. Holloway
National Institute of Standards and Technology; 325 Broadway; Boulder, CO 80305
1. Corresponding author: ph. (303) 497 3652, email: remley@boulder.nist.gov

Mike Worrell, Phoenix Fire Department; Phoenix, AZ
Andy MacFarlane, Medaire, Inc.; Phoenix, AZ

To assess in-building radio coverage, in 2004 the City of Phoenix Fire Department carried out extensive testing of their radio systems. They deployed firefighters in standard configurations in a variety of buildings, and rated on a scale of 1 to 5 the audio quality of the received signals. To provide a link between the qualitative ratings and absolute field strength, NIST staff later carried out a series of measurements side-by-side with the Phoenix firefighters. The calibrated data from the NIST tests enables translation of the larger set of Phoenix Fire Department data into transferable values useful to industry, standards organizations, and other public-safety groups. We report here on a subset of these tests that compare analog and digitally modulated signals at 800 MHz.

1. Introduction

Reliable communication between emergency responders operating in hazardous situations is critical to both the safety of personnel and the success of their mission. Their radio communications equipment must be extremely reliable, and the communications functions provided must be predictable. To assess their in-building radio communications under field conditions, the Phoenix Fire Department (PFD) conducted extensive testing of new and existing radio systems deployed in configurations common to the department [1].

Testing focused on the ability of firefighters deployed in a building interior to be able to communicate with a command position on the exterior of the building. An analysis of fire-ground communications was performed for each building studied. All responses were based on the Phoenix Fire Department Standard Operating Procedures (SOPs) for the various National Fire Protection Association (NFPA) building types (see Appendix).

During the testing process, the National Institute of Standards and Technology (NIST) became aware of the Phoenix Fire Department testing. NIST developed measurement techniques to determine the absolute received signal strength at locations near PFD personnel during a set of tests. A team from NIST spent a week measuring radio signal levels. The NIST measurements were later calibrated in a post-processing step to find the electric field strength. Comparing these field strengths to the qualitative ratings provided by Phoenix Fire personnel allowed us to "translate" the larger set of Phoenix Fire Department tests into approximate quantitative values that are transferable to other locations and enables technical evaluation or comparison of different radio systems.

Among the tests carried out were those using point-to-point 700/800 MHz analog simplex and 700/800 MHz digital simplex. These tests provided an opportunity to compare radio reception when both digital and analog modulation were used. We report on these tests below.

2. The Phoenix Fire Department Tests

The Phoenix Fire Department tests were performed over an eight-week period. The tests were conducted in 30 buildings that consisted of the five different NFPA construction types (see Appendix). Approximately 1,500 talk paths were tested. The same test participants were used throughout the testing to provide test consistency in the grading process.

A command structure was developed for each response. Analyzing the command structure and determining the logical fire ground communications paths led to the development of the talk matrix. A test script was developed for each unique communications path identified in the talk matrix. Each communications path was categorized as either fire ground communications or wide-area communications, with the majority of the paths being fire ground. A typical layout is shown in Fig. 1.

Each building was pre-planned for the test session. Personnel were placed in the buildings to represent fire companies on an incident response. The personnel

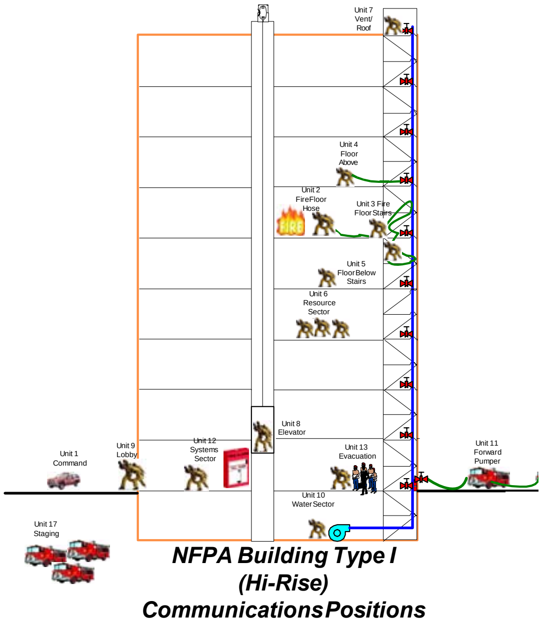


Figure 1: Typical firefighter deployment in a high-rise building. Radio links between all deployed positions were tested by PFD.

followed the test plan and graded each communication path. Each path was graded on a 1-5 scale with 1 representing poor communications and 5 being the best audio quality. Participants also determined whether the communications were useable on the fire ground on a pass/fail basis.

Upon completion of testing each day, the building's test results were entered into a spreadsheet containing the bidirectional grades for each communications path. Using the spreadsheet data, histograms were created for each building type and path category. NFPA Type 1 (concrete and steel) buildings were found to have the most variability in the histograms, and are also the buildings with the highest-risk environments, most complex interiors and largest operational structure. The histograms in the following are for fire-ground operations in Type 1 buildings.

Test results showed that the 700/800 MHz analog simplex channels provided clear, consistent communications in most test situations. This is shown in Fig. 2 by the high number of "5" ratings in the histogram. The test participants preferred this mode of operation over all others. Before testing commenced we expected that there would be differences in penetration capabilities between the 450 MHz and 700/800 MHz RF frequency bands. However, the penetration differences between the RF bands were negligible for the buildings we tested.

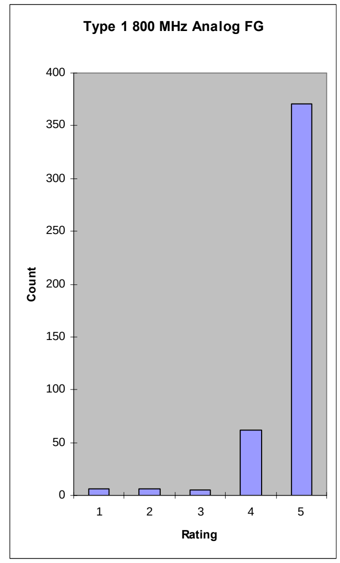


Figure 2: PFD test results: Analog simplex signals received in the local fireground (FG) over the course of all tests. A rating of 5 represents the clearest audio transmission.

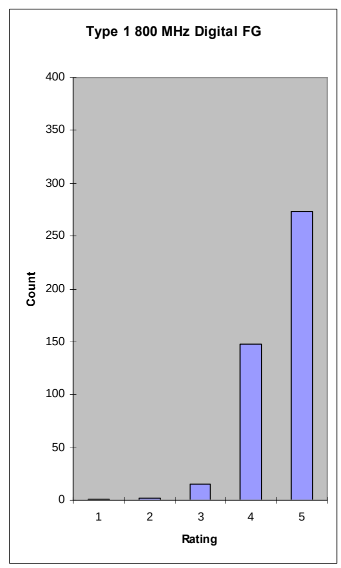


Figure 3: PFD Test Results: Digital simplex signals received in the local fireground (FG) over the course of all tests. A rating of 5 represents the clearest audio transmission.

The 700/800 MHz digital simplex channels (see Fig. 3) provided consistent communications. Test participants did note that the majority of transmissions had some level of digital distortion as reflected by the large number of "4" ratings and an increase in "3" ratings. The typical level of distortion did not render the communications unusable. However, users did encounter more situations where a repeat broadcast was needed to interpret the message.

Digital modulation did outperform analog modulation in many low signal-level situations. The digital mode would

Figure 4: NIST calibrated measurements were taken in close proximity to Phoenix Fire personnel using the system shown on the left side of the photo.

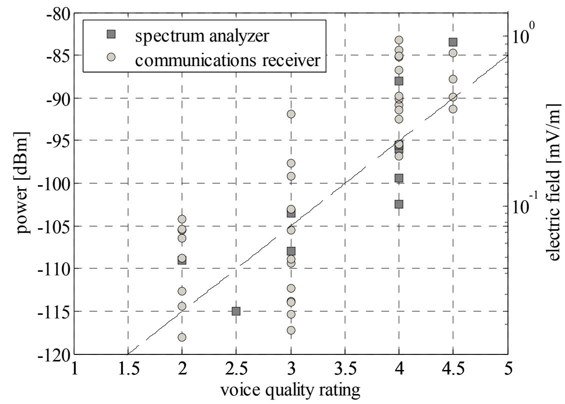


Figure 5: Measured average power and field strengths corresponding to voice quality evaluations for 860 MHz analog transmissions. The dashed line is a linear fit of all the data.

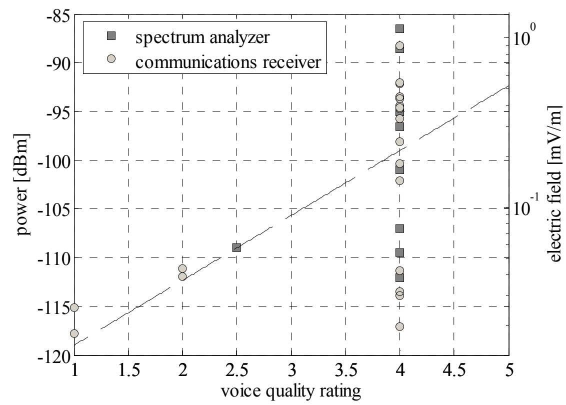


Figure 6: Measured average power and field strengths corresponding to voice quality evaluations for 860 MHz digital transmissions. The dashed line is a linear fit of all the data.

provide understandable communication when the same analog path would be scratchy and barely readable.

3. The NIST Tests

The goal of the NIST collaboration with the Phoenix Fire Department was to assign measured power and absolute electric field strength levels to the various qualitative ratings. While rating schemes such as the one described above are common, they suffer limitations inherent to "subjective" scales: they do not enable technical evaluation or comparison of different radio systems and it is difficult to compare ratings carried out by other groups in similar experiments.

The NIST measurement system consisted of a communications receiver, antenna system, and laptop computer. This system was developed to detect very weak signals in a cost effective way. The calibrations involved in these tests are described in [2,3]. The receiver system was used to collect data on the ground floor of an eight-story commercial building in Phoenix. The transmitter – a portable radio identical to those used by the Fire Department – was carried up and down a staircase on the opposite side of the building.

To ensure that the NIST experiments covered the entire range of voice quality ratings described in the Phoenix tests, we carried the transmitters to places with good and bad radio reception. At various locations in the building, an audio and silent test count were performed to allow nearly simultaneous PFD evaluation of the voice quality and NIST measurement of the power in the modulated carrier. The quality of the audio transmission was evaluated by a firefighter located at the listening station; both test counts were recorded using the communications receiver system for later calibration and evaluation. The silent test counts were also monitored with a spectrum analyzer. This allowed us to have a backup set of measurements should one system fail, and also provided an independent verification of the communications receiver measurements. Among other tests carried out, analog and digital transmissions at 860 MHz were investigated and are the ones reported on here.

Scatter plots of the measured field strengths corresponding to voice quality evaluations are shown in Figures 5 and 6 for the 860 MHz analog and 860 MHz digital transmission respectively. The linear fit to the data in each graph may be used to estimate the field strength required for a particular quality of transmission. For example, in Fig. 5 all of the analog transmissions rated 3 correspond to received power levels below -90 dBm.

Unlike the analog case, the digital signal scatter plot in Fig. 6 shows that the majority of transmissions were given a voice quality level of 4, even for quite weak measured signal levels. This is the "digital shelf" effect noted by the PFD earlier. Note that all the quality ratings of 3 or worse were for signals with power levels of lower than -105 dBm.

The number of ratings of lower audio quality in the NIST tests is greater than that for the Phoenix Fire Department tests shown in Figs. 2 and 3. The NIST-led experiment was designed to intentionally cover the entire range of audio quality levels, whereas the Phoenix Fire tests reflected more realistic communications scenarios.

Factors such as multipath, the approximate one-meter difference in position between firefighters and the receiver system, repeatability of transmitter position, and the subjective nature of the audio quality assessments likely contribute to the spread of data for a given audio quality rating in Figs. 5 and 6. The graphs shown here are thus an estimate of the range of signal strengths that would result in a certain voice quality being received in the same room.

4. Summary

Tests that compared audio quality of received signals from analog and digitally modulated radio systems were part of an extensive study carried out by the Phoenix Fire Department. The purpose of this study was to assess effectiveness and deployment of various firefighter radio systems in the field. Results showed that both analog- and digitally-modulated systems work in a majority of situations in an NFPA Type 1 building (concrete and steel). However, digitally modulated signals were discernable at lower signal levels than those for their analog counterparts. When both types of signals could be received, the firefighters preferred the audio quality of the analog signals.

To investigate the link between the subjective ratings used in the Phoenix Fire Department tests and received signal strength levels, NIST carried out measurements of absolute field strength in an eight-story building in Phoenix in which poor signal transmission quality had been observed in previous tests. Audio quality ratings were provided by members of the Phoenix Fire Department who had participated in the earlier tests. Plotting the received signal levels vs. audio quality ratings provided a link between quantitative and the qualitative results. This link allows for comparison of the Phoenix Fire Department test data with data collected by other groups. Furthermore, having absolute signal strengths enables the data to be used in the technical evaluation of different radio systems, which may be of use to industry, standards organizations, and other public-safety groups.

Appendix—NFPA Building Types

Type 1: Fire-Resistive Construction
Reinforced concrete and structural steel
Type 2: Non-Combustible/Limited Combustible Construction.
a) Metal-Frame covered by metal exterior walls
b) Metal frame enclosed by concrete block, non-bearing exterior walls
c) Concrete block bearing walls supporting a metal roof
Type 3: Ordinary Construction/Brick and Joint Construction
Type 4: Heavy Timber Construction
Type 5: Wood Frame Construction

Acknowledgement

This work was funded in part by the Department of Justice Community Oriented Police Services (COPS) and the Department of Homeland Security through the NIST Office of Law Enforcement Standards.

References

[1] Mike Worrell and Andy MacFarlane, "Phoenix Fire Department Radio System Safety Project," *Final Report*, Oct. 2004. Report available at http://www.ci.phoenix.az.us/FIRE/radioreport.pdf Copyrighted material used by permission.

[2] M. Rütschlin, K. A. Remley, R. T. Johnk, D. F. Williams, G. Koepke, C. Holloway, A. MacFarlane, and M. Worrell, "Measurement of weak signals using a communications receiver system," *Proc. Intl. Symp. Advanced Radio Tech.*, Boulder, CO, March 2005, pp. 199-204.

[3] M. Rütschlin, K. A. Remley, D. F. Williams, R. T. Johnk, G. Koepke, D. Camell, M. Worrell, and A. MacFarlane, "Field-strength measurement of weak-modulated signals in buildings," *in process*.

RFID-Assisted Indoor Localization and Communication for First Responders

L. E. Miller, P. F. Wilson, N. P. Bryner, M. H. Francis, J. R. Guerrieri, D. W. Stroup, L. Klein-Berndt
National Institute of Standards and Technology (NIST)
Contact: L. E. Miller, 301-975-8018, 301-590-0932 (fax), lmiller@antd.nist.gov

An indoor localization and communication project is described that proposes to use RFID tags, placed in the building beforehand, as navigation waypoints for an inertial navigation system carried by a first responder. The findings from the first year of the planned three-year project are summarized.

1. Introduction

RFID (radio-frequency identification) devices commonly are attached to persons or to moveable objects so that the objects can be tracked using fixed readers (special-purpose radios) at different locations. In this project, supported by the NIST Advanced Technology Program (ATP), we are exploring a novel application of the "flip side" of this practice based on the concept that detection of an RFID device in a known, fixed location by a moving reader provides a precise indication of location for tracking the person or moving object that is carrying the reader. The research aims to evaluate the exploitation of this concept to implement a low-cost, reliable means for tracking firefighters and other first responders inside buildings, where navigation using GPS is not reliable—indeed, the GPS signal may have been disabled temporarily to prevent exploitation by terrorists [1]. The research will also consider the use of building-related information stored in RFID devices placed at fixed on-site locations to aid the responders in their mission, as well as to describe the room layout or other context of the device, thereby minimizing the need for accessing remote databases through the communication system.

Previous research and development for indoor localization includes that of a wireless network that integrates communications, precise tracking, and data telemetry, based on ultrawideband (UWB) technology, for use in hospital and manufacturing environments [2]. In contrast, the system envisioned by this new project is intended for an environment that is potentially much less "friendly" to RF propagation—the in-building environment of first responders that may contain smoke, dust, or flames—and is intended to leverage advances in ubiquitous RFID tag technology, in combination with recent advances in miniaturized inertial sensors, to develop a low-cost tracking system that does not depend upon the stability of the RF environment over relatively large distances to derive range from precision timing. The "philosophy" of the proposed RFID-assisted system also involves reducing the dependence on RF links to external data sources by exploiting the capability of RFID tags to store critical building information for retrieval when it is needed, where it is needed.

This project is a joint effort by components of three NIST laboratories: the Wireless Communication Technologies Group of the Information Technology Laboratory (ITL) and the Fire Fighting Technology Group of the Building and Fire Research Laboratory (BFRL) in Gaithersburg, Maryland; and the Radio-Frequency Fields Group of the Electronics and Electrical Engineering Laboratory in Boulder, Colorado.

1.1 Concepts Motivating the Study

Indoor Navigation Cannot Depend on GPS. It cannot be assumed that GPS position solutions will be available to first responders in an indoor mission-critical situation. Even if the GPS signals are not blocked or obscured for tactical advantage, the reception of GPS signals inside most buildings is not reliable.

Inertial Sensors Can Track Location, Motion. In addition to, or in place of, GPS, the position of a first responder inside a building can be tracked using inertial sensors such as accelerometers and gyroscopes. Non-inertial sensors such as magnetometers and barometers can also be used in conjunction with dead reckoning to develop positions of a first responder in motion.

RFID Fixes Can Enhance the Accuracy of Inertial Tracking Systems. Inertial tracking systems inherently drift over time and produce errors in position, especially for inexpensive and lightweight systems. Corrections to the position solution at points along the path of the first responder can limit the maximum error to an acceptable level. Corrections, in the form of the insertion of known locations that have been reached, can be developed automatically by the detection of an RFID device, either by correlating the identity of the device with a table of locations or by reading the device's location from data stored on it.

1.2 Approach to the Study

At the outset of the study, the overall approach was described as follows: In addition to the RF propagation environment of buildings in emergency situations, the research will consider several operational scenarios consisting of (1) the strategy for RFID deployment, (2) the tracking method, and (3) the options for presenting location information to the user and communicating this information to a monitoring station. The RFID deployment and tracking aspects of the scenarios to be studied will include:

The tradeoffs involved in the choice of RFID devices for this application, including cost, ease of programming, suitability for emergency environments, and data capacity.

Use of relatively few RFID location reference points to correct or calibrate an inertial navigation or other localization system to maintain sufficient accuracy during a first responder incident.

Use of multiple RFID location reference points to furnish data for tracking without the use of inertial sensors.

The emphasis will be to make maximum use of information and to leverage software to simplify hardware implementations. The presentation and communication aspects of the scenarios to be studied will include:

Informing the user (only) of position (stand-alone mode), assuming any communication is provided by a separate system.

Informing the user, other team members, and an incident commander of their positions via an ad hoc network of radio terminals that combine RFID reading and radio communication.

Providing the user with directions for safe exiting of the building.

1.3 Project Milestones and Plans

The project goals are rather ambitious. But the milestones do serve to show how the work is intended to sequence and how the three groups collaborate.

1.3.1 FY 2005 Milestones

A. Define critical parameters of firefighter localization and in-building informational requirements in typical scenarios that relate to the building RF propagation environment and to the number and placement of RFID tags in buildings, as well as the type of data to be stored on the tags. (BFRL, EEEL)
B. Evaluate inertial and dead-reckoning navigation techniques and device options (including MEMS-based sensors) regarding their accuracy, availability, and suitability for integration with an RFID reader on a small platform for location updating. (ITL)
C. Analyze the requirements for the number of RFID tags and their placement to achieve desired localization accuracies, as a function of navigation techniques and device options. (BFRL, ITL)
D. Evaluate options for RFID technologies—including both tags and readers—to use for location updating of a navigation system implemented on a small, battery-powered device similar to a handheld computer or PDA (personal digital assistant). (EEEL, ITL)
E. Establish a project web page. (ITL)[1]
F. Document/publish interim results. (ITL, EEEL, BFRL)

1.3.2 FY 2006 Milestones

A. Select RFID tag and reader technologies and develop a prototype reader for use in this application. (EEEL)
B. Develop embedded software for acquisition of data from the RFID reader and use of that data to perform location updates and to display the location on a handheld computer as well as building information derived from RFID tag data. (ITL, EEEL)
C. Conduct preliminary experiments in NIST's Large Fire Facility to evaluate the performance of RFID-assisted localization devices in structures of simple geometry. (BFRL)
D. Evaluate options for interfacing the localization device with an ad hoc wireless communication network (ITL, BFRL)
E. Document/publish interim results. (ITL, EEEL, BFRL)

1.3.3 FY 2007 Milestones

A. Develop embedded software for interfacing an RFID-assisted localization device with an ad hoc wireless communication network. (ITL, BFRL)
B. Integrate RFID reader, navigation hardware and software, and ad hoc communication system for prototype testing. (EEEL, ITL)
C. Identify test sites/buildings and conduct tests to demonstrate the operation of the prototype localization and communication system in burning or smoke-filled building environments. (BFRL, ITL)
D. Develop embedded software for directing the user to the nearest RFID-tagged exit. (ITL)
E. Document/publish final results. (ITL, EEEL, BFRL)

[1] The project web page is found at http://www.antd.nist.gov/wctg/RFID/RFIDassist.htm

2. Summary of FY 2005 Accomplishments

2.1 Firefighter Localization Parameters

Firefighter location and in-building information requirements have been developed, as representing perhaps the most demanding environment for an RFID-assisted localization and communication system. The implications for placement of tags have been identified as the survival of the functioning of the tag in the environment.[2] The number of RFID tags is still an open question, depending on a tradeoff between practicality and the desired localization resolution—in many first responder scenarios, it would be a much-needed advancement to be able to identify which room in a building a given first responder is located. This question will continue to be considered in later phases of the project.

Firefighter location and in-building information requirements may be grouped in terms of (a) building type, (b) temperature environment, (c) radio attenuation factors, and (d) desired location resolution.

Building type refers both to the building's construction, which relates to its resistance to fire, and to factors affecting communications in the building. The building type, along with classification of the building use (e.g., residential or industrial), is the primary parameter in the description of the fire event scenario. Building and fire codes classify buildings according to the type of construction and the fire resistance of the various load-bearing and non load-bearing elements, such as exterior and interior walls, columns, beams and girders, and floor construction. There are five types of building construction identified in the various codes [3, 4]:

Type I Buildings are classified as Fire-Resistive. Most of these buildings are used as high-rise office buildings, shopping centers, or residential units. They will be either reinforced concrete or structural steel. Structural members will be approved noncombustible or limited-combustible materials with specified fire resistance ratings. Any steel construction members must be protected to withstand prescribed test temperatures for fire resistance.

Type II Buildings are classified as Noncombustible. These buildings can be used for example as office buildings, warehouses, or automobile repair shops. There are three basic types of non-combustible buildings: metal frame covered by metal exterior walls (Butler Buildings), metal frame enclosed by masonry as non-bearing exterior walls, and masonry bearing walls supporting a metal roof. These metal roofs can be either solid steel girders and beams, or lightweight open-web bar joists or a combination of both with corrugated metal sheathing. The structural members are noncombustible or limited-combustible materials.

Type III Buildings are classified as Ordinary. Type III buildings are often called ordinary buildings or combustible/noncombustible. The majority of buildings probably fall into the Type III category. These buildings can be office buildings, retail stores, mixed occupancy, dwellings, or apartment buildings. These buildings usually have non-combustible bearing walls and combustible roofs. Usually, the exterior walls are concrete, concrete block, or brick. Interior, non-load bearing walls can be made of wood.

Type IV Buildings are classified as Heavy Timber. The exterior and interior walls and structural members that are portions of walls must be of approved non-combustible or limited combustible materials. Interior structural members, including walls, columns, floors, and roofs, are large dimension solid or laminated wood timbers. The exterior walls are typically masonry. These buildings exist primarily in the New England area.

Type V Buildings are classified as Wood Frame. Wood frame buildings generally are constructed in one of five methods: log, post and beam, balloon, platform, and plank and beam. A wood frame building can be used for many different purposes, such as single-family dwellings, multiple-family dwellings, restaurants, or retail stores. The major structural members are typically composed of wood and the exterior walls are combustible.

The communications factors associated with the building type include whether the building has a pre-wired communication system and whether the construction of the building is such that radio signals may not penetrate the building adequately (*e.g.*, steel or metal).

The temperature environment of a building during a first responder event is described in terms of zones that correspond to degrees of exposure to heat flux and therefore to risk of injury [5]. Tables 1–4 give examples of fires in the different temperature and heat flux zones (Zone I to Zone IV).

Radio attenuation factors are those affecting the transmission of radio signals into and out of a burning building. These factors include

Presence of water in the air, due either to combustion products or the fire suppression water. 100% relative humidity can be expected.

Smoke particulates in the air, usually carbon. Typically one gram per cubic meter.

Charged particles in the air that are ions from combustion processes.

[2] High temperature testing of the functionality of typical RFID tag systems will be conducted in FY 2006, using FY 2005 DHS funds that have been made available in part on an interest in this ATP project.

Table 1. Examples of Zone I environments.

Group, Year	Designation	Description	Ranges
USFA, FEMA; 1992 [6]	None		
Abeles Project Fires, 1980 [7]	Class 1	Overhaul, up to 30 minutes	Temperature to 40 °C Flux to 0.5 kW/m^2
IAFF; Based on Abeles Project Fires 1980	Class1	Overhaul, up to 30 minutes	Temperature to 40 °C Flux to 0.5 kW/m^2
FRDG, FEU 1995 [8]	Routine	Elevated temperature, no direct thermal radiation, 25 minutes	Temperature to 100 °C Flux to 1.0 kW/m^2
Coletta, 1976 [5]	None		
Abbott, 1976 [9]	None		

Table 2. Examples of Zone II environments.

Group, Year	Designation	Description	Ranges
USFA, FEMA 1992 [6]	Routine	One or two objects burning	Temperature 20 C to 60 °C Flux 1.0 to 2.1 kW/m^2
Abeles Project Fires, 1980 [7]	Class 2	Small fire in a room, 15 minutes	Temperature 40 °C to 95 °C Flux 0.5 to 1.0 kW/m^2
IAFF; Based on Abeles Project Fires 1980	Class 2	Small fire in a room, 15 minutes	Temperature 40 °C to 93 °C Flux 0.5 to 1.0 kW/m^2
FRDG, FEU 1995 [8]	Hazardous	Elevated temperature and direct thermal radiation, 10 minutes	Temp. 100 °C to160 °C Flux 1.0 to 4.0 kW/m^2
Coletta 1976 [5]	Routine	Fighting fires from a distance	Temperature to 60 °C Flux 0.4 to 1.25 kW/m^2
Abbott 1976 [9]	Routine	Fighting fires from a distance	Temperature to 70 °C Flux 0.5 to 1.7 kW/m^2

Table 3. Examples of Zone III environments.

Group, Year	Designation	Description	Ranges
USFA, FEMA 1992 [6]	Ordinary	Serious fire, next to a room in flashover; 10 to 20 minutes maximum	Temp. 60 °C to 300 °C Flux 2.1 to 25 kW/m^2
Abeles Project Fires, 1980 [7]	Class 3	Totally involved fire, 5 minutes	Temp. 95 °C to 250 °C Flux 1.0 to 1.75 kW/m^2
IAFF; Based on Abeles Project Fires, 1980	Class 3	Totally involved fire, 5 minutes	Temp. 93 °C to 260 °C Flux 1.0 to 1.75 kW/m^2
FRDG, FEU 1995 [8]	Extreme	Rescue, retreat from flashover or backdraft	Temp. 160 °C to 235 °C Flux 4.0 to 10.0 kW/m^2
Coletta 1976 [5]	Hazardous	Outside burning room or small building	Temp. 60 °C to 300 °C Flux 1.25 to 8.3 kW/m^2
Abbott 1976 [9]	Ordinary	Outside burning room or small building	Temp. 70 °C to 300 °C Flux 1.7 to 12.5 kW/m^2

Table 2.4 Examples of Zone IV environments.

Group, Year	Designation	Description	Ranges
USFA, FEMA 1992 [6]	Emergency	Severe and unusual, 15 to 30 seconds for escape	Temp. 300 °C to 1000 °C Flux 25 to 125 kW/m^2
Abeles Project Fires, 1980 [7]	Class 4	Flashover or backdraft, up to 10 seconds	Temp. 250 °C to 815 °C Flux 1.75 to 42 kW/m^2
IAFF; Based on Abeles Project Fires, 1980	Class 4	Flashover or backdraft, up to 10 seconds	Temp. 260 °C to 815 °C Flux 1.75 to 42 kW/m^2
FRDG, FEU 1995 [8]	Critical	Could be encountered briefly	Temp. 235 °C to 1000 °C Flux 10 to 100 kW/m^2
Coletta 1976 [5]	Emergency	Not normally encountered, may be during flashover	Temp. 300 °C to 1000 °C Flux 8.3 to 105 kW/m^2
Abbott 1976 [9]	Emergency	Not normally encountered, may be during flashover	Temp. 300 °C to 1100 °C Flux 12.5 to 208 kW/m^2

Thermal layers that could reflect or refract radio waves.
Construction materials and their various attenuation properties. Adverse effects can be from the metal facing on insulation, metal rebar in concrete, metal siding, and solar window treatments.

The desired resolution of indoor location information during a first responder event varies according to the scenario, that is, whether the building is residential or industrial. Tables 5 and 6 correlate the location resolution in meters to the accuracy in locating personnel and escape openings.

Table 5. Location parameters for residential scenarios.

Resolution in meters	Location		Escape	
	X-Y Direction	Z Direction	X-Y Direction	Z Direction
100	City Block +/-	10 floors +/-		
10	Front or rear of house	3 floors +/-	Structure +/- (Townhouse)	Floor +/-
1	Room	Floor +/-	Correct Wall	Window or Door
0.1	Location in Room	Correct Floor	Location on wall	Height of window or door

Table 6. Location parameters for industrial scenarios.

Resolution in meters	Location		Escape	
	X-Y Direction	Z Direction	X-Y Direction	Z Direction
100	Building +/-	10 floors +/-		
10	Section of Bldg	3 floors +/-	Section of Bldg	Floor +/-
1	Room	Floor +/-	Correct Wall	Window or Door
0.1	Location in Room	Correct Floor	Location on wall	Height of window or door

2.2. Navigation Techniques and Devices

The surveys and evaluations of navigation techniques and inertial navigation sensor technologies have been completed, with the major findings summarized below. On the basis of the surveys, a particular dead-reckoning module (DRM) has been selected for testing. A report containing survey details and tutorial materials on navigation has been drafted; its completion awaits testing of the selected DRM so that the test results may be included.

The most widely used navigation system today is the Global Positioning System (GPS), which enables position determination through the measurement of time delays of signals from multiple satellites in known (moving) positions; the time delay measurements are based on cross-correlating received satellite signals with local replicas to identify the signals' digital code position in time relative to the common reference. The difficulty in using GPS indoors and in urban "canyons" is that the line of sight to the GPS satellites is obscured or severely attenuated. Without four good satellite signals, the GPS position solution is inaccurate. Also, with weak signals, the GPS receiver continually loses lock and must spend an inordinate amount of time in attempting to acquire the signals.

Prior to the establishment of GPS, of course, many techniques and devices for navigation have been used. Today's navigation devices implement some very old navigation techniques, such as dead reckoning and waypoint navigation. Dead Reckoning (DR) is the process of estimating position by advancing a known position using course, speed, time and distance to be traveled—in other words, figuring out where one will be at a certain time if he holds the planned speed, time and course [10]. The usefulness of the technique depends upon how accurately speed and course can be maintained on a given "tack;" in the air and on the sea, the selection of fixed speed and course for relatively long periods of time are feasible, while on land or inside buildings the duration of the tack may need to be relatively short due to maneuvers that are required by the terrain or building layout. As illustrated in Figure 1, the uncertainty of the DR position grows with time, so that it is necessary to check the position regularly with a "fix" of some kind (perhaps an RFID tag as envisioned by this project).

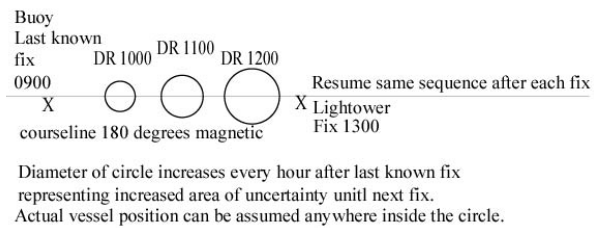


Figure 1 Dead reckoning in open ocean (from [11]).

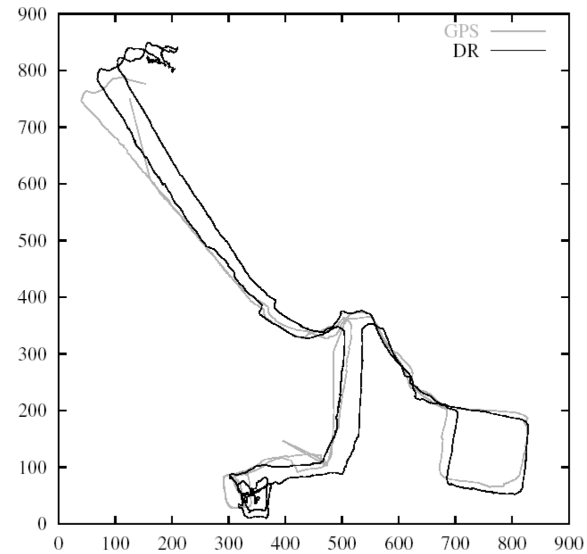


Figure 2 Positions developed by a DR system vs. GPS over a 4 km urban trail (from [12]).

Various systems are being proposed for "pedestrian navigation" utilizing DR techniques and small compasses. For example, [12] describes a small DR unit that utilizes a two-axis compass, a three-axis compass, or a rate gyroscope to track a walking person's heading. The heading produced by the 3-axis sensor is least affected by deviations in the person's orientation—very important in the case of first responders, who often do not walk "normally" in the course of their work. For computation of speed (actually, displacement as a function of time), the system in [12] uses an accelerometer to detect the person's steps; on average, the distance covered by a step was found to be quite consistent, even though the detection process can be rather subtle.

The outdoor performance of the system in [12] with a 2-axis magnetic compass and step detection using an accelerometer in an urban "tourist area" test is shown in Figure 2 in comparison to GPS. The scales in the figure are in meters. Although the DR positions were generally in agreement with those developed by GPS, they were often significantly different, due in part to unknown magnetic effects along the path, probably from the presence of an underground electric utility substation along the path. The standard deviation of the position area for the test was about 20 m. The authors conclude, "This case illustrates the susceptibility of magnetic compasses to localized magnetic fields, particularly in an urban environment. In these circumstances a gyroscope solution is clearly advisable."

Another personal navigation product based on dead reckoning and step counting is a DRM that integrates a GPS receiver with a magnetic compass and other sensors [13–16] and is described as follows [14]:

> The Dead Reckoning Module (DRM®) is a miniature, self-contained, electronic navigation unit that provides the user's position relative to an initialization point. The DRM® is the first commercially available practical implementation of a drift-free dead reckoning navigation system for use by personnel on foot. It is specifically designed to supplement GPS receivers during signal outages. You still know where personnel are located even when GPS is blocked by nearby buildings, heavy foliage, or even inside many structures. The DRM contains a tilt-compensated magnetic compass, electronic pedometer and barometric altimeter to provide a continuous deduced position. A microprocessor performs dead reckoning calculations and includes a Kalman filter to combine the dead reckoning data with GPS data when it is available. The filter and other proprietary algorithms use GPS data to calibrate dead reckoning sensors for a typical dead reckoning accuracy of 2% to 5% of distance traveled, entirely without GPS. Options for the system integrator include a selection of voltage input ranges, CMOS or RS232 interface, data logging, and special software functions. In addition to horizontal position data, compass azimuth, tilt (pitch and roll), and barometric altitude are available.

For improved stability and accuracy, a version of the DRM can be obtained that includes a gyroscope. Figure 3 from [13] shows an example of the improvement in DRM performance using a gyroscope.

The relatively small size and cost of these personal navigation devices is made possible by the development in recent years of very small inertial and non-inertial sensors. For example, Micro-machined (MEMS) rate gyros based on vibration are available; a diagram of a semiconductor rate gyro based on a MEMS tuning fork is shown in Figure 4. The principle of a vibrating gyro is very simple: a vibrating object (such as a tuning fork) tends to keep vibrating in the same plane as its support is rotated. It is therefore much simpler and cheaper than is a conventional rotating gyroscope of similar accuracy. In the engineering literature, this type of device is also known as a *Coriolis vibratory gyro* because as the plane of oscillation is rotated, the response detected by the transducer (usually a piezo

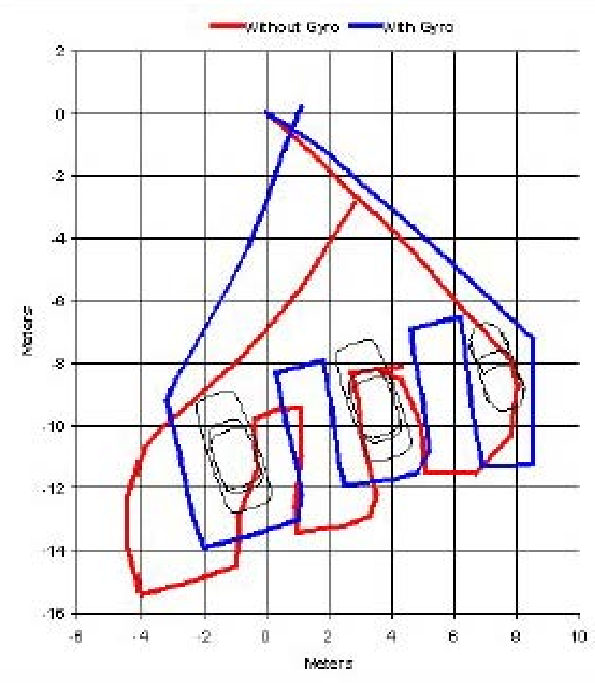


Figure 3. Manufacturer's demo of gyro-stabilized dead reckoning module improvement (from [14]).

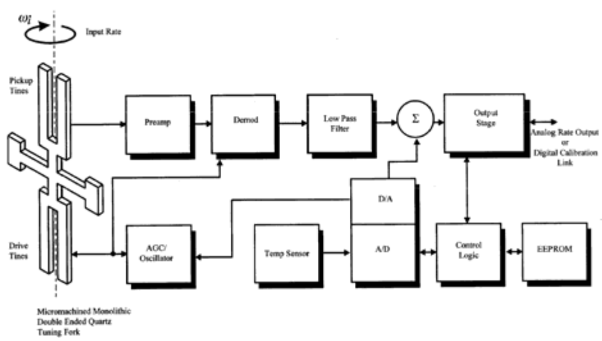


Figure 4. Block diagram of a semiconductor gyro based on MEMS technology (from [17]).

electric device) results from the coriolis term in its equations of motion. [16]

There is some question as to whether the step-counting algorithms in current personal navigation devices are sufficiently sophisticated to adapt to the irregular stepping patterns of firefighters while doing their work. Even if they are not very accurate, it is possible that they are good enough to preserve a useful track if they are periodically updated by accurate position information (see below). For the purposes of this study, ITL has placed an order for the DRM described above so that it can be tested under various conditions relevant to first responder scenarios, and eventually integrated with an RFID reader for obtaining position fixes from RFID tags.

2.3. Number and Placement of Tags

This milestone is predicated on the results of milestones A and B in that it assumes respectively that a required localization accuracy or range of accuracies has been stated and that the accuracies of various navigation techniques have been formulated in terms of the spatial density of waypoints and/or frequency of navigation fixes.

Although general information on the accuracy of potential navigation techniques for use indoors was developed, it was realized during the project that it was premature to analyze the final accuracy of an RFID-assisted inertial system at this time.

One factor involved is that the accuracy of the dead reckoning module to be tested is unknown under the conditions to be expected in a first responder scenario, particularly the effect of irregular walking on the step-counting algorithm.

Another factor involved is that there is no device available to first responders at this time that provides indoor location with any reliable degree of accuracy—the firefighting community would consider the ability just to identify the floor of the building on which a firefighter is located would be a step forward. Therefore, it is not appropriate to focus on analysis of the potential accuracy of a sophisticated system yet.

However, the project team did agree that it would be useful to evaluate the requirements for placement of RFID tags in order to indicate location from the tags alone. Using just RFID tags to derive the indoor location of a first responder has certain advantages that are known in advance. The "you are here" event of detecting a particular RFID tag in a building provides a positive indication of not only which floor of the building the first responder is on, but also which room or work area he or she is nearest to, assuming the ability to correlate the data on the RFID tag with building information or the existence of this information on the tag itself. That is, this result can be expected if there is no ambiguity in the case of the detection of more than one RFID tag.

In the industrial and supply chain applications of RFID technology, the expectation is that many tags will be within the range of the RFID reader, so that the standards for RFID devices provide protocols and procedures for resolving the signals from multiple tags. However, no attempt is made to determine the locations of the multiple tags, other than that they are all within the range of the reader.

Since the "read range" of RFID tags is very dependent on the specific RFID technology that is being used, some RFID tags and readers were procured and testing was performed to determine the read range and related parameters.

2.4 Options for RFID Technologies

The RFID technologies referred to in this milestone include not only the currently available devices, standards, and frequencies, but also any technique by which a person in a building can automatically detect that he is in a known location.

Discovery of a technique for deducing the orientation of a person at the time of detecting a device in a known location would be significant information for the study. For system studies, the evaluation of RFID technologies should include availability, cost, portability, and power consumption in addition to the functional parameters of read range and data capability.

NIST-Boulder has developed an RFID test bed at 13.56 MHz that complies with the ISO 14443 and 10373-6 standards. We have evaluated the 13.56 MHz passive tags. The tags and readers communicate via magnetic coupling which limits the read range to less than 10 cm for typical readers (simple loop antenna). More complex loop systems can be used to extend this range, as shown in Figure 5, but the practicality of such systems for first responder navigation remains under investigation. The key challenge is that because of the large wavelength of approximately 22 m, these systems operate in the near field with the magnetic field strength dropping as one over the distance cubed.

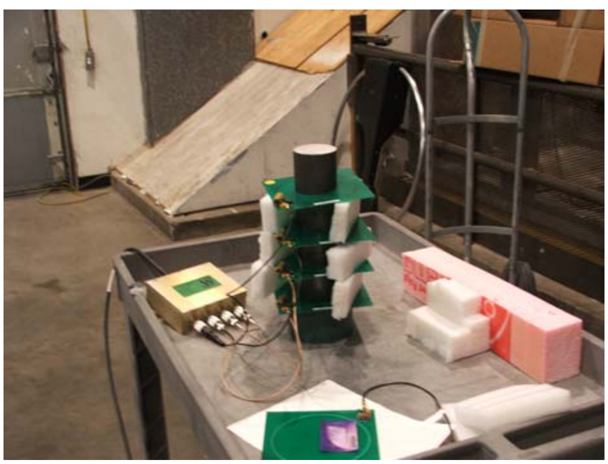


Figure 5. Loop antennas used to activate and read 13.56 MHz tags.

For higher frequencies, such as 400 MHz and 900 MHz, wavelengths are about 0.3 – 0.8 m. These systems operate in the far field. Coupling is via the electric field and the field drops more slowly (as one over the distance). This results in an anticipated read range of several meters for passive tags. It has been noted [19] that 600 MHz – 2 GHz is the best frequency band for propagation in buildings.

We have also started to evaluate an active tag RFID system that operates at 433 MHz, shown in Figure 6. This system requires an additional computer and power supply. The tags are battery powered. The tag and reader communicate using an electric field and the read range is over 30 meters. The reader can be set to recognize the tag responding with most power. This should be the tag closest to the first responder unless the battery power is low.

We have ordered 900 MHz RFID systems with software development kits. Anticipated delivery is mid-November 2005. These are passive tag systems. One uses a hand held reader that weighs only 1 kg and has a read range of approximately 3 meters (see Figure 7).

There is on-going work in the industry to optimize the performance and reduce the costs of RFID antennas and readers [20–24].

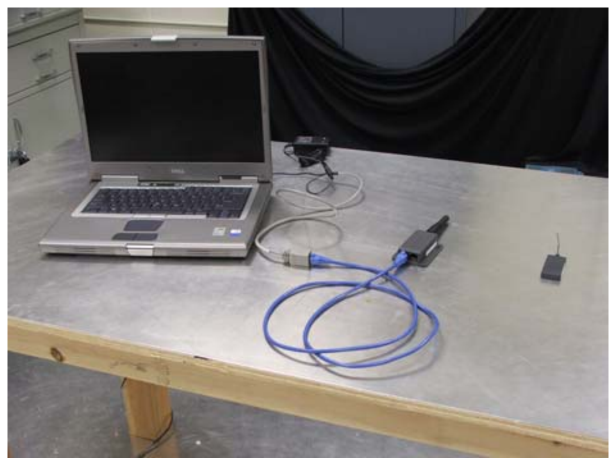


Figure 6. RFID system using active tags at 433 MHz.

Figure 7. Handheld 900 MHz reader.

3. References

[1] "Bush prepares for possible GPS shutdown," http://abcnews.go.com/Politics/wireStory?id=335054&CMP=OTC-RSSFeeds0312

[2] "Time Domain and GE Corporate R&D Announce Technology Partnership to Develop Ultra Wideband -UWB- Wireless Indoor Network," http://www.bbwexchange.com/news/archive/archived_broadband_wireless_files_2001/timedomain042401.htm

[3] NFPA A 220 Standard on Types of Building Construction, National Fire Protection Assoc., Quincy, MA 02269-9101, 2006.

[4] NFPA 5000 Building Construction and Safety Code, National Fire Protection Assoc., Quincy, MA 02269-9101, 2006.

[5] C. Coletta, I. J. Arons, L. E. Ashley, "Development of Criteria for Firefighters Gloves," NIOSH-77-134-A; GIDEP-E072-2273; Natl. Inst. for Occupational Safety and Health, Morgantown, WV, February 1976.

[6] ——, "Minimum Standards on Structural Fire Fighting Protective Clothing and Equipment: A Guide for Fire Service Education and Procurement," U.S. Fire Admin., Emmitsburg, MD, 1992.

[7] F. J. Abeles, "Project Fires, Volume 2: Protective Ensemble Performance Standards, Phase 1B," NASA CR-161530, National Aeronautics and Space Admin., Washington, DC, May 1980.

[8] J. A. Foster and G. V. Roberts, "Measurements of the Firefighting Environment – Summary Report," Central Fire Brigades Advisory Council Research Report number 61, 1994, Home Office Fire Research and Development Group, *Fire Engineers Journal*, United Kingdom, September 1995.

[9] N. J. Abbott, T. E. Lannefeld, and R. E. Erlandson, "Exploratory Development of Coated Fabric for Firefighters' Protective Clothing. Final Report. April 1974-March 1975," AFML-TR-75-72; Fabric Research Labs, Dedham, MA, July 1975.

[10] ——, http://www.deadreckoning.com/

[11] W. Newberry, "Dead Reckoning: A Skill All Navigators Should Master," Pilothouse Online at http://www.pilothouseonline.com/IS2V1_00/Lessons/main0001.htm

[12] C. Randell, et al., "Personal Position Measurement Using Dead Reckoning," *Proc. 2003 ISWC* at http://www1.cs.columbia.edu/~drexel/CandExam/Randell2003_ISWC_DeadReckoning.pdf

[13] C. T. Judd, "A Personal Dead Reckoning Module," *Proc. ION GPS '97*. Online at http://www1.cs.columbia.edu/~drexel/CandExam/DRM_ION97paper.pdf

[14] http://www.ssec.honeywell.com/magnetic/datasheets/gyrodrm.pdf

[15] ——, "DRM-III OEM Dead Reckoning Module for Personnel Positioning," Point Research Corp. brochure. Online at http://www.pointresearch.com/PDFs/DRM3.pdf

[16] ——, "Vibrating structure gyroscope," Wikipedia article online at http://www.answers.com/topic/vibrating-structure-gyroscope

[17] A. M. Madni and L. E. Costlow, "A third generation, highly monitored, micromachined quartz rate sensor for safety-critical vehicle stability control," *Proc. IEEE 2001 Aerospace Conference*, Vol. 5, pp. 2523–2534.

[18] ——, "Waypoint," Wikipedia online article at http://en.wikipedia.org/wiki/Waypoint

[19] A. E. Fathy, discussion during "Through-Wall Imaging and Sensing" July 7, 2005, session 99, 2005 IEEE Antennas and Propag. Soc. Internatl. Symp. and USNC/URSI Natl. Radio Science Mtg.

[20] S. M. Weigand, "Compact Microstrip Antenna with Forward-Directed Radiation Pattern for RFID Reader Card," July 6, 2005, session 64, 2005 IEEE Antennas and Propag. Soc. Internatl. Symp. and USNC/URSI Natl. Radio Science Mtg.

[21] L. Ukkonen, D. Engels, L. Sydänheimo, M. Kivikoski, "Reliability of Passive RFID of Multiple Objects Using Folded Microstrip Patch-Type Tag Antenna," July 6, 2005, session 64, 2005 IEEE Antennas and Propag. Soc. Internatl. Symp. and USNC/URSI Natl. Radio Science Mtg.

[22] D. Bechevet, T.P. Vuong, S. Tedjini, "Design and Measurements of Antennas for RFID, Made by Conductive Ink on Plastics," July 6, 2005, session 64, 2005 IEEE Antennas and Propag. Soc. Internatl. Symp. and USNC/URSI Natl. Radio Science Mtg.

[23] C. Cho, H. Choo, "Design of UHF Small Passive Tag Antennas," July 6, 2005, session 64, 2005 IEEE Antennas and Propag. Soc. Internatl. Symp. and USNC/URSI Natl. Radio Science Mtg.

[24] P,V, Nikitin, K.V.S. Rao, S. Lam, "Low Cost Silver Ink RFID Tag Antennas," July 6, 2005, session 64, 2005 IEEE Antennas and Propag. Soc. Internatl. Symp. and USNC/URSI Natl. Radio Science Mtg.

Extensible Software for Automated Testing of Public Safety P25 Land Mobile Radios

Julie Kub, Electronics Engineer and Eric Nelson, Electronics Engineer
Institute for Telecommunication Sciences (ITS), National
Telecommunications and Information Administration (NTIA)
325 Broadway, Boulder, CO 80305
Julie Kub: email: jkub@its.bldrdoc.gov, phone: 303.497.4607
Eric Nelson: email: enelson@its.bldrdoc.gov, phone: 303.497.4445

Abstract: One of the most prominent digital Land Mobile Radio (LMR) technologies used by public safety agencies, Project 25, is built upon an expansive suite of standards defining numerous open interfaces. As the P25 standard has matured and greater numbers of subscriber units and fixed station equipment have reached the market, increasing complaints of non-interoperability and substandard performance have arisen. In response, P25 users and manufacturers are forming a P25 Compliance Assessment Program. One element of this program requires electrical performance measurements on P25 portable and mobile radios. These tests will be conducted using an automated test software application developed at the Institute for Telecommunication Sciences. This paper describes background information on the compliance program, discusses the technical approaches taken in application development, and provides a detailed overview of the test application's functionality and capabilities.

1. Introduction

Whether responding to cataclysmic events (such as natural disasters and terrorist attacks) or performing day to day operations, it is critical that public safety personnel be able to communicate. Recognizing that cellular and land-line communication systems are often overloaded and unusable when they are needed most, public safety officials have continued to demand their own dedicated, reliable LMR systems. For a variety of reasons, public safety agencies have been slow to adopt digital modulations; most still use analog frequency modulation. However, the Federal Communications Commission's (FCC's) mandate of a transition to narrowband systems to address spectrum congestion has spurred the deployment of spectrally efficient digital modulations. In 1989 the Association of Public Safety Communications Officials (APCO) International established Project 25 (P25) to work with the Telecommunications Industry Association (TIA) to facilitate the development of standards-based narrowband digital LMR systems with open interfaces. With the maturation of the Project 25 suite of standards, the number of P25 radios on the market has recently increased dramatically. Unfortunately, operational field locations and a host of test laboratories have reported numerous interoperability and performance deficiencies. Following on the heels of September 11th, these continued reports of deficiencies and non-interoperability spurred the United States Congress to direct a formal compliance assessment program be established for Project 25 equipment procured by the Federal government and recipients of Federal grants. The following two statements are excerpted from Congressional reports:

<u>Department of Homeland Security Congressional Report</u>
The conferees direct the Office of Interoperability and Compatibility (OIC) to work with the National Institute of Standards and Technology [NIST] and the U.S. Department of Justice [DoJ] to require, when Project 25 equipment is purchased with such funds, the equipment meets the requirements of a conformity assessment program. The conferees further direct such a conformity assessment program be funded by this appropriation and be available by the end of fiscal year 2006 [1].

<u>Department of Commerce Appropriations for FY06</u>
The Committee also directs that, within this report, OLES [Office of Law Enforcement Standards] identify a process to ensure that equipment procured using Federal grant dollars complies with the requirements of the identified standard(s). At a minimum, the Office of Interoperability and Compatibility [OIC] within the Department of Homeland Security should consider working with NIST and DOJ to require that all grant dollars for interoperable communication be used for Project 25 compliant equipment that meet the requirements of a conformity assessment program [2].

Accordingly, NIST's Office of Law Enforcement Standards (OLES) working on behalf of the Department of Homeland Security's SAFECOM program in conjunction with the Institute for Telecommunication Sciences (ITS) has spearheaded the creation of a Project 25 Compliance Assessment Program. Program managers in consultation with members of the APCO 25 Interface Committee (which represents both P25 user and manufacturer interests) have identified three essential elements of a compliance assessment program, namely, tests for conformance, interoperability, and performance. Efforts to implement performance tests have proceeded quite rapidly due to the availability of published test procedures. ITS engineers have been engaged for a number of months in a process to create the P25 Radio Performance Measurement (RPM) software suite of performance tests. This software represents a significant technical contribution to P25 radio testing. This paper focuses on the efforts to automate the performance testing element of the compliance assessment program. Interoperability and conformance testing are not discussed in this paper, because they require significant manual operations which do not lend themselves to automation.

Once software automation of the performance tests is complete, third-party independent laboratories will perform the actual tests. Many qualified laboratories such as those designated as FCC's Telecommunication Certification Bodies already have significant experience testing land mobile radios for compliance with FCC Parts 22 and 90. With the RPM software, laboratories will be able to rapidly expand their existing accreditation to include digital radio tests via NIST's National Voluntary Laboratory Accreditation Program (NVLAP). ITS will act as a technical bridge between TIA, NIST NVLAP, and participating laboratories.

2. Automated Software for Performance Testing

Performance tests are natural candidates for automation. First, uniform test protocols implemented through automated test procedures can ensure consistent methodologies amongst the various participating laboratories. Also, automation encapsulates the highly sophisticated testing techniques and equipment setup requirements of P25 digital radio testing and thus reduces the level of experience required of engineers and technicians at the independent laboratories. Finally, automated testing can either drastically reduce testing time through the use of efficient testing algorithms or enable exhaustive testing which would otherwise be prohibitive. For instance, bench tests conducted at ITS indicated that an exhaustive execution of the Spurious Response Rejection procedure required bit error rate measurements at approximately 100,000 interference frequencies. Manually conducting such an exhaustive test would be inconceivable but is readily accomplished via unattended test procedures. To promote the use of uniform test methods NIST/OLES directed ITS to develop automated test software for participating laboratories. The software suite will be provided through traditional software distribution channels. As a result, P25 radio manufacturers need not develop their own internal performance testing applications, thus promoting industry and competition.

3. Identifying LMR TIA-102 Digital P25 Measurements

Project 25 radio performance tests and accompanying requirements are documented in TIA-102.CAAA[3] and TIA-102.CAAB[4], respectively. The requirements consist of 18 receiver and 20 transmitter measurements. From these, a total of 10 receiver and 8 transmitter measurements were selected, based on their practical significance and utility, for inclusion in the performance testing component of the P25 Compliance Assessment Program. The proposed tests were reviewed by experienced LMR test engineers and accepted by the APCO 25 Interface Committee. Several test procedures were omitted since the characteristics they measured are readily evaluated by observation (e.g., audio quality and audio levels tests). Other procedures can be tested using existing FCC measurement procedures on the analog section of the radios. The 18 measurements selected for the performance test suite are being combined into one testing application, the P25 Radio Performance Measurements (RPM) tool. The tests that compose the P25 RPM tool are listed in Table 1.

Table 1. Digital P25 Automated Measurements

Receiver Measurements		Transmitter Measurements	
Section	**Procedure**	**Section**	**Procedure**
2.1.4	Reference Sensitivity	2.2.5	Modulation Emission Spectrum
2.1.5	Faded Reference Sensitivity	2.2.8	Adjacent Channel Power Ratio
2.1.6	Signal Delay Spread Capability	2.2.9	Intermodulation Attenuation
2.1.7	Adjacent Channel Rejection	2.2.12	Transmitter Power and Encoder Attack Time
2.1.8	Co-Channel Rejection	2.2.14	Transmitter Throughput Delay
2.1.9	Spurious Response Rejection	2.2.15	Frequency Deviation for C4FM
2.1.10	Intermodulation Rejection	2.2.16	Modulation Fidelity
2.1.11	Signal Displacement Bandwidth	2.2.18	Transient Frequency Behavior
2.1.17	Late Entry Unsquelch Delay		
2.1.18	Receiver Throughput Delay		

4. General RPM Hardware Requirements

The receiver measurements contained in the RPM require one or more radio frequency (RF) signal generators which emit either desired or interference type signals. These signals are combined and fed into the LMR under test by direct conduction into its antenna terminal. Most of the receiver measurements (except 2.1.17 and 2.1.18) require continued observation of a radio's bit error rate (BER) which is usually driven to a target value of 5% under a variety of simulated conditions. Significant BER averaging is required to achieve consistent results while the input signal is gradually adjusted until the target BER is obtained. A typical measurement takes 5-10 minutes while the longest receiver measurement must run unattended for several hours.

Only a handful of RF signal generators on the market are capable of generating the Compatible 4 Level Frequency Modulation (C4FM) required to perform tests on P25 receivers. Fewer still meet the phase noise requirements of TIA-102.CAAA. Two test procedures (2.1.5 and 2.1.6) also require a fading channel simulator. ITS acquired two different models of signal generator which satisfied the above requirements.

As an example of a receiver measurement, Figure 1 shows a block diagram of TIA-102.CAAA, Section 2.1.10, the Intermodulation Rejection measurement.

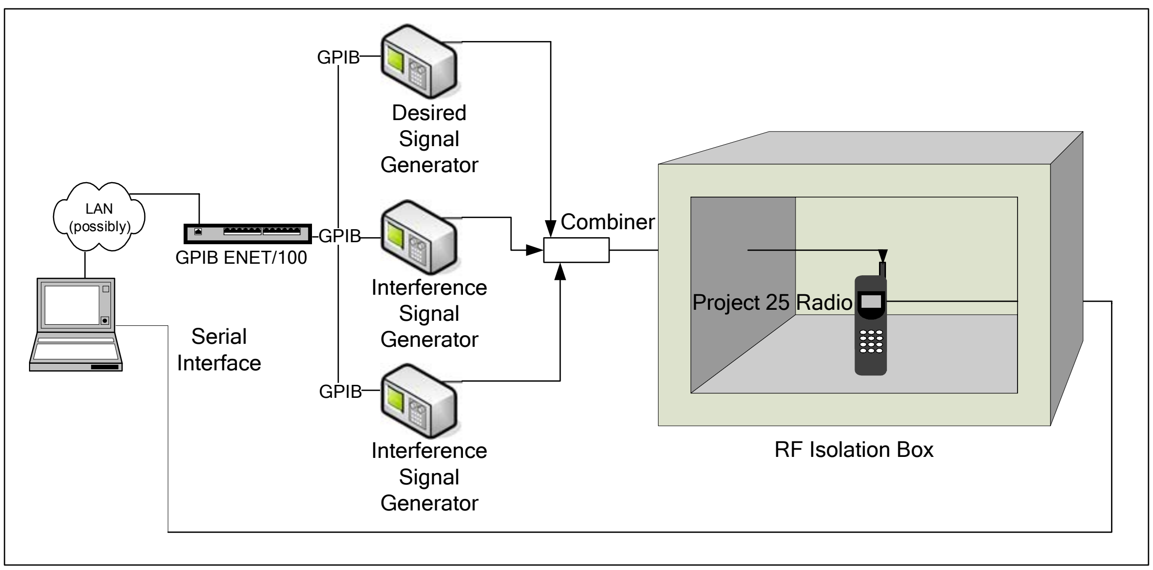


Figure 1. Example Receiver Measurement Block Diagram (2.1.10, Intermodulation Rejection)

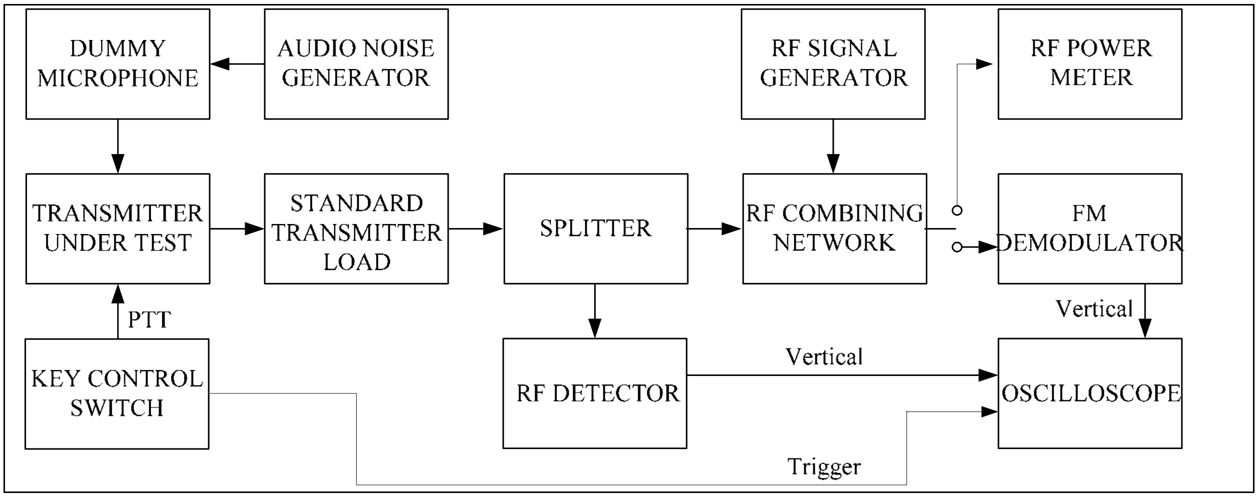


Figure 2. Transmitter Measurement Equipment without Using RTSA[3]

The automated transmitter measurements, on the other hand, are generally much less sophisticated and faster to execute (typically less than five minutes per measurement). All of these measurements can be performed using a single instrument—a real time spectrum analyzer (RTSA) with multi-domain functions. In some instances this instrument effectively replaces several other instruments as shown in Figure 2 (equipment other than the RTSA) and Figure 3 (RTSA only). Most of the transmitter tests are accomplished by configuring the device under test to transmit a particular bit sequence and asserting push to talk for a brief period of time. The RTSA captured waveform is then examined for compliance with the standards.

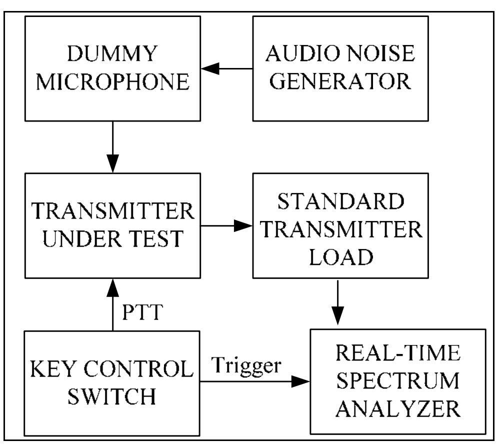


Figure 3. Transmitter Measurement Equipment Using Only RTSA

Both digital and receiver measurements are controlled via a computer running the RPM software. The test equipment, i.e., RF signal generators and spectrum analyzer, is controlled using a General Purpose Interface Bus (GPIB) to Ethernet controller. The computer's serial port is used to monitor BER data (for receiver tests) and for controlling the radios (transmitter tests). Per the recommended practice, the LMR under test will be housed in a table top RF shielded test fixture while the entire suite of test equipment will be contained in an RF shielded enclosure. TIA-102.CAAA contains detailed specifications for all eighteen transmitter and receiver measurements.

5. General RPM Software Requirements

Since the test procedures, especially the receiver tests, follow a consistent pattern, the opportunity to re-use code was apparent. Object-oriented code was used extensively throughout the RPM software for clarity, code re-use, and modularity. Every single measurement module uses a standard template containing state machines, objects, code templates and event-driven graphical user interfaces.

The RPM employs extensive data-driven design by using three databases: one for instrument commands and information, one for measurement information, and one for storing the results of all 18 measurements. The software development effort follows standard software design practices including the creation of requirements, design, and test plan documentation, and the utilization of a bug tracking database. The RPM's core measurement structure is easily extensible. Additional measurements are easily constructed from existing code with minor modifications to the measurement database and the core engine.

6. Typical Program Flow

Figure 4 shows the RPM automation state flow diagram. Of particular importance is the generic nature of the RPM software; in fact, the RPM flow chart could describe any measurement automation system that requires GPIB test equipment control and measurement automation. The flowchart shows common features of software initialization, selecting measurements to test, resetting and ensuring that all test equipment is responding, selecting the inputs for all measurements, running a batch of measurements unattended, and then performing clean-up functions. Batch measurement processing is made possible by up-front measurement procedure selection allowing several procedures to be run without further user input – a critical time-saving design feature.

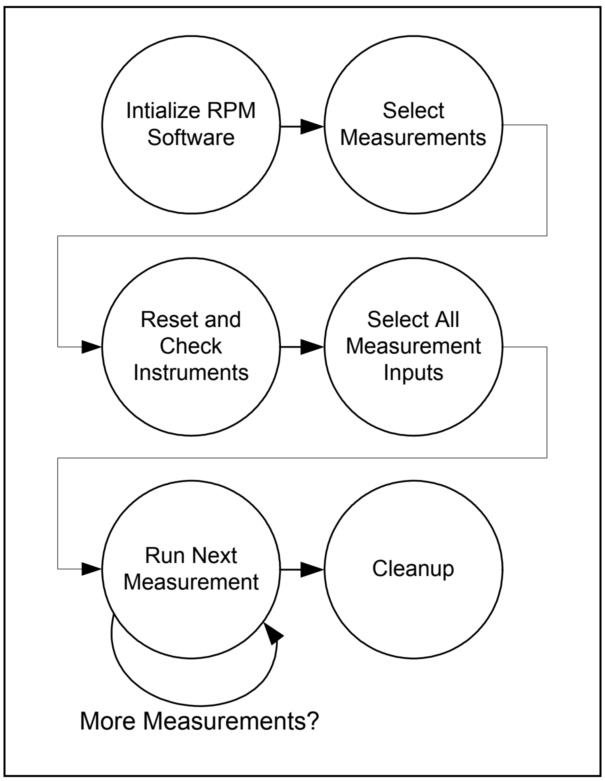


Figure 4. RPM Software Automation Flow Chart

7. The Finished Product

The RPM system contains a number of Graphical User Interfaces (GUI) which display measurement parameters or intermediate measurement data represented in graphical and/or tabular forms. A typical user interface is illustrated in Figure 5 which shows the "Intermodulation Rejection" measurement. All measurement GUI's also contain status indicators at the bottom of the display and command buttons in the lower right which enable the user to manually pause and cancel the measurement. Care was taken to ensure that "cancel" requests from any GUI are handled promptly.

8. Extensibility

The RPM system is easily extended to incorporate new test equipment and measurement modules. For instance, a digital storage oscilloscope or communications analyzer could easily be added to the program. Hypothetically, any measurement which employs GPIB controlled test equipment can be accommodated by inserting test modules into the core code.

9. Conclusion

RPM software development is nearing completion and deployment is slated to begin by the end of 2006. The software will be made available by NIST. It will be used by accredited radio test laboratories as a part of the Project 25 Compliance Assessment Program to test P25 LMRs. Use of the RPM software will minimize testing costs for manufacturers. The resultant performance test reports will be contained in a central repository which will include interoperability and conformance tests as well. Data within this repository will assist public safety procurement officials in making informed purchasing decisions. Designed using industry best practices, the RPM software is extensible and could potentially be applied to a wide array of automated testing applications well beyond its initial design scope.

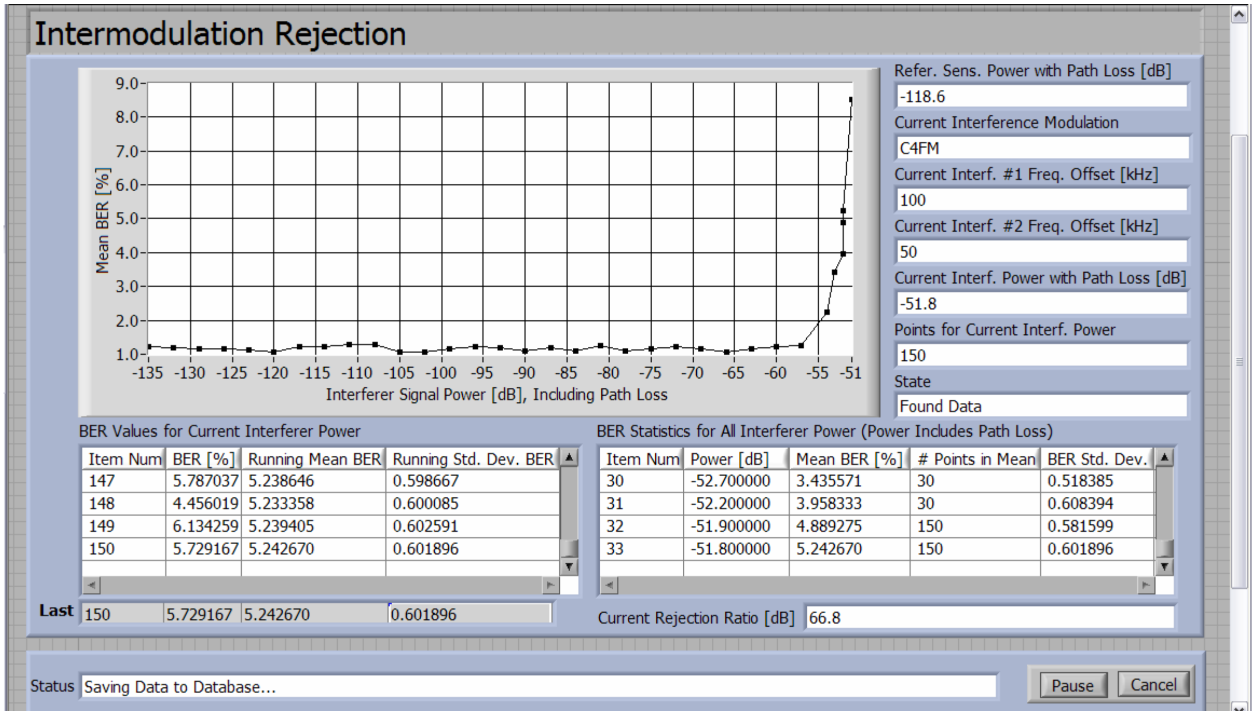


Figure 5. Digital Receiver Intermodulation Rejection Measurement GUI

10. References

[1] 109 Congress Report, House of Representatives, 23-669, 1st session, 109-241, "Making Appropriations for the Department of Homeland Security for the Fiscal Year Ending September 30, 2006, and for Other Purposes", submitted by Mr. Rogers of Kentucky, "Interoperability and Compatibility" section, September, 2005.

[2] Senate Report 109-088, "Departments of Commerce and Justice, Science, and Related Agencies Appropriations Bill, 2006".

[3] TIA-102.CAAA-A, TIA Standard, Digital C4FM/CQPSK Transceiver Measurement Methods, Sections 2.1 and 2.2, November 2002.

[4] TIA-102.CAAB-A, TIA Standard, Land Mobile Radio Transceiver Performance Recommendations – Project 25 – Digital Radio Technology C4FM/CQPSK Modulation, Sections 3.1 and 3.2, September 2002.

Flexibility in Frequency Management: The New Frequency Bill in the Netherlands

Lilian Jeanty
Radio Communications Agency the Netherlands
Phone: +31 50 5877 163
Email: lilian.jeanty@at-ez.nl

Paper unavailable at time of printing.

Why Unlicensed Use of the White Space in the TV Bands Will Not Cause Interference to DTV Viewers

Michael J. Marcus
Marcus Spectrum Solutions
+1-301-229-7714
mjmarcus@alum.mit.edu

Paul Kolodzy
Kolodzy Consulting

Andrew Lippman
MIT Media Lab

Abstract: On May 13, 2004, the Federal Communications Commission approved a Notice of Proposed Rulemaking (NPRM) proposing to allow a new generation of wireless devices to utilize vacant television channel frequencies in each market. This so-called TV band "white space" consists of frequencies that are allocated for television broadcasting but are not actually in use in a given area. This paper explains how the concept proposed by FCC, with minor modifications, can be made to work without causing harmful interference to DTV reception

The FCC's proposed rules[1] in Docket 04-186 are intended to make way for unlicensed spectrum technologies, such as Wi-Fi, to utilize the prime TV band spectrum to offer wireless broadband services. Wi-Fi technology has become very popular at higher frequencies, and has had a positive impact on the growth of broadband services. However, the bands used for Wi-Fi do not have appropriate radio propagation characteristics to serve low population densities. Lower-frequency spectrum, such as that used for TV broadcasting, is capable of traveling longer distances at a given power level, and can better penetrate obstacles such as buildings and trees.

The FCC's proposal would promote both spectrum efficiency and wireless broadband deployment. The TV band has been called a "vast wasteland" of underutilized spectrum. Even after the completion of the DTV transition – and the reallocation of TV channels 52-to-69 – an average of only seven full-power DTV stations will be operating on channels 2-to-51 in the nation's 210 local TV markets. Only a fraction of the 294 MHz of prime spectrum allocated to DTV services will actually be utilized in most markets.

Thus, the proposed use of "white space" TV channels could have a particularly great impact on the growth of information services in rural areas, where such empty channels are readily available. In urban areas, where less "white space" is available, this spectrum would also be useful because of the great demand for wireless broadband services and because of the ability of the TV band spectrum to penetrate buildings and objects within buildings better than the higher bands.

The FCC was clear in this NPRM that any devices certified to operate in the TV white spaces would be required to use *new "smart radio" technology* that would not interfere with television reception. Nevertheless, the National Association of Broadcasters (NAB) and other broadcast industry representatives, in comments filed at the FCC and in communications with Congress, have objected to the FCC's proposal, claiming that unlicensed devices operating on vacant channels in the TV band would cause harmful interference to television broadcasts and other uses of licensed TV band channels.

This Issue Brief responds to the broadcast industry's allegations, addressing each of the industry's concerns about interference. The paper concludes that interference-free unlicensed use of the white space *is* practical with today's technology. While some of the issues raised here are novel, the FCC as an "expert agency" should be able to handle them as it handles other cutting-edge spectrum problems. Indeed, the FCC is required by statute to avoid harmful interference with licensed TV broadcasts – and its NPRM describes several different ways to protect the dwindling number of over-the-air TV viewers from interference, as described below.

Unlicensed Devices: 350 Million and Booming

Unlicensed devices have been authorized by the FCC since 1938. A Consumer Electronics Association study quoted by the FCC estimates that there are over 350 million unlicensed devices in the US and that annual hardware sales are in the multibillion dollar range.[2] The earliest unlicensed devices were remote controls for radio receivers. Today's unlicensed devices range from the ubiquitous cordless telephone to garage door openers to home security systems to Wi-Fi wireless local area networks.

All of these systems comply with general rules established by the FCC to ensure that they do not cause interference to licensed systems and Federal Government systems.[3] Some unlicensed devices operate at very low power so they can coexist with higher power licensed users in the same band,[4] while others (e.g., Wi-Fi) operate in bands that are largely devoid of licensed users.[5] Before a new model of unlicensed device can be sold, it must be authorized--that is, it must be tested by a third party and shown to comply with technical standards established in FCC Rules.[6] The FCC enforces its technical rules for unlicensed devices through both this equipment authorization program and through its statutory jurisdiction over the marketing of devices "capable of emitting radio frequency energy...in a sufficient degree to cause harmful interference to radio communications."[7]

The FCC's technical rules have been primarily focused on preventing interference with licensed users. Unlike some other governments, the FCC has not attempted to steer the market by mandating specific services or technologies. This light-handed regulation has enabled a dynamic market for unlicensed devices to develop, as innovators bring to market new devices for new applications. Perhaps the best known example of this dynamic innovation on unlicensed bands is the explosive growth of Wi-Fi technology. The Telecommunications Industry Association estimates that sales of Wi-Fi equipment in 2004 reached $4.35 billion, and predicts spending on Wi-Fi infrastructure equipment will increase to $7 billion in 2008, a 12.6 percent annual increase.[8] The development and popularization of Wi-Fi technology was built on a 1985 FCC decision[9] to allow unlicensed devices in three bands--then best known for being the "home" of microwave ovens--provided they used "spread spectrum" technology to minimize interference.

The FCC Proposal for Unlicensed Sharing of TV Spectrum Without Harmful Interference

The Commission's May 2004 NPRM proposed to allow unlicensed devices to operate on unused TV channels, often called "white spaces." As the FCC noted in its NPRM, this spectrum would be ideal for unlicensed broadband because it has better radio propagation characteristics than the present Wi-Fi bands and can tolerate higher power devices without causing interference. These characteristics allow wireless broadband providers to achieve better-quality coverage of larger areas using less infrastructure, significantly reducing the cost of broadband deployment. A recent study by Intel confirms this, showing that the capital costs of covering a rural area with wireless broadband service in the TV band would be one-fourth those needed to achieve the same coverage using licensed MMDS spectrum in the 2.5 GHz band (which sits adjacent to the current unlicensed "Wi-Fi band" at 2.4 GHz).[10]

The NPRM proposes unlicensed operation under one of three alternative schemes intended to prevent interference to television reception:

I. "*Listen-Before-Talk*" (*LBT*): Sensing the presence of TV signals by the unlicensed device in order to select channels not in use. This concept, also described as dynamic frequency selection (DFS), has already been adopted by the International Telecommunications Union (ITU) and the FCC for sharing of the 5 GHz spectrum between unlicensed systems and military radar.[11] Technical protocols to avoid interference are negotiated between industry and the military.

II. "*Geolocation/Database*": Location sensing and consultation with broadcast database. In this scheme, an unlicensed device would contain location-sensing technology, such as a Global Positioning System (GPS) receiver. The device would cross-check its own location with an internal database of TV transmitter locations in order to verify that it was a minimum distance from a TV transmitter.

III. "*Local Beacon*": Reception of a locally transmitted signal that identifies which TV channels may be used in the local area for unlicensed use. In this scheme, low power local signals, possibly controlled by local broadcasters, would indicate directly which channels were free for use.

The FCC NPRM proposes possible use of any of these methods as acceptable ways of avoiding interference to licensed broadcast users, and recognizes that the final rules might only allow for one or two of these independent alternatives. The remainder of this Issue Brief will discuss basic technical issues that have been raised in the FCC proceeding and then specific points made by the broadcast industry lobby in recent communications with Congress and the FCC.

I. Broadcaster Interference Concerns are Unfounded or Readily Avoidable with Established Technologies

This section will address basic technical issues associated with the three alternatives. The proponents of this NPRM, including academics and equipment manufacturers, have shown in their comments that any of the three alternatives may be both effective and practical. While the original FCC proposal might not have been flawless, the remaining issues can be resolved through the normal rulemaking process at the FCC. Indeed, this is why Congress adopted the Administrative Procedures Act[12] in order to have a give and take between regulators and concerned parties before rules are adopted reflecting the overall public interest.

A. Listen-Before-Talk (LBT) Alternative; Avoiding the "Hidden Node" Problem

The broadcast interests have focused much of their concern about the NPRM on alleged vulnerabilities in the *LBT* alternative (Alternative I above), in which unlicensed devices must first "listen" and sense the presence of TV signals in the area before transmitting. They point out that, as shown in Figure 1, an unlicensed device could be in the shadow of a building and be shielded from the TV signals, while a TV antenna at the top of the building might get a good signal.[13] This is known in the technical literature as the "hidden node" problem. Indeed, studies have shown that in both urban and rural areas, where buildings and terrain serve as obstacles to TV signal penetration, there exist many "shadow" spots in which TV signals may be weakened or totally diminished.

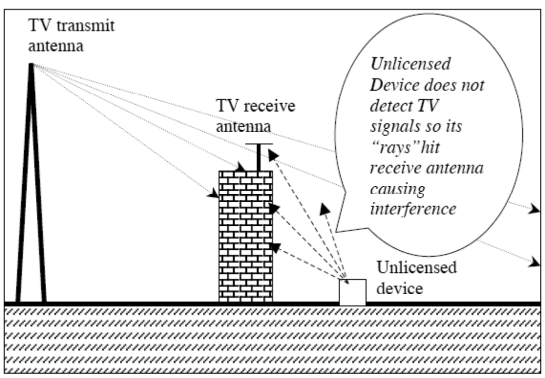


Figure 1 - The "Hidden Node" Problem

Therefore, the broadcast interests claim that unlicensed devices using this alternative are likely to miss detecting TV signals due to shadowing, and thus will cause interference to nearby TV receivers that have adequate signal strength.

The comments of the broadcast industry (and even the FCC's NPRM) assume that the detector part of the unlicensed devices in the *LBT* alternative would be about as sensitive to radio-frequency emissions as are normal TV receivers. But this need not be the case. Research presented at a February 2003 FCC-sponsored seminar demonstrated that a detector optimized for a specific class of signals (e.g., TV signals) can be orders of magnitude more sensitive than a normal receiver.[14] The Commission had previously taken note of this research in its NPRM on cognitive radio,[15] but inexplicably did not address it in this unlicensed NPRM. Similarly, the reply comments of the broadcast community have steadfastly ignored the applicability of this technology, which was mentioned repeatedly by various parties in the comment phase of the FCC rulemaking.[16]

It has also been pointed out in the comments that cooperative sensing of TV spectrum by multiple unlicensed devices could, in effect, improve sensitivity of TV signal detection significantly. Such cooperative sensing can be used in conjunction with very sensitive detectors for even more sensitivity gain.[17]

The use of very sensitive receivers could solve the hidden node problem. *The FCC could simply set a sensitivity value for detectors that would give a high confidence that usable TV signals would not be missed*, and then verify during the equipment authorization process for each model of unlicensed device whether that sensitivity level was met.

B. Geolocation/Database Alternative: Need to Keep FCC Data Up to Date

The broadcast interests also raise concerns about a second alternative means to avoid interference with TV reception on nearby channels: Geolocation and automated checking against a database of frequency assignments (Alternative II). Broadcasters have pointed out that geolocation systems such as GPS do not generally work indoors and hence could not reliably determine location. They also point out that the FCC databases on broadcast stations are not 100% accurate and are sometimes slow in catching up to transmitter frequency location changes – a more common problem now during the DTV transition.

We acknowledge the validity of these comments, but note that all of these concerns can be addressed with minor modifications to the proposed rules. The final rules should require that unlicensed devices must make iterative geolocation checks within a specified time interval in order to continue transmitting on a given frequency.

With respect to the broadcaster claims about the reliability of geolocation technologies, it is important to note that there are advanced GPS technologies used in some cellular telephone systems that actually *do* work indoors.[18] Furthermore, once the DTV transition is complete, it will become technologically feasible to conduct indoor geolocation using multiple DTV signals, instead of the satellite technology used in current GPS systems. Indeed, geolocation could even become a new product for broadcasters.

In regards to the accuracy of FCC transmitter databases,

outdated, allowing manual data entry problems to compromise the accuracy of transmitter location information. We call upon Congress and the FCC to recognize that such technology issues limit the potential of the multibillion dollar industries the FCC regulates and upgrade FCC databases so that they can be viewed as highly reliable. Regardless, if Congress mandates a "hard date" for the end of the DTV transition as it is expected to do, spectrum use will become more stable and the problems of updating the present FCC systems will become manageable.

C. Local Beacon Alternative: Control Signal Rules Can Avoid False Positives

With respect to the *Local Beacon* alternative (Alternative III above), the broadcast interests point out that the NPRM did not specifically propose what type of short-range radio signals should be used to broadcast channel availability information. Absent specific rules, a long-range transmitter might indicate availability of a certain channel and be received in an area far away, where that channel is not really available. For example, a signal transmitted in the AM broadcast band could have a range of hundreds of miles at night, and would be inappropriate for carrying information about which empty TV channels could be used in a given area. We agree with the broadcast interests on this point, but the problem could be simply resolved by rules specifying that the radio channel used to convey TV channel availability information must have a range comparable with the geographic validity of the channel availability information.[19]

D. Channel Availability

Some broadcast interests have questioned whether there will be significant channel availability for unlicensed use in major urban areas during the DTV transition. This concern is unwarranted. Even in urban areas, where there are fewer unused channels, there is likely to be substantial channel availability during the transition. Also, as Intel has argued, just because a particular channel may not be available throughout an urban region, doesn't mean it won't be available in *parts* of an urban area. Furthermore, the issue of channel availability during the DTV transition is likely to be short-lived. It now seems likely that the DTV transition will be ended by a date certain in the not too distant future – and the transition issue will simply go away.

Most importantly, there is no doubt that in rural areas—where unlicensed access to the TV band white space would make the most difference for affordable broadband deployment—there *is* spectrum available now and there will be for the foreseeable future. The proponents of this proposal do not seek a guarantee on how much spectrum will be available in a given location at a given time, and are willing to take their risks with the basic FCC proposal and their own analysis.

II. Other Concerns Expressed by Broadcast Community and Responses from NAF *et al.*

The broadcast industry has vehemently opposed the NPRM with multiple allegations that the proposals would cause serious harm to broadcast reception, cable television (CATV) reception, and to wireless microphones used in broadcast program production.[20] These allegations are addressed in turn below. The order of discussion here follows that of the April 8, 2005 letter sent by a broadcast industry consortium, the Coalition for Spectrum Integrity, to Senate Commerce Committee Chair Ted Stevens (R-AK).[21] After the discussion of these points, we address the issues raised in a recent web-based video from the broadcast lobby.

A. "Interference to 73 Million TV Sets"

The FCC has previously noted that only a steadily declining minority of households with televisions are actually dependent on over-the-air signal reception, and that more than 85% of American households with televisions subscribe to cable or satellite services,[22] and thus could not possibly be affected by interference from nearby unlicensed devices.[23] Nonetheless, the broadcast lobby asserts that permitting unlicensed broadband devices to operate on vacant TV band frequencies will cause a range of interference problems. The industry commissioned a Canadian laboratory study to corroborate these claims.[24] However, the results produced by the study were created under unrealistic conditions, such as certain combinations of channels and antennas pointing directly at each other. (This study also implicitly introduces the broadcast lobby's trick of using ultrawideband transmitters, permitted by a loophole in the original FCC proposal, to simulate the proposed unlicensed devices. This tactic forms the basis of a lobbying video released by the broadcast industry, discussed in Section I, below.)

B. DTV Disruption Issue

Broadcasters have claimed that implementation of the proposals would create consumer confusion and delay the penetration of DTV receivers needed to reach the 85% consumer take-up threshold mandated in current law before broadcasters would be required to cease analog transmissions. There is no evidence for this assertion. Concerns have also been raised that uncertainty about this rulemaking might make small local stations delay making final channel selections and converting to DTV. However, it now appears likely that the DTV transition end-date will be mandated, rendering this issue moot. Congress is expected to pass legislation this year that will end the DTV transition by a date certain, as well as to subsidize digital-to-analog converters and an education campaign aimed principally at the 15% of households still relying on over-the-air reception.

The broadcast community's statement that unlicensed devices may cause "interference to newly purchased DTV receivers, which may cause consumers to return their new TV sets," similarly lacks a factual basis. Today's DTVs are far more capable of handling and rejecting any potential interference than older analog sets, which are sus
impairments that pass through directly to viewers in the form of ghosts, snow, and interference patterns in the video display. To suggest that new DTVs are somehow more susceptible to potential interference than other TVs is questionable logic.

C. Public Safety Interference

The *Geolocation/Database* and *Local Beacon* alternatives in the FCC proposal use local information, such as location and databases of facilities, in deciding what channel to use. Thus unlicensed systems using these techniques could readily avoid channels 14-to-20 in the handful of markets in which they are used for public safety. The *LBT* alternative requires more complexity to avoid public safety use of channels 14-to-20 since lower power, intermittent public safety communications are harder to detect than high power, full time TV broadcasting. However, technology already exists that allows unlicensed devices to detect and avoid military radar – which is a far harder task that detecting public safety communications. The FCC can solve this problem simply by requiring a long listening period on public safety channels before they can be declared vacant.[25] Similarly, the FCC can decide to require that unlicensed devices operating on certain frequencies include the ability to recognize a priority-in-use signal transmitted by public safety systems.

D. Newsgathering and Sports Programming Production

Although not generally known, broadcasters and certain other entities are allowed to use vacant TV channels for "low power auxiliary stations" (e.g., wireless microphones) with nominal licensing under the provisions of federal regulation.[26] While this use is officially licensed, this spectrum has not been auctioned and it bears many similarities to unlicensed use except that it is reserved for a narrow group of eligible devices. These devices are used at studios, but are sometimes used at sports events and other outdoor news events.

The broadcast interests raise concerns that the wireless microphones used by broadcasters on vacant TV channels might receive interference from unlicensed devices using the *LBT* alternative. While the FCC minimized this problem in the NPRM,[27] it is a difficult problem to solve in a manner that is transparent to existing users of such wireless microphones because the microphones operate at a lower power, do not necessarily have signal formats enumerated by regulations, and do not have a formal channel plan.

But it is not at all clear that such devices should continue to have exclusive access to this spectrum. The continued exclusive access of this small group of devices to large blocks of valuable spectrum for very occasional use, independent of marketplace forces, is anachronistic and inconsistent with spectrum policies enacted by Congress and implemented by the FCC in the past two decades. The FCC should perhaps revisit why broadcasters and the narrow group of eligible entities specified in FCC regulations are granted *sole access* to the "white space" spectrum in the TV band for a use that does not involve broadcasting directly to the public. When these policies were adopted decades ago, there was no other alternative to allow use of this "white space" except manual coordination among a small group of broadcast licensees. However, today's technology has increased the demand for this type of spectrum and permits cognitive radio alternatives such as those in the NPRM. Why should wireless microphones not be treated as unlicensed devices?

Even if this anachronistic use of the white spaces is continued, however, the *Local Beacon* scheme would protect wireless microphones, as local broadcasters would control the signals indicating which channels were available in a given area at a given time. There are also compromises available that could protect users of such microphones and allow the proposed unlicensed use: the FCC could, for example, adopt a transition plan that exempts unlicensed devices from certain TV channels for a transition period. Following this period, it could then grant full interference protection to eligible wireless microphone users that transmit a low power beacon signal in the vicinity of an operating wireless microphone, and having a comparable coverage area to that microphone, indicating which TV channel the microphone was using.[28] In this way, the broadcasters would have preferential (but not sole) access to the TV band.

In the past, traditional land mobile radio technology (i.e., walkie-talkies) did not provide the audio quality required for broadcasting. Now, however, high-speed 3G cellular technology could offer broadcast-quality audio for program production with a minor variant of standard technology—that is, *if* the broadcasters were willing to pay for such a service.[29] However, the present availability of "free" spectrum for this limited group of eligible entities discourages cellular firms from developing such 3G offerings. Shouldn't the broadcasters' use of this spectrum for auxiliary purposes be subject to the same marketplace forces that apply to other spectrum users in order to ensure the highest and best use of limited spectrum resources?

E. Interference with "Theaters, Churches, and School Events"

Broadcast interests have also raised the concern that unlicensed use of the TV band might interfere with spectrum use at theaters, churches, and social events. However, they have failed to explain why these entities are even using this band, as theaters, churches and schools are not permitted to use the TV band spectrum. It appears that wireless microphone vendors have been selling their products to customers who cannot lawfully use them – and some now want to rely on those unlawful sales to prevent use of the spectrum for wireless broadband. Mass-market wireless microphones are capable of operating on the adjacent low-power Private Land Mobile Radio band, in which theaters, churches and schools *are* eligible to obtain licenses. Instead of using the TV bands, these users should use the lawful adjacent band. The equipment vendors who created this confusion should be required to help clear it up.

F. Will the Proposal "Permanently Chill Investment" in Spectrum?

The FCC proposal focuses on unlicensed sharing of channels 2-51 of the TV band spectrum, which has and will continue to have plentiful white space. The proponents of this rulemaking do not seek to expand this proposal to cover unlicensed sharing of the spectrum covering channels 52-69, which is to be licensed for non-broadcast use.

The broadcast community suggests that any regulatory change allowing unlicensed access to empty TV channels would deter investment in spectrum. Proponents of the proceeding would counter that unlicensed sharing of the TV band below channel 52 would in fact have precisely the opposite effect. By providing access to frequencies favorable for cost-effective rural coverage on an unlicensed basis, the proceeding would increase the economic incentive for deploying broadband wireless service in areas currently unserved or underserved by existing licensed wireless and wired broadband providers.

The statement of broadcast interests that "once unlicensed devices are permitted in a licensed band, there is no way to remove them" is overly dramatic and does not reflect contemporary technology. PC users routinely update their operating systems and other software to get the latest version. Demonstration versions of software with fixed expirations are common. The FCC should require that the internal software used by unlicensed devices to share the TV band white space be capable of being updated at a specified interval, so that the FCC will be able to modify the operating criteria of these systems based on experience, and even turn them all off if it so chooses. While this approach may be difficult for some types of transmitters, the transmitters in this proposal are expected to be connected to the Internet on a regular basis, and thus could check for software updates without requiring user intervention.

G. Interference to Cable Services

The allegations of the broadcast interests here fit into two sub-issues dealing with cable headends and in-home wiring. Translator stations, which pick up and rebroadcast signals in remote areas to extend a station's coverage area, raise similar issues as cable headends. Although translator stations are not specifically brought up in the broadcast industry's allegations, they are included in this discussion because of their relevance.

1. Cable Headend and Translators

Cable television systems (CATV) usually use over-the-air reception of TV signals to collect the signals for redistribution to their subscribers. The Commission's "must-carry" rules[30] result in obligations to carry certain signals that in some cases are quite weak. Thus some cable headends in rural areas have high antennas on mountaintops aimed at distant stations in order to receive these very weak signals. Some TV translators in rural areas have similar receiving systems.

Indeed, a wireless ISP using solely the *LBT* alternative with an antenna on a hillside close to a cable system headend or translator antenna might fail to notice a weak TV signal and thus cause interference to a CATV headend or translator. This type of interference could be prevented by requiring WISPs, at least in rural areas, to use options other than the *LBT* alternative for detecting vacant channels. While an end user of an unlicensed consumer device operating in the TV band might also cause interference for the same reason, this is very unlikely, because rural CATV headends and translators are explicitly located to give them a line of sight that avoids nearby populated areas in order to ensure good reception. Also, headend and translator antennas are often highly directional—that is, they are optimized to receive signals from certain angles so as to acquire as much of the distant signals as possible. This directionality desensitizes the headend receiver to any off-axis interference generated locally by a portable consumer device operating within proximity.

At present, there is no reliable database that contains the sites of rural headends or translators or the channels they receive. As was suggested by the National Translator Association,[31] there are benefits here and in other applications to encouraging formal registration of cable headends and translators in order to improve spectrum management in general. Similar optional registration of CATV headend satellite receivers has been in place for more than 20 years to ensure that they do not face interference from other spectrum users.[32]

The final rules should require that wireless ISPs at high elevations in rural areas must check the database of headend and translator input locations and avoid any use of frequencies used by headends and translators in their area. Translators and headends that choose not to register would receive no guaranteed protection.

2. Wiring Issue

The Canadian laboratory study used by the broadcast interests to demonstrate unlicensed devices interfering with over-the-air TV signals (see Section A above) concluded that even CATV users might face interference due to unlicensed signals entering "leaky" cables in their wall.[33] In order to demonstrate this, the laboratory had to aim a directional antenna at a cable at a distance of one meter. Furthermore, the cable used was of a type that is not used by the CATV industry and is not even sold by the largest US electronics retailers. Finally, in order to get this result, it was necessary that all the unused cable connections in the house had to be left "unterminated"--that is, without either a TV connected to it or an inexpensive, thimble-sized "terminator" device. These unrealistic test conditions render the findings of the study extremely unreliable at best.

"Eglin AFB Incident"

In its letter to Sen. Stevens and in filings at the FCC, the broadcast interests have repeatedly quoted a news story from *USA Today* reporting unlicensed device interference to military radars at Eglin Air Force Base in Florida in early 2005.[34] The FCC, with the concurrence of NTIA and DoD, adopted rules in 2003 to require unlicensed devices operating in the 5.250-5.350 GHz band to employ dynamic frequency selection (DFS) to ensure protection of military radar systems.[35] This DFS technology is related to the cognitive radio technology that would be used for unlicensed use of the TV band, but the technological problems associated with reliably detecting a single radar pulse of less than a millionth of a second duration versus detecting a TV signal, which is on continuously, are very different.

Furthermore, these DFS systems are not even available yet because FCC, NTIA, DoD, and interested parties are still negotiating the details of the testing to verify compliance.[36] Thus, DFS-equipped unlicensed devices could not have caused the problem at Eglin AFB because they are not yet available. The delay in developing consensus for testing methods mentioned by the broadcast interests is not a problem; rather, it is a sign that the spectrum policy-making system is behaving responsibly in delaying final implementation until a consensus is reached on the difficult issues.

It is puzzling why the broadcast interests also included in their letter to Sen. Stevens a copy of the FCC's February 15, 2005 public notice dealing with garage door openers possibly receiving interference near military bases.[37] While garage door openers and military communications systems share the same frequency, the priority is very clear and is given for all unlicensed devices in federal regulations: unlicensed devices can not cause interference to licensed systems and must accept any interference caused by licensed systems. Potential unlicensed devices operating in the TV band would be subject to the same requirement. The likelihood of interference is very different in the case of military systems--which are not designed to avoid problems with garage door openers--and the proposed unlicensed devices in the TV band, which are specifically designed to use all available technology to avoid creating interference.

III. "Your Neighbor's Static"

In August 2005, the Association for Maximum Service Television (MSTV), an arm of the broadcast lobby, released a video on its website alleging to show the interference that would be caused by unlicensed devices operating in the TV band.[38] Ignoring standard scientific methodology, MSTV did not include any details to show how an independent observer could reproduce its results; it stated simply that the device demonstrated was "an FCC-compliant unlicensed device," and could cause interference to DTV sets at distances up to 78 feet and to analog TV sets up to 452.7 feet "even through multiple walls."

Informal discussions with an individual involved with the production of the video reveal that the simulated unlicensed device exploited a longstanding loophole in FCC Rules that has never caused a problem using real transmitters in the field.[39] The device demonstrated is - MHz-wide noise generator (covering the bandwidth of nine TV channels) – essentially an ultrawideband transmitter. This device would normally be forbidden by existing and proposed FCC Rules, but the loophole permits it to be used in existing unlicensed bands in conjunction with a more powerful signal limited to 6 MHz.

The present FCC rules were written two decades ago when test instrumentation was less advanced than it is today. As broadcasters well know, the FCC's rulemaking contemplates new rules and device certification requirements that will be designed specifically to avoid interference with broadcast reception. This loophole in the Part 15 unlicensed rules, which would theoretically permit ultrawideband emissions in TV spectrum, can be closed once and for all if the FCC includes in its Report and Order in this proceeding an additional easily-measured total limit on power in the TV bands for out-of-band emissions.[40]

Conclusions

The FCC made a reasonable and important proposal in May 2004 to give unlicensed access to under-utilized TV band frequencies to devices that meet rigid technical specifications. The FCC has proposed several alternative means to ensure there would be no harmful interference to television reception or to public safety operations, as required by law. The ability of "smart radio" technologies to avoid interference is well-established, and technology industries have suggested additional improvements. A comprehensive record has been established at the FCC. Legislation that mandates an end to the DTV transition will have the side effect of removing a major uncertainty affecting this proposal. The other concerns about interference raised by the broadcast interests in this proceeding can be easily resolved through normal rulemaking.

Acknowledgements

This paper was written for the New American Foundation's Wireless Future Program, http://www.spectrumpolicy.org/

References

[1] FCC Notice of Proposed Rulemaking in Docket 04-186, http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-04-113A1.pdf

[2] "Unlicensed and Unshackled: A Joint OSP-OET White Paper on Unlicensed Devices and Their Regulatory Issues" OSP Working Paper Series, Federal Communications Commission, http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-234741A1.pdf

[3] The FCC has agreed with the National Telecommunications and Information Administration (NTIA), which regulates Federal Government use of spectrum pursuant to 47 USC 305, that it will coordinate with NTIA all rule changes that might cause interference to Federal Government radio systems. *See* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-230835A2.pdf

[4] *See* 47 CFR §§15.205,209

[5] *See* 47 CFR 15.247

[6] For most types of equipment, this authorization is done by a Telecommunications Certification Body accredited by the FCC or a foreign counterpart of the FCC pursuant to a mutual recognition agreement. However, for new classes of equipment FCC usually insists on retaining "hands on" control of final approval of new models until there is a consensus with industry on how the testing is to be done and interpreted. *See* http://www.fcc.gov/oet/ea/ - sec0

[7] 47 USC 302(a),(b) FCC issued $350,000 in fines in 2004 for equipment marketing violations. *See* http://www.fcc.gov/eb/reports/Jan2005.pdf

[8] "Wi-Fi and WiMAX Key Drivers of Wireless Equipment Spending, with Market Reaching $29.3 Billion by 2008", Telecommunications Industry Association press release, 4/21/05, http://www.tiaonline.org/media/press_releases/index.cfm?parelease=05-24

[9] *See* Report and Order, Docket 81-413, 1985

[10] Kibria, Masud and Chris Knudsen, "Capital Expenditure Implications of Spectrum Assets in Semi-rural Environments," Intel Corporation, August 23, 2005, p3.

[11] *See* 47 CFR 15.407(h)

[12] 5 USC 553

[13] At the frequencies used for TV broadcasting, radio signals do not act like rays of light with clear shadows. But obstacles such as buildings and terrain do result in some shadowing and signal decrease.

[14] The Powerpoint presentation given by Dr. John Betz of MITRE Corp. is available at

http://www.fcc.gov/realaudio/presentations/2003/021203/featuredetection.pdf One illustration of lower threshold for detection versus good reception can be found in tuning an analog TV set with over-the-air reception. One can notice which channels have distant and weak signals by seeing rolling snowy signals that can not be viewed as local signals can.

[15] The Commission had previously taken note of this research in its NPRM on cognitive radio The Commission stated, "there are techniques that can be used to increase the ability of a sensing receiver to reliably detect other signals in a band which rely on the fact that it is not necessary to decode the information in a signal to determine whether a signal is present. … For example, sensing can be made more sensitive by using bandwidths much smaller than a 6 MHz TV channel and/or can look for specific features of the TV signal such as the visual and audio carriers." *Notice of Proposed Rule Making and Order* in ET Docket No. 03-108, 18 FCC Rcd 26859 (2003). At para. 20 and fn. 35 http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-03-322A1.pdf

[16] *See* Comments of Michael Marcus, Sc.D. http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6516482949 and Shared Spectrum Company http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6516982986

[17] Comments of Adaptrum, Docket 04-186, at para. 17-19 http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6516482775

[18] An example is Qualcomm's SnapTrack technology which is used in cellular E-911 systems. *See* http://www.snaptrack.com/impact/index.jsp

[19] Radio propagation is generally has a large random component due to the same reflections that cause "ghosts" in analog TV reception. Therefore, the fix to this problem must state that the statistical confidence limit in the coverage area of the beacon signal must match the validity of the channel availability data to a high confidence limit such as 99%.

[20] Examples of the comments filed by broadcast TV interests are : http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6517610710, http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6517587197, and http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6516983613

[21] This letter is on file at FCC. *See* http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6517610710

[22] FCC has reported that in 2003 66.1% of US TV households had cable service and 19.1% had DBS service. See http://www.fcc.gov/Reports/fcc2006budget_main.pdf at p. 12. Independently, a Consumer Electronics Association study this summer disputed the NAB's number of 73 million affected sets finding only 33.6 million sets with over-the-air reception. http://www.ce.org/press_room/press_release_detail.asp?id=10764) See also Speeding the DTV Transition: Facts and Policy Options, New America Foundation Issue Brief, May 2005, at p. 1 http://www.newamerica.net/Download_Docs/pdfs/Doc_File_2389_1.pdf

[23] However, there is a separate issue of preventing interference to certain rural cable TV headend receivers which will be discussed later in this paper.

[24] The study is quoted, in part, in the 11/30/04 Comments of NAB and MSTV but has never been published in enough detail that its results could be reproduced. *See* http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6516883657

[25] TV channels are used continuously for all or most of the day. Public safety channels are more intermittent. There are 120 pairs of 25 kHz channels in a 6 MHz TV channel. These generally have a high transmitter (repeater) on each pair. If none of these 120 pairs have any detectable signal after a monitoring period of, say 1 hour, one can be certain one is not in an area where the channel is used for public safety.

[26] *See* 47 CFR 74.832(a),(c),(d)

[27] *See* para. 38 of NPRM

[28] A possible format for this beacons signal would be to have it imitate the narrow band pilot tone that is an integral part of DTV signals and thus would be looked for by devices searching for the presence of DTV signals. The beacon should be limited to having a comparable coverage area to the microphones in use.

[29] In rural areas there may not 3G availability for several years. But there is little wireless microphone use by broadcasters in such areas and they could either use microphones based on the unlicensed systems in this proceeding or microphones using Private Land Mobile spectrum for which they are eligible.

[30] *See* 47 CFR §§76.51-76.65

[31] *See* http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6516883675

[32] *See* 47 CFR 25.131 Only satellite receivers which have an optional license receive protection against C band terrestrial users.

[33] *Supra* Note 25

[34] Sen. Stevens letter at p. 4. Article referenced is "High Speed Net, Wi-Fi Interfering with Military Radar", *USA Today*, Jan. 28,2005

[35] *See 5 GHz U-NII Report and Order*, ET Dkt No. 03-122, 18 FCC Rcd 24484 (2003)

[36] *See* FCC Public Notice, OET Clarifies Equipment Authorization Policy for Unlicensed National Information Infrastructure (U-NII) Devices Operating in the 5 GHz Band, January 26, 2005, http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-05-175A1.pdf

[37] FCC Public Notice, Consumers May Experience Interference To Their Garage Door Opener Controls Near Military Bases, February 15, 2005 http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-05-424A1.pdf

[38] Video available at: http://www.tvtechnology.com/dlrf/one.php?id=986

[39] The NPRM stated that the out-of-band emissions of the proposed new unlicensed devices must comply with longstanding §15.209(a) of the FCC Rules. This is the same limit that applies to a variety of other unlicensed devices and which for 25 years has also applied to millions of home computers. (The computer limits in §15.109(a) are the same as this limit in the UHF TV band.) While many unlicensed devices are subject to a peak power limit in §15.37(b), the wording of the proposal implicitly exempted the proposed devices from the peak power limit that would have prevented the effect shown in the video.

[40] TV receivers are uniquely subject to this ultrawideband interference and the ultrawideband rules, §15.501,525 forbid UWB in TV bands. TV receivers have a very wide tuning range to accommodate all channels and try to achieve high sensitivity (which equivalent to low "noise figure") using modestly priced components. There is a basic tradeoff between sensitivity and rejection of undesired signals in the same band and consumer-grade TV receivers, as demonstrated in the video, have a susceptibility to ultrawideband signals which do not occur in real environments outside the laboratory.

Reclaiming the Vast Wasteland
The Economic Case for Re-Allocating to Unlicensed Service the Unused Spectrum (White Space) Between TV Channels 2 and 51

By J.H. Snider*
New America Foundation
Phone: 202-986-2700, Fax: 202-986-3696
snider@newamerica.net

On May 12, 2004, the FCC issued a Notice of Proposed Rulemaking (NPRM) proposing unlicensed use of unused TV channels 2-to-51 (Docket 04-186). Although incumbent TV broadcast license holders and their vendors opposed the NPRM, leaders of the high tech community, including Intel and Microsoft, came out strongly in favor of it and submitted compelling technical evidence that unlicensed devices could be introduced into the unoccupied TV guard bands without causing harmful interference to neighboring occupied bands. On October 26, 2005, the House Commerce Committee approved a digital TV bill with language asking the FCC to complete the rulemaking.

Having lost in the FCC rulemaking, the TV broadcasters have shifted their locus of battle to Congress, where they are now trying to kill the FCC's NPRM. As a second best strategy, they have also lobbied Chairman Martin to table the rulemaking. Congress needs to ensure that the NPRM is completed and that, given the record established in the NPRM, the burden of proof remains on the broadcasters when they seek to block the use of this extremely valuable public resource.

Background

The broadcast TV band is famously underutilized, mostly because of the large number of vacant TV channels[1] known as "guard bands" (alternately known as "white space" or "taboo channels") that have historically served as an interference buffer between local TV stations. But just as air conditioning technology made the Southwest into prime real estate, digital technology is transforming the TV guard band spectrum into prime spectrum real estate. Indeed, one of the major debates of the digital TV (DTV) transition concerns how this so-called TV band "white space" will be divvied up. It's in the TV broadcast industry's interest to keep others out of the white space and gradually win free access to it for itself.

Guard band spectrum has historically been the buffer between local broadcast TV stations, protecting them from harmful interference. With analog TV technology, for example, if channel 15 is used in one market, then channels 14 and 16 cannot be used in the same market and channel 15 also cannot be used in adjacent, surrounding markets. How much guard band spectrum is available around the country? There are 210 local TV markets in the United States. Each is currently allocated 67 channels (channels 2-to-69, excluding channel 37 for radio astronomy and medical telemetry). Of these, the average market only uses approximately seven high-power channels (a high-power channel is one that covers its entire market, whereas a low-power one may only cover a small fraction of the market). Since large markets such as New York City have many more high-power stations than small markets such as Burlington, Vermont, the population weighted average number of channels per market is higher, approximately 13 stations.[2] In either case, the ratio of unused to used channels is high--more than five to one. It is no wonder that many have called the TV band spectrum a vast "wasteland" of underutilized spectrum.[3]

Digital technology allows many of these guard bands to be used. For example, during the DTV transition, each existing broadcaster has been loaned a second channel so it can simultaneously operate an analog and digital channel. At the end of the digital TV transition, broadcasters must give back one of their two channels. This has fueled debate about what to do with those freed up channels.

Broadcasters have already laid claim to some of the guard band spectrum separate from the loaned channel they must return to the public after the digital TV (DTV) transition. For example, as part of the DTV transition, many local TV stations have been allowed to expand their coverage areas, thus eating into the guard band spectrum in adjacent markets.[4] In addition, TV producers, including TV stations and cable TV operators, have also been granted exclusive use of guard band spectrum for very low-power devices that can be used in TV production.

The FCC's current TV allotment plan mandates that after the DTV transition, channels 52-to-69 will be freed up in all 210 local TV markets in the United States. Four of these channels are being reallocated for public safety agencies, while ten others are likely to be auctioned for exclusive, licensed use by commercial wireless firms. However, even

* *J.H. Snider is a Senior Research Fellow at the New America Foundation, and the author of* Speak Softly and Carry a Big Stick: How Local TV Broadcasters Exert Political Power *(iUniverse, 2005).* This paper is a condensed version of a Working Paper that can be found on New America's website at www.spectrumpolicy.org.

after channels 52-to-69 are returned, substantial guard band spectrum will remain, especially in small TV markets, on the 49 channels from channels 2-to-51. The difference is that these freed up channels will not be contiguous. For example, an unused channel in Baltimore may be in use in the adjacent markets of Washington, DC and Philadelphia.

Until recently, it was thought that non-contiguous spectrum allocations would have very little economic value--just like forty scattered quarter-acre real estate parcels may be less valuable for commercial development than a contiguous ten acre lot. Why would a manufacturer want to produce a wireless device that couldn't be used nationally? How would it be possible to make a portable radio device that would work in Baltimore on a particular channel but wouldn't work in Philadelphia on the same channel, even if transported there? Accordingly, the guard band channels that would continue to be allotted market-by-market in Swiss cheese fashion after the digital TV transition generated relatively little commercial interest.

However, the technological environment has rapidly changed. With the advent of low-power, "smart radios" providing broadband service, the ability of localized wireless broadband operators to utilize non-contiguous spectrum has dramatically increased. High-tech companies, including Intel and Microsoft, have used their substantial technological and economic credibility to argue that such "smart radios" are the perfect application for this Swiss cheese guard band spectrum. In response, the FCC issued an NPRM proposing unlicensed use of unused TV channels 2-to-51, subject to strict equipment certification requirements to avoid harmful interference with DTV reception.

The broadcasters have fought tooth and nail to oppose this use of the guard band white space. Publicly, they have argued that unlicensed use of this spectrum will cause intolerable interference with existing TV stations, thus slowing down the DTV transition and perhaps even rendering all over-the-air television unusable. Privately, they have sought to win free access to this guard band spectrum for themselves. Responses to these actions are discussed in depth in two companion papers issued by the New America Foundation.[5] Briefly, these papers argue that the broadcasters' technical comments are without merit, and call attention to the broadcasters' below-the-public-radar strategy to win free rights to white space, including the unpublicized transfer of $6 billion worth of TV guard band spectrum to broadcast industry licensees since 1997. Holding up competing uses of spectrum until the government eventually gives up and allocates all the spectrum rights to local TV broadcasters is a clever lobbying strategy. But it's not one that Congress and the FCC should reward.

This paper is structured in three sections. Section 1 argues that the white space should be reallocated from broadcast to broadband use. Section 2 explains the technological and economic forces behind the shift from licensed to unlicensed use. Section 3 provides an overview of non-economic arguments for unlicensed use.

I. From Broadcast to Broadband

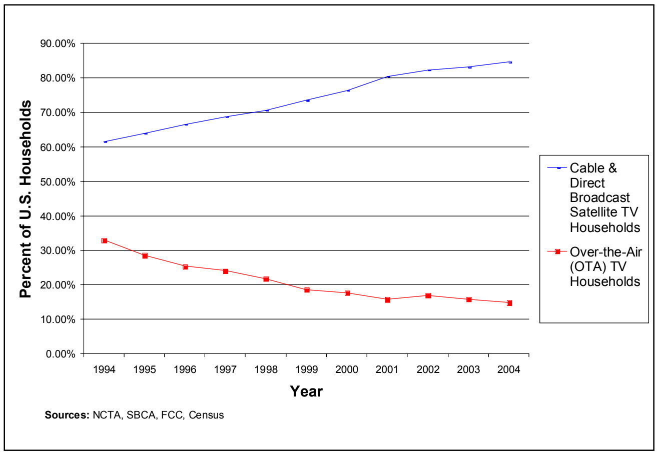
Since the mid-1980s, prominent telecom policy analysts have been arguing that broadcasting is a misuse of low-frequency spectrum. In the mid-1980s, Nicholas Negroponte, founder of the MIT Media Lab, popularized the idea of the Negroponte Switch: the idea that video services such as broadcast TV would migrate to wired telecommunications, and audio services such as telephone calls would migrate to wireless telecommunications.[6] In the original formulation of the Negroponte Switch, stationary services (such as broadcast TV) should use wires; and mobile services (such as talking while driving or roaming within your house) should use spectrum. At the same time, the FCC initiated a proceeding—later defeated by the National Association of Broadcasters (NAB) on the grounds the spectrum would be needed to transition to HDTV—to reallocate 168 MHz of unused broadcast spectrum to non-broadcast services.[7]

In 1990, George Gilder wrote a book titled *The Death of Television*, which elaborated on this basic idea that conventional broadcast TV was a great misuse of spectrum.[8] Since then, there have been dozens of telecom analysts that have made much the same argument.[9]

The two underlying economic reasons why over-the-air (OTA) TV broadcasting is a misuse of low-frequency spectrum are fairly simple. First, over-the-air broadcasting has close yet superior substitutes. Most notable, both satellite and cable TV can provide the same programming as local broadcast TV but with more reliable signal quality (e.g., hills and buildings don't degrade images), greater geographic coverage (in the case of satellite, the entire continental U.S.), and more programming choice (as many as 100 times more channels of the same resolution). This reality has resulted in the continuing decline in demand for over-the-air broadcast TV. From 1970 to 2005, the percentage of US television households relying exclusively on over-the-air reception for their TV has declined from essentially 100% to less than 13%,[10] with a drop of about 14 percentage points coming in the last decade alone.[11] This drop is remarkable, given that it has happened despite huge government subsidies to preserve over-the-air TV and despite the fact that an additional fee is required to view identical local broadcast TV programming over cable or satellite TV. So far, the figures for *digital* OTA TV are even more dismal. As of 2004, 40.4% of Americans had access to digital TV but only 2.7% of those relied on broadcast DTV. The rest relied on cable DTV (50.7%) and satellite DTV (46.6%).[12]

This is not to say that over-the-air broadcasting does not retain a niche, especially among households with low demand for TV or those who cannot get either satellite or cable service for some reason. However, this niche is getting smaller for fundamental technological and economic reasons. Figure 1 depicts the decline of terrestrial over-the-air TV and the rise of cable and satellite TV.

***Figure 1*– The Decline of Over-the-Air Television**



Second, the opportunity cost of continuing to use low frequency spectrum for broadcasting has become increasingly evident. The demand for broadband Internet information services is skyrocketing. Americans want high-speed anywhere/anytime/anything information services, which conventional digital broadcasting cannot deliver but which the low-frequency spectrum broadcasters occupy is ideally suited to provide. This is reflected in the fact that nobody purchases low-frequency spectrum today to provide conventional, fixed broadcast TV services, digital or otherwise; for this type of spectrum, the market values mobile, interactive, Internet-based information services. Congress has conceded as much in its DTV transition plans for the future of spectrum occupying channels 52-69. No member of Congress is arguing that those channels should be allocated for more conventional digital broadcast TV service. Figure 2 depicts the decline of over-the-air broadcasting and the rise of broadband.

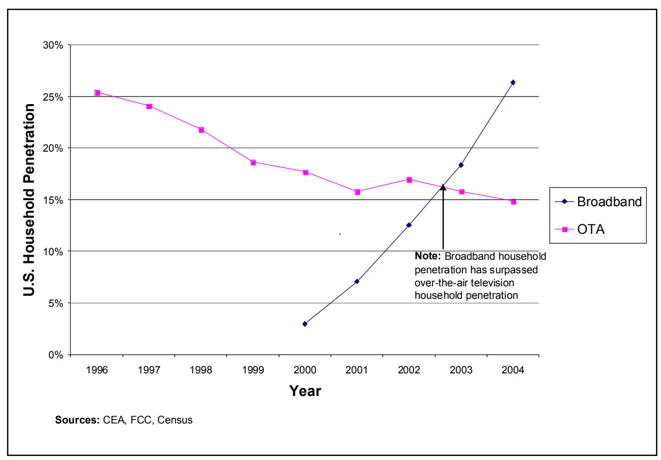
The economics favoring low-frequency spectrum for non-broadcast services is based on the physical characteristics of the spectrum. Low-frequency spectrum is better suited for mobility because its waves are longer and can thus better pass through objects such as walls, foliage, and weather.[13] All terrestrial mobile telephone services, for example, are located below 3 GHz (the lowest 1% of the radio spectrum). If cell phone service went out every time someone passed a tree or building, its utility would be minimal. Similarly, WiFi service would be much less valuable if it couldn't pass through walls, furniture, people, pets, and other common household obstructions.

Higher-frequency spectrum is primarily used for line-of-sight applications such as a direct connection between a satellite and a home satellite dish or a point-to-point microwave link used as a backhaul between a building rooftop and a fiber node several miles away linked to the Internet backbone. The primary reason that high-frequency spectrum sells for much less than low-frequency spectrum is that it competes with close substitutes from wired services. Instead of using a point-to-point microwave link, for example, a company can use an optical fiber link and get the same or better service. In contrast, there are no wired substitutes for portable service.

Another major physical advantage of low-frequency spectrum is that it requires less energy than high-frequency spectrum to cover the same distance. The large waves that characterize low-frequency spectrum lose less energy when they pass through objects. As a result, they can cover greater distances with the same power. This, in turn, means that battery-powered devices can be less expensive, longer-lived, smaller, and lighter. In the emerging era of ubiquitous, portable wireless devices, this can be a great advantage.

Lastly, lower-frequency devices require fewer cell towers – and hence substantially lower infrastructure costs – to cover a given geographic area. This is a corollary of the power observation above. If power is held constant, then coverage is enhanced with lower frequency spectrum. This savings in tower expense is especially important in rural areas where broadband service is less constrained by the amount of spectrum and more constrained by the cost of additional or higher towers to reach residents. An Intel study estimates that a rural cell tower transmitting at 700 MHz can cover more than four times the territory of the same tower transmitting at 2.5 GHz.[14] Assuming that towers are the fundamental constraint on rural broadband deployment, low-frequency spectrum for broadband can reduce rural broadband deployment costs by 75% or more.

***Figure 2* – The Rise of Broadband and Decline of Over-the-Air Television**



II. From Licensed to Unlicensed

When household, business, and government entities consider low-power terrestrial wireless applications, they have increasingly come to the conclusion that unlicensed spectrum offers them service at lower cost and higher quality than licensed spectrum. Already, tens of millions of American households and more than two-thirds of U.S. businesses use unlicensed WiFi technology —a remarkable feat for a product that only became generally available five years ago. Other popular unlicensed technologies enjoying explosive growth include Bluetooth, Zigbee, and UWB (ultra-wideband). Given that the FCC and Congress have strongly favored licensed products in the amount and quality of spectrum they have allocated, this feat is all the more remarkable—and an achievement that the legions of Washington lobbyists seeking more spectrum rights for licensed carriers have done everything they can to sweep under the rug.

What explains the shift from licensed to unlicensed spectrum services? Most people would agree it is inefficient for the Federal government to sell toll booth rights to third parties to collect payment when anybody uses local real property such as public roads, private homes, or businesses. It turns out that the same economic logic is being played out with spectrum rights. Real property may be physical in a way that spectrum is not, but the underlying economic logic is surprisingly similar—and becoming more so. As I will argue, for many good reasons the world is moving toward networks of low-power devices, such as household WiFi, enterprise WiFi, municipal WiFi, and highway WiFi. Forcing households, enterprises, and local governments to purchase spectrum rights from a third-party license holder for strictly localized, low-power uses of spectrum (a Federal government mandate that acts, in effect, like a hidden tax) needlessly adds cost while also often reducing quality of service.

Unlicensed devices are generally found in four types of locations: homes, workplaces (including offices, hospitals, college campuses, and warehouses), retail establishments (including coffee shops, hotels, libraries, and airports), and public rights of way (including municipal roads, highways, and subways). As Figure 3 suggests, a growing variety of private and public sector institutions are deploying wide area wireless broadband networks on *un*licensed frequencies.

Most important from a policy perspective, unlicensed devices can be either low-power or high-power. It takes more energy to transmit over larger distances, so—all other things being equal—lower-power devices cover a smaller geographic area than higher-power devices. FCC-approved lower-power unlicensed devices usually focus their energy within the property lines of a particular entity. An example of a small area device would be a WiFi router covering a home; an example of a large area device would be a cell tower covering a square mile.

Small area devices can be networked together to cover a large area, usually still focused within the property lines of a particular entity. Thus, there are two types of unlicensed large area networks: one type comprised of high-powered devices and the other type constituted of many low-power devices. Failure to recognize this distinction between the two types of large area networks has been the source of great public confusion and chicanery by advocates of more licensed spectrum. It is typically the basis on which they create a straw man argument that unlicensed service cannot provide large area coverage without chaos stemming from a "tragedy of the commons"—the mismanagement of a free resource that becomes degraded through overuse. But, as we shall see, this argument reflects a profound misunderstanding of the growing importance and ubiquity of networked small area devices.

Figure 3 - Sampling of Wide Area Unlicensed Networks

Universities
Dartmouth College, Hanover, NH
Carnegie Mellon University, Pittsburgh, PA
United States Military Academy, West Point, NY

Hotels (all with Free WiFi)
Best Western
Courtyard (Marriott International Inc.)
Doubletree Hotels (Hilton Hotels)

Hospitals
Baycrest Centre for Geriatric Care, Toronto, Canada
Children's Memorial Hospital, Chicago, Illinois
John C. Lincoln Hospital, Phoenix, Arizona

Manufacturing, Distribution, and Inventory Management
Biggs' Hypermarket, Mason and Harrison, Ohio
Nine Mile Point Nuclear Station
Nike, Memphis, Tennessee

K12 Schools
Lincoln Unified School District, Stockton, California
Arlington Independent School District, Arlington, Texas
Fairfax County Public Schools, Fairfax, Virginia (available in more than 200 schools)

Retail
ALLTEL Stadium, Jacksonville, Florida (host of 2005 SuperBowl)
Barnes & Noble Bookstores, hundreds of locations
Starbucks, thousands of locations

Municipalities, Outdoor Public Access
Philadelphia, Pennsylvania (planned)
Corpus Christi, Texas
Chaska, Minnesota

Municipal, Outdoor Public Safety
Lower Valley Public Safety Network, Yakima County, Washington
City of Aurora Police and Fire Departments, Aurora, Colorado
City of San Mateo Police Department, San Mateo, California

Convention and Sports Centers
American Airlines Center, Dallas, Texas
Connecticut Convention Center, Hartford, Connecticut
William A. Egan Civic and Convention Center, Anchorage, Alaska

Airlines (only international travel)
Lufthansa
Japan Airlines
Korean Air

Airports
Atlanta International Airport, Atlanta, Georgia
Baltimore-Washington International Airport
Boston, Logan International Airport, Boston, Massachusetts

Other
Marinas (Beacon WiFi supplies WiFi service to more than 100 boat marinas)
RV Parks, (Boingo supplies WiFi service to hundreds of RV parks)
Flying J truck stops (hundreds of locations)

Consider municipal WiFi, the fastest growing and most high-profile type of low-powered wide area network.[15] These unlicensed networks can traverse great distances via public roads and other public rights of way. For example, Philadelphia's plan to build a franchised municipal WiFi system will cover the entire 135-square-mile footprint of the city.[16] And the Canamex highway WiFi network in Arizona may cover more than 500 miles before it is complete.[17]

Tens of thousands of other large spaces, including college campuses, hospitals, malls, warehouses, stadiums, K-12 schools, amusement parks, and office buildings, have been building networks of small area devices that collectively cover large areas. Similarly, thousands of Wireless Internet Service Providers (WISPs) have been providing unlicensed coverage to households and businesses in rural areas where the signal passes through a lightly populated area, often in a focused beam.

A basic rule of thumb in spectrum allocation is that unlicensed spectrum is more efficient for small area devices (including networks of small area devices that collectively cover large areas) and licensed spectrum for large area devices (such as broadcasting). Even advocates of licensed spectrum have been extremely careful not to explicitly argue in public that they should be allowed to take possession of spectrum rights within property contour lines. Instead, they have sought to divert attention with misleading claims related to potential interference, enforcement problems, and tragedies of the commons. It is therefore of great significance for spectrum policy that emerging economic forces strongly favor the use of low-frequency small area devices as a substitute for low-frequency large area devices.

The Shift to Lower-Power Wireless Devices in the Lower Frequencies

During the early years of radio, the most prominent terrestrial wireless services tended to send signals over great distances. Moreover, they used single, relatively high-power devices to do so. At the beginning of the 20th century, for example, the most famous demonstration of radio's utility was a terrestrial transmission across the Atlantic Ocean from England to the United States. Later, TV and radio broadcasters typically used a single transmission tower to cover thousands of square miles. Early cell phone companies, too, typically covered many square miles with a single transmitter. Vividly demonstrating the diminishing size of cells, New York City recently leased out its 18,000 light posts, each a potential cell site for up to a half-dozen wireless vendors. See Figure 4 for the growth of cell towers. This growth has largely been driven by the need to subdivide cells to increase information capacity. Since each cell can reuse spectrum, the information capacity of a cellular network is directly proportional to the number of cells.

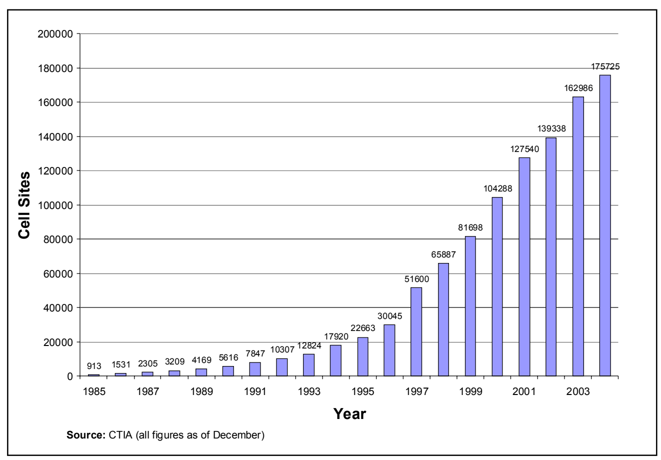
One major economic force leading to the growth in terrestrial low-power wireless communications is that high-power wireless service has close wired substitutes. Over time, optical fiber is moving closer and closer to the premises. Optical fiber is relatively expensive to deploy but is otherwise a superior technology to wireless for backhaul--that is, linking small area networks to the Internet backbone. Fiber's capacity is huge, and it has excellent quality of service. For example, a single strand of optical fiber has more information carrying capacity in a direct point-to-point communication than the entire radio spectrum. For this reason, the major telephone companies and cable operators are bringing high-speed fiber lines to the neighborhood and eventually to the premises in every high-density area in the United States.

Nevertheless, wireless communication remains a highly valued complement to wired communication. As wired communication nears the individual, it loses its quality advantage because it cannot provide anytime, anywhere (i.e., mobile) service.

As wires approach the individual, their cost advantage also tends to diminish. For example, the cost of digging a trench on a major city street can be shared by tens of thousands of customers; that is, it has great economies of scale. But by the time the wire gets to the premises, the cost of laying the wire can only be shared by the relatively small number of people at the wire's destination.

For these reasons of both quality and cost, the long-term economic logic of the terrestrial communications system is to bring wires as close as possible to the individual, but leave the last part of the communications link wireless.

Figure 4 **- Growth in Cell Sites**



A second major economic force leading to lower power devices is the growing opportunity cost of large wireless cell sites. Just as demand for Internet backbone capacity is skyrocketing, so is demand for spectrum capacity. People want faster, higher fidelity, interactive communications and they don't want to have to be plugged in to access it. At the same time, the supply of spectrum is fixed. Carriers can purchase rights to use additional spectrum. But since the supply of spectrum is not infinite, this ultimately means robbing Peter to pay Paul.

The long-term strategy, then, must be to expand the information carrying capacity of spectrum, especially low frequency spectrum. Carriers can do this by employing more efficient data compression technology or developing more advanced modulation technologies to squeeze more bits of information on a single electromagnetic wave. Such strategies are useful as far as they go, but they are strictly limited. The most efficient long-term strategy to increase the information carrying capacity of spectrum is to geographically subdivide it so that it can be reused in different geographic areas. ArrayComm CEO Martin Cooper has estimated that more than 97.5% of the increase in spectrum capacity since 1960 has come from reducing the geographic coverage area of cells.[18] One way to subdivide geographic coverage is with directional antennas that point signals in a specific direction and thus can reuse spectrum in different directions. Another way is to subdivide cells to cover smaller and smaller areas, with each cell able to reuse the same spectrum.

The extent of this dilemma is illustrated by today's mobile telephone services. Even the most advanced services are currently struggling to provide 1 Mbps of mobile service. For example, the Verizon Wireless 3G service (called "EV-DO") only provides *mobile* broadband users up to 700 kbps—and that is under highly optimistic conditions. To provide service at 10 Mbps, 100 Mbps, or more, Verizon Wireless would have to migrate to ever smaller cell sizes, which helps explain the demand for wireless sites on New York City's light posts. With mobile telephone service or today's typical broadband services, higher speeds may not be critical. But as Americans spend ever increasing amounts of time on the Internet accessing ever higher-bandwidth applications, the demand for spectrum bandwidth will continue to skyrocket, requiring ever shrinking geographic coverage.

Now consider this thought experiment that highlights the underlying economic logic. Assume that the cost of a low-power wireless transmitter drops to zero while demand for bandwidth increases to infinity. The economic equilibrium derived from such assumptions would be an infinite number of infinitesimal cell sites.

Of course, these assumptions, as stated, are unrealistic. The cost for wireless transmitters will not drop to zero, and the demand for bandwidth will not grow infinitely. However, the cost of factory ordered WiFi chips has already dropped to $5/each in high-volume purchases and that number could drop to pennies within a few years. Fry's Electronics already sells a WiFi access point at retail for $19.95. In contrast, a high-power TV transmitter may still cost over $1 million. Meanwhile, Verizon, Comcast and others are already building wired networks to homes and businesses with a planned capacity of 100 mbps or more. Using today's conventional state-of-the-art mobile telephone cell architecture, there isn't enough low-frequency spectrum in the universe if the thousands of households within a cell must all share the same spectrum and expect to receive 100 mbps wireless service. Thus, although these assumptions are unrealistic, they do highlight a fundamental economic force driving cell architecture.

Another advantage of low power is less battery usage. As portable devices grow in popularity, efficient battery use grows in importance. Physics dictates that the greater the distance a wireless device must send its signal, the greater the power it must use as well as the corresponding size, weight, and cost of batteries.[19] Low-power also opens up the possibility of solar-powered WiFi, which is useful for a host of military, scientific, and municipal applications, as well as in disaster relief, developing countries, and highly rural areas, where there is unreliable or no electricity.[20]

Similarly, physics dictates that the amount of energy required to send information is a function of the number of bits sent. Every additional bit requires more energy. When telephone-quality audio bits are the predominant type of bits sent, power usage is relatively low. But as we move into a world of CD-quality voice communication, interactive video, and other high bandwidth applications, hundreds of times more power may be needed. When the bits are coming from a battery-operated portable device, this becomes a major problem. One way to address it is with lower power links between the transmitter and receiver.

Another advantage of lower power is more comprehensive coverage. The conventional wisdom is that pervasive computing and communications requires a high-power wireless network. But, in fact, the opposite is the case. Wide area networks tend to miss many spaces blocked by impenetrable barriers such as hills, buildings, and elevator shafts. Mobile telephone service, for example, is frequently unavailable within commercial buildings and homes, especially in low-density areas. That's why major commercial buildings and underground public transportation systems often have their own very small local area cells.[21] J.D. Power calculates that 3 out of 100 cell phone calls has a quality of service problem.[22] But it doesn't calculate the much greater number of calls that aren't made because people have learned not to expect service.

Another advantage of lower power is more precise coverage. Let's say a local government wants to cover its public spaces, including the public roads that link every house and business in its territory. Low-power allows it to do this without interfering with other, nearby low-power users unless those users seek access to its network. Many municipal WiFi networks, for example, are designed in default mode to focus their coverage within public rights-of-way.

Another advantage of low-power is greater security. Wired communications are more secure than wireless communications because of the confined space in which they operate; it's necessary to dig up a wire to intercept a shielded, buried wired communications link. But the last wireless leg of a communications link is relatively easy to intercept with any device in its coverage area. Thus, the smaller the coverage area—for example, a corporate campus vs. an entire city—the more secure the connection.

All this analysis does not deny that there are economic advantages to large cell sites, most notably the higher costs associated with more cell sites. This economic logic is most striking in rural areas that are range limited rather than capacity limited. In rural areas, cells cover large distances but few people, so there isn't enough demand to justify subdividing cells. For example, only in such areas does WiMax's boast of providing 70 mbps of service over a radius of 30 miles make any sense. In a dense urban area like New York City, the same WiMax transmitter would only provide a trickle of service—perhaps at an even lower speed than a dialup modem—and probably miss the vast majority of people due to the obstruction of large buildings.

Consequently, rural areas with low population density will continue to have larger cell sizes than urban areas with high population density. But as the demand for wireless information soars and the cost of low-power wireless equipment plummets, the economic tradeoffs between low-power and high-power devices—even in rural areas—shift decisively to the advantage of low power.

Links Between Low-power Devices, Unlicensed Spectrum, and Economic Efficiency

The essence of unlicensed spectrum is decentralized, local control of spectrum rights. Confidence is placed in local property owners to figure out how best to use their spectrum rather than the Federal government, which is ill equipped to determine the needs of tens of millions of homeowners, millions of businesses, and tens of thousands of municipalities. It turns out that this local control has many beneficial economic consequences in terms of increased innovation, lower costs, and higher-quality service. To the extent that the Federal government has allowed such local control, it has been embraced by homeowners, businesses, and local governments on the demand side, and venture capitalists, entrepreneurs, and manufacturers on the supply side.

Figure 5 compares the growth in devices manufactured to operate on unlicensed spectrum with mobile telephone authorizations for licensed cellular and PCS bands. Observe that the unlicensed 2.4 GHz band--the largest unlicensed band in the prime low-frequency spectrum below 3 GHz--has more than 25 times the number of authorizations as the mobile telephone bands. This is despite the fact that the mobile telephone bands occupy far more spectrum (more than twice as much) and far better spectrum (the 2.4 GHz band has both a higher frequency than the mobile telephone bands and is nicknamed the "junk" band because unlicensed devices must accept interference from a host of other devices that use that band, including licensed devices and dumb, non-telecommunications emitters such as microwave ovens).

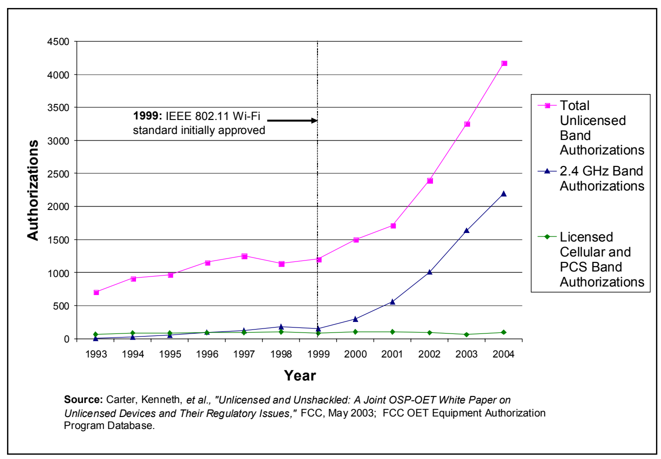
Also note that most of the unlicensed growth has occurred since 1999. That growth is primarily attributable to the development of smart unlicensed devices, such as WiFi, in the 2.4 GHz band. Previously, dumb unlicensed devices such as garage door openers dominated the unlicensed market. It is also noteworthy that unlicensed devices, like licensed devices, are overwhelmingly located in the low frequencies. At these frequencies, the equipment is cheaper, can be positioned without regard to physical obstacles like walls, and uses less battery power.

Now let's look more closely at the economic advantages of unlicensed spectrum.

Lower Barriers to Entry for Manufacturers. For manufacturers of wireless products, unlicensed spectrum has lower barriers to entry, leading to more competition and innovation. With licensed technology such as mobile telephone service or public safety communications, entrepreneurs must first get permission from the license holder before launching their innovation. This creates a number of problems.

Many manufacturers consider securing rights to use licensed spectrum from private parties as similar in difficulty to getting rights to use spectrum from the FCC. Like government license holders, private license holders may create huge bureaucratic obstacles before granting permission to use their spectrum, and the outcome may be highly uncertain. In the high-tech world, a delay of six months in getting a product to market can be the difference between success and failure.

Figure 5 **– FCC Device Authorizations for Licensed and Unlicensed Bands, 1993-4004**



Many licensed bands employ proprietary technologies with large license fees that discriminate against small companies. For example, license fees to use W-CDMA, a popular cellular telephone standard, may be 30% of the total product cost for a small manufacturer but as little as zero percent for a large manufacturer with more negotiating power and its own patents to barter.[23] When small players have to pay a 30% premium for the same product, it discourages innovation. WiFi is an open standard, so is not burdened by such royalty payments.

Entrepreneurs also worry about holdup problems and uncompensated appropriation of their ideas. In addition to a royalty, the licensee may insist on a cut in the profits of any successful innovation and may choose to compete with the entrepreneur if the innovation proves especially lucrative. Consider Ibiquity, the new digital radio standard for the AM and FM bands. The large commercial radio broadcasters insisted that they get a fee for any radio device sold using spectrum where they had a license. Thus, they banded together to create a company, Ibiquity, that would develop an exclusive proprietary standard for their spectrum band. The commercial broadcasters were genuinely interested in studying other companies' proposed radio standards. But the bottom line was that if the technology used their spectrum, they wanted control of it—a demand that would discourage many entrepreneurs.

In seeking negotiating leverage, a spectrum license holder may also reveal the idea to competitors, thus eliminating the entrepreneur's first mover advantage. In fast moving high tech markets, this advantage is often critical to profitability.

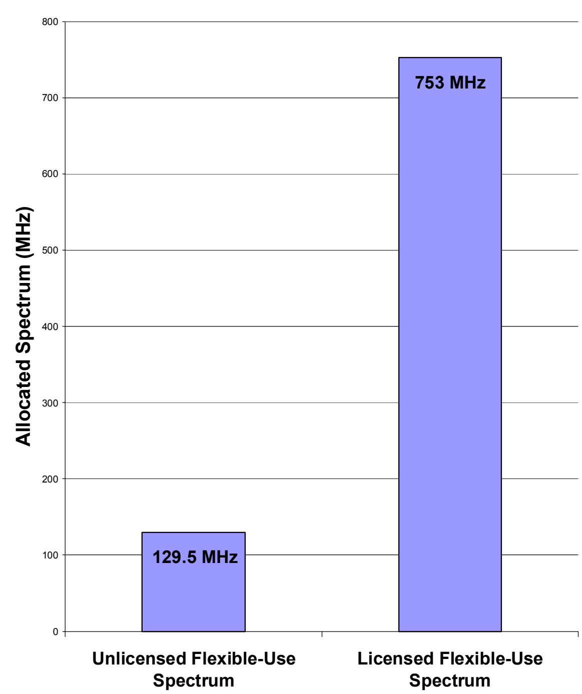
As a case study on the influence of licensing barriers to entry on market structure, compare the level of competition and innovation in the mobile telephone and unlicensed bands in the prime spectrum below 3 GHz. The mobile telephone band is a good reference case because that is where the most licensed spectrum activity takes place. In addition, the mobile telephone bands will shortly control at least five times as much spectrum as the unlicensed bands (See Figure 6).[24]

As in many other licensed bands, no mobile telephone handset manufacturer can sell a product within a particular band without first getting permission from the licensed carrier in that band. Getting such permission usually involves developing a unique model for the licensed carrier and selling it through the licensed carrier's approved retail channel. As a result of these and other economic incentives, fewer than ten handset makers, including Nokia, Motorola, Samsung, Sony Ericcson, and LG, control 99% of the U.S. retail handset market.

In contrast, there are hundreds and probably thousands of manufacturers now selling unlicensed devices, despite the fact that the mobile telephone industry is more than two decades old and the new industry of smart, unlicensed devices barely five years old. These companies include Dell, Scientific Atlanta, Intel, HP, Linksys, D-Link, Panasonic, Sony, Starkey Laboratories, Kodak, Canon, Nikon, Sony, Microsoft, Hexagram, Sharper Image, Nortel, Cisco, Motorola, Toyota, BMW, Zensys A/S, Logitech, Connexion, Lumin, Tropos, BelAir, Ember, Chipcon, Freescale, Vocera, Avaya, Colubris, Spectralink, CardioNet, Crossbow Technology, General Electric, Palm, Nintendo, and Honeywell.

A major reason these companies exist is that they sell highly differentiated products targeted at narrow market niches. Indeed, most of these companies the public has never heard of precisely because they are targeted to such narrow market niches. Consider the mobile video surveillance system developed by ODF Optronics, an Israeli company. The product consists of a ball that a public safety official (e.g., police, fire, or military) can throw into a building and on a remote screen monitor receive a 360-degree view of the room. The entire worldwide market for this product may be tiny compared to the market for a mobile telephone handset. But that doesn't mean the product isn't extremely valuable and capable of saving many lives.

***Figure 6* – Licensed vs. Unlicensed Flexible Spectrum Under 3 GHz**



Sources: Kobb, Bennett, *Wireless Spectrum Finder* (2001), FCC, Legg Mason, 47CFR §25.2

Lower Barriers to Entry for Carriers. Just as there are lower barriers to entry for manufacturers, there are lower barriers to entry for carriers. Unlike wide area networks, there are minimal economies of scale in local area networks. This is true whether the networks are wired or wireless. Again, contrast mobile telephone and unlicensed markets. Mobile telephone service is dominated by just four carriers: Verizon Wireless, Cingular, Sprint Nextel, and T Mobile. In contrast, thousands of carriers have emerged in the unlicensed space in the US alone. These include between 4,000 and 6,000 WISPs providing WiFi service to mostly rural areas;[25] more than 85 municipal and regional governments providing WiFi networks for public use and/or government and public safety agency use (with at least 34 more networks planned or under construction);[26] and more than 20,000 coffee shops, airports, truck stops, and many other retail businesses in America.[27] (See Figure 3 above for a sampling of carriers.)

Lower Usage Costs for End Users. An increasing number of household, business, and government entities have access to wired broadband connections via DSL, cable, and fiber.

When these entities look for wireless service on or near their premises, unlicensed usually becomes the obvious low cost solution. For example, why should an entity pay Verizon Wireless $60 per month per individual (plus about 15% in taxes) for wireless data service when its premise is already linked to high speed wired service and can add a wireless component for zero dollars per month per individual? This largely explains the significant pressure on mobile telephone carriers to introduce dual mode handsets that can carry both licensed and unlicensed communications. The carriers hate this idea because up to 40% of the minutes used by their customers are made in household and business premises where WiFi is likely to be used.[28] In addition, there is the threat that free or low cost WiFi will be strung on more roads, thus depriving mobile telephone companies of their bread and butter revenue. WiFi networks are also open networks whereas mobile telephone networks are mostly closed, which means that mobile telephone operators would be likely to lose content and transaction revenues that they can currently monopolize.

Finally, American carriers have been especially resistant to genuinely open dual handsets because more than 50% of the mobile telephone market is controlled by two operators, Cingular and Verizon, which also have wired networks. When consumers switch to WiFi calls, the operators will not only lose toll minutes on their wireless networks but also toll minutes on their wired networks. Still, the business pressure is becoming so great that dual mode WiFi phones are expected to become widespread within the next few years.

Lower Equipment Costs. A number of factors have led unlicensed equipment to have lower equipment costs than most licensed equipment. These include lower royalty rates and greater economies of scale. Unlicensed chips are designed for flexible use and mass consumer markets, so are relatively inexpensive even if installed in a highly specialized product. Contrast, for example, the cost of public safety and WiFi equipment. A Motorola public safety phone costs in the vicinity of $3,000 whereas a WiFi access point with comparable technological sophistication costs only about 1% of that, or $30. Most of the difference is simply due to economies of scale.

For mobile telephone technology, the production economies of scale are comparable to WiFi. But the equipment costs for entities larger than a household tend to be much greater. This is because of the need to install redundant equipment from multiple carriers. Many markets have four to six mobile telephone carriers. To get ubiquitous in-building coverage for all potential licensed users, an entity needs to install equipment from each of these vendors. This may be cost effective for large, heavily trafficked entities such as sports stadiums and malls. But for smaller entities, such as the vast majority of businesses and local governments in the U.S., standardizing on a single WiFi standard may be more efficient.

Higher Quality for End Users. In real world applications, unlicensed spectrum has many quality advantages over licensed mobile telephone spectrum. These are the same advantages leading to the growth of low-power devices, and include better coverage, faster speeds (due to more efficient use of spectrum), smaller devices (due to less need for power and smaller batteries), more security, and higher quality of service. Consequently, the most demanding wireless users, notably large, sophisticated businesses, are shifting to unlicensed for reasons of quality as well as cost.

A major advantage of unlicensed spectrum for business is greater control, including tight integration with corporate PBXs, which are widely perceived to allow for better transferring, parking, monitoring and filtering of calls than mobile telephone networks. Businesses are increasingly seeking to have on-premise mobile employees, and they want those employees to be able to carry their work and use the same PBX features, including internal extension number, wherever they go. With WiFi, they can do this whether the employee is working at the corporate campus, telecommuting from home, or working out of a hotel. This is especially important in businesses, including hospitals, hotels, warehouses, retail stores, and universities, where a large fraction of employees are constantly moving around.

Businesses also want more control over quality of service. A large fraction of mobile telephone calls are dropped. When the CEO of a major corporation is making a wireless call to a vital client, he doesn't want the call dropped because a teenager two miles away is chatting with his girlfriend. The mobile telephone company doesn't offer him a way to ensure his call gets through. But through integration of a VOIP WiFi phone into his PBX, he can do that.

Businesses also want more control of internal security. Both licensed and unlicensed wireless devices now have similar encryption technology to prevent unauthorized access to information. But high-power out-of-building mobile telephone signals are much more vulnerable to hackers and corporate espionage.

Businesses also want more control of coverage. Only a small percentage of businesses have complete on-premise mobile telephone coverage. Elevators, basements, nearby buildings, steel or concrete walls, and factory machines are just a few of the obstacles that typically pose barriers to ubiquitous coverage.

Businesses also want high speeds where they need it. Security, medical, and warehouse personnel may have a need for high speed images and video on the go. For example, a doctor in an emergency room may highly value the ability to download a patient MRI sixty times faster via an unlicensed (WiFi) than a licensed (mobile telephone) network. Indeed, the extra speed may be the difference between life and death for a patient.

Many products and services wouldn't even exist without unlicensed spectrum. Today, the vast majority of wireless products are only manufactured to use unlicensed spectrum. For example, the Sony Portable Playstation video game player, the Kodak EasyShare digital camera, and the Dell Axim personal digital assistant have built in WiFi to connect to the Internet but no mobile telephone links. The reason is obvious. Manufacturers can include a WiFi chip for about $5/device, users don't have to pay usage charges, and the speed of connection is faster. Embedding a mobile telephone in one of these products is possible, but in practice has proven prohibitively expensive for most consumer markets.

III. Non-Economic Arguments

This paper has focused on the economic arguments for unlicensed spectrum. But there are also First Amendment, universal service, public safety, and takings clause arguments for unlicensed spectrum.

First Amendment. Spectrum is the 21st century's medium for speech. Decentralized control of this medium fosters robust free speech, a fundamental value long recognized in the United States for its economic and democratic value. Along these lines, it is revealing that one of the best indicators of whether a country supports unlicensed use of spectrum is whether it is a dictatorship. The 15 countries in the world that require a license to use WiFi are Bahrain, Belarus, China, Cuba, Democratic Republic of the Congo, Kazakhstan, Macau, Mongolia, Myanmar, Oman, Pakistan, Sri Lanka, Ukraine, Vietnam, and Zimbabwe.[29] Every European and North American country allows unlicensed WiFi.

Universal Service. America is now 13th in the world in broadband penetration.[30] Low-frequency, unlicensed spectrum is critical to bringing affordable broadband services to poor, undeserved communities. This is a major factor explaining the explosion in both urban (municipal) and rural (WISP) WiFi deployments. The low-cost, high-quality calculus of unlicensed spectrum has proven to be an unbeatable formula for bridging the broadband divide. For example, Philadelphia's WiFi franchisee, Earthlink, is offering broadband service to low income households for $10/month, less than 25% of the cost of the broadband service offered by its cable franchisee, Comcast.

Public Safety. A rapidly growing number of municipal and county public safety agencies are using unlicensed spectrum to build out high-speed mobile data networks, despite the fact that they have free access to licensed spectrum. First responders are driven to use unlicensed spectrum for four primary reasons. First, real world public safety agencies have limited budgets. Second, unlicensed equipment is less expensive, primarily because it is mass produced for all market segments, not just public safety. Third, telecommunications is a fixed cost business, so sharing infrastructure costs across multiple market segments reduces the costs any one market segment must pay. Fourth, numerous public safety products are only designed to use unlicensed spectrum, primarily because many public safety equipment entrepreneurs have recognized there is no net advantage to using licensed spectrum. Insofar as unlicensed spectrum results in more public safety services being purchased on a limited budget, unlicensed spectrum results in more lives saved, which are presumably priceless.

Takings Clause. Allowing the Federal government to take control of local spectrum rights via a tacit form of eminent domain (that is no less consequential because it deals with the invisible airwaves rather than real property) and then give away those rights to a handful of the largest and most politically powerful companies in the U.S. (albeit in the name of "deregulation," "spectrum flexibility," "investment certainty," and other Orwellian claims) should be an outrage to all Americans because it is a takings of their property without just compensation.

Of course, "property rights" to electromagnetic spectrum must be tempered by free speech and anti-monopoly considerations. We don't allow local governments to unduly control acoustic speech on public property (imagine the outrage if a local government banned people from freely talking with each other while using public property such as a street or park!). Similarly, we should not allow local governments excessive control of electromagnetic speech.

The same principle applies to private property. For example, the FCC's over-the-air reception device (OTARD) rules prevent a landlord from extracting a monopoly rent from a tenant for installing a relatively inconspicuous antenna to pick up a signal. This principle should also apply to unlicensed devices.

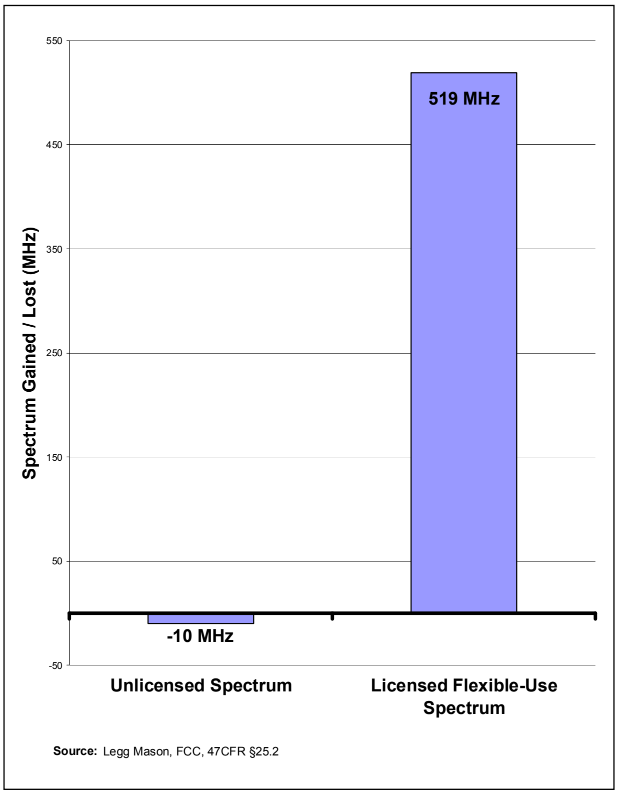
Conclusion

This paper has argued that the best use of the TV guard band white space is for unlicensed broadband services. Driving the analysis are the unique propagation characteristics of the low-frequency TV band and the growing economic importance of low-frequency, low-power spectrum applications, as exemplified by the growth of home WiFi, enterprise WiFi, and municipal WiFi.

Obviously, there continues to be an economic case for terrestrial broadcast and licensed spectrum. However, that case is getting weaker while the economic case for broadband unlicensed spectrum is getting stronger.

The policy implication of this analysis is that a new balance must be struck between the allocation of licensed and unlicensed spectrum. Specifically, the balance should be shifted to favor unlicensed spectrum—especially in the most valuable lower-frequency spectrum. In fact, spectrum policy has done just the opposite. It has extended the duration of licenses and dramatically shifted spectrum allocation in favor of licensed use. See Figure 7.[31]

***Figure 7* – Reallocations of Spectrum Below 3 GHz Since November 2002 FCC Spectrum Policy Task force Report**



In the context of the digital TV transition, the choice Congress and the FCC now face is even simpler: 1) warehouse the unused guard bands, or 2) make them available for public use. These frequencies are the crown jewel of the information age. They should be put to good use. The moment has come to stop wasting them in what amounts to one of the great political disgraces and economic tragedies of our times.

Endnotes

1 Also described as “taboo channels.”

2 Thomas W. Hazlett, "The Wireless Craze: An Essay on Airwave Allocation Policy," *Harvard Journal of Law and Technology* (Spring 2001).

3 Mark Lewyn, “Airwave Wars,” *Business Week*, July 23, 1990.

4 J.H. Snider, *Speak Softly and Carry a Big Stick: How Local TV Broadcasters Exert Political Power* (New York: iUniverse, 2005), Appendix D: Valuation of Guard Band Spectrum Acquired by Broadcasters (1997-2004).

5 For a discussion of the broadcasters’ technical arguments, *see* Michael J. Marcus, Paul Kolozy and Andrew Lippman, “Reclaiming the Vast Wasteland: Why Unlicensed Use of the White Space in the TV Bands Will Not Cause Interference to DTV Viewers,” New America Foundation, Wireless Future Program Issue Brief #17, October 2005. For a discussion of the history of broadcast-industry attempts at securing access to guard band channels for themselves, *see* J.H. Snider, “Myth vs. Facts: A Response to Broadcast Industry Misinformation Concerning Possible Interference from “Smart” Wi-Fi Devices Using Vacant TV Channels,” New America Foundation Fact Sheet, September 2005.

6 Stewart Brand, *The Media Lab: Inventing the Future at MIT* (New York, N.Y.: Viking, 1987), Nicholas Negroponte, *Being Digital*, 1st ed. (New York: Knopf, 1995).

7 This is described at length in Joel Brinkley, *Defining Vision: The Battle for the Future of Television*, 1st U.S. ed. (New York: Harcourt Brace, 1997).

8 George F. Gilder, *Life after Television* (Knoxville, Tenn.: Whittle Direct Books, 1990).

9 E.g., Thomas W. Hazlett, "The U.S. Digital TV Transition: Time to Toss the Negroponte Switch," (Washington, DC: AEI-Brookings Joint Center for Regulatory Studies, 2001), Bruce M. Owen, *The Internet Challenge to Television* (Cambridge, Mass.: Harvard University Press, 1999), Snider, *Speak Softly and Carry a Big Stick: How Local TV Broadcasters Exert Political Power*. Craig Moffett, "Cable and Satellite: Search Versus Browse," in *Bernstein Research Call* (New York City: Sanford C. Bernstein & Co., 14 July 2005).

10 *See* “Comments of the Consumer Electronics Association,” FCC Media Bureau Docket 04-210, Inquiry Into Over-the-Air Broadcast Television Viewers, August 2004, p.2.

11 *See* FCC *Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming*, “Eleventh Annual Report” (February 2005), Table B-1 and “Sixth Annual Report” (January 2000), Table C-1. Household OTA penetration is calculated as the percentage of households without MVPD service.

12 “OECD Communications Outlook 2005,” (Paris, France, OECD Publishing, 2005), p. 222.

13 *See* Remarks of Ed Thomas, “DTV 201: How the DTV Transition Can Move the Nation from Broadcast to Broadband,” New America Foundation & House Future of American Media Caucus Luncheon, September 7, 2005. Available at: http://www.newamerica.net/Download_Docs/pdfs/Doc_File_2547_1.pdf.

[14] Masud Kibria and Chris Knudsen, "Capital Expenditure Implications of Spectrum Assets in Semi-Rural Environments," (Hillsboro, Oregon: Intel, 4 August 2005).

[15] E.g., *see* Jesse Drucker et al, "Google's Wireless Plan Underscores Threat to Telecom," *Wall Street Journal*, 3 October 2005, p. A1.

[16] *See* "Wireless Philadelphia™ Business Plan," Wireless Philadelphia Executive Committee, February 2005, p.12.

[17] Eliot Cole, "Wi-Fi the Highway," Mobile Government, June 2005, pp. 22-25.

[18] *See* "Cooper's Law" at www.arraycomm.com. See also J.M. Vanderau et al., "A Technological Rationale to Use Higher Wireless Frequencies," (Washington, DC: U.S. Department of Commerce, February 1998), p. 10, and Toru Otsu et al., "Network Architecture for Mobile Communications Systems Beyond IMT-2000," *IEEE Personal Communications*, October 2001, p. 33.

[19] *Supra* Note 12.

[20] E.g., *see* Lumin Innovative Products at http://www.luminip.com.

[21] E.g., *see* FCC Tenth Report in the Matter of the Annual Report and Analysis of Competitive Market Conditions With Respect To Commercial Mobile Services, WT Docket No. 05-71, released 30 September 2005, p. 51.

[22] Dan Meyer, "Operators Make Call-Quality Gains in J.D. Power Study," *RCRWireless*, 8 August 2005, p. 13.

[23] Mike Dano, "Royalties Remain an Industry Mystery," *RCRWireless*, 12 August 2005, p. 1.

[24] Licensed mobile telephone bands included in this chart are WCS (2305-2320 MHz, 2345-2360 MHz); former government bands transferred to commercial use (1390-1395 MHz, 1432-1435 MHz); the Crown Castle band (1670-1675 MHz); Response TV (217-218 MHz, 219-220 MHz); CMRS (220-222 MHz), Broadband PCS (1850-1915 MHz, 1930-1995 MHz); Narrowband PCS (901-902 MHz, 930-931 MHz, 940-941 MHz); Cellular (824-849 MHz, 869-894 MHz); ESMR (817-824 MHz, 862-869 MHz); AWS (1710-1755 MHz, 1915-1920 MHz, 1995-2000 MHz, 2020-2025 MHz, 2110-2155 MHz, 2155-2175 MHz, 2175-2180 MHz); Ancillary Terrestrial Use of the MSS band (2000-2020 MHz, 2180-2200 MHz, 1626.5-1660.5 MHz, 1525-1559 MHz, 1610-1615.5 MHz, 1621.35-1626.5 MHz, 2487.5-2493 MHz), ITFS/MDS rebanding (2495-2690 MHz); and TV band auction spectrum (698-710 MHz, 722-740 MHz, 747-762 MHz, 777-792 MHz).

Unlicensed bands included in this chart are the 900 MHz band (902-928 MHz) and 2.4 GHz band (2400-2483.5 MHz).

[25] Research in late 2004 by the Wireless Internet Service Provider Association (WISPA) found that there are between 4000 and 6000 WISPs in the United States, serving between 1.5 and 2 million customers. *See* Wireless Internet Service Provider Association (WISPA) comments to FCC in Public Notice 04-163, September 28, 2004. WISPA surveyed the leading suppliers of equipment used by WISPs to find out their customer numbers. Accounting for overlap (WISPs that buy equipment for more than one supplier), WISPA estimated that there are 4000-6000 WISPs in the US. Customer estimates were made by surveying customer premise equipment (CPE) manufacturers for numbers of CPE units sold.

[26] Esme Vos, "Muniwireless.com Second Anniversary Report," July 2005, pp.1,5. Available at: http://muniwireless.com/reports/docs/July2005report.pdf. For an up-to-date mapping of current and planned municipal and regional wireless networks, *see* http://news.com.com/Municipal+broadband+and+wireless+projects+map/2009-1034_3-5690287.html.

[27] "2005 Telecommunications Market Review and Forecast" (Arlington, Virginia: Telecommunications Industry Association, 2005), p. 165.

[28] "Mobile and Fixed Services," *Information Week*, 10 October 2005.

[29] Robert Horvitz, "Preliminary Results of the Open Spectrum Foundation's Global Survey of WiFi Regulations," Unpublished Paper, September 2005. For more information, see http://www.openspectrum.info.

[30] S. Derek Turner, "Broadband Reality Check: The FCC Ignores America's Digital Divide," (Washington, DC: Free Press, August 2005).

[31] Licensed Mobile Telephone spectrum additions since 2002 include AWS (1710-1755 MHz, 1915-1920 MHz, 1995-2000 MHz, 2020-2025 MHz, 2110-2155 MHz, 2155-2175 MHz, 2175-2180 MHz); PCS Expansion (1910-1915 MHz, 1990-1995 MHz); Ancillary Terrestrial Use of the MSS band (2000-2020 MHz, 2180-2200 MHz, 1626.5-1660.5 MHz, 1525-1559 MHz, 1610-1615.5 MHz, 1621.35-1626.5 MHz, 2487.5-2493 MHz), ITFS/MDS rebanding (2495-2690 MHz); and TV band auction spectrum (698-710 MHz, 722-740 MHz, 747-762 MHz, 777-792 MHz).

Note that MSS providers still must provide a satellite service in addition to any terrestrial service they might choose to provide.

The FCC has reclaimed 10 MHz of unlicensed spectrum from the Unlicensed PCS band (1910-1920 MHz).

Spectrum Regulations for Ad Hoc Wireless Cognitive Radios

D. J. Cohen
D. Stewart [1]
Alion Science and Technology Corporation
8100 Corporate Drive, Suite 400
Lanham, MD 20785

***Abstract*-Ad hoc wireless networks will soon be utilized for public safety, military, sensor, and wireless 802.11 networks. These ad hoc networks will be peer to peer and not include base stations. The communication method used in these networks will likely be dynamic spectrum access cognitive radio. This paper reviews the US domestic and International spectrum regulation issues requiring attention to enable deployment of ad hoc wireless cognitive radios. An important requirement is gaining the spectrum regulatory community acceptance that machine readable spectrum access policies (both allocations and technical parameters) can be implemented in mobile transceivers on a real time basis to accomplish spectrum sharing. Other important regulatory and implementation issues are also addressed. The emphasis in the paper will be the identification of alternatives.**

I. INTRODUCTION

Utilizing cognitive dynamic spectrum access in ad hoc networks can increase the amount of spectrum available to these networks thereby improving communications performance and spectrum efficiency. Researchers such as those involved in the DARPA Next Generation (XG) Communications program hope that by using underutilized spectrum, cognitive radio will provide a 10 times spectrum capacity improvement [1].

Potential users of cognitive ad hoc wireless LAN/MAN technologies include public safety, military, homeland defense, and commercial wireless organizations. This paper identifies for these users the US domestic and international spectrum regulatory issues requiring attention. One instance of dynamic spectrum access that has been incorporated into national and international spectrum regulations is Dynamic Frequency Selection (DFS). Wireless broadband access devices implementing Dynamic Frequency Selection (DFS) must detect the presence of the primary user of a channel and automatically switch to another channel to minimize interference to primary user systems. Today (2005) regulations exist that require certain wireless broadband access devices to implement DFS services in the 5 GHz spectrum [2].

Regulatory changes supporting expanded application of cognitive dynamic spectrum access should be evolutionary, starting at a national/regional level (e.g. US and Europe) and move to a global basis. The paper is directed to an identification of a number of regulatory alternatives.

II. DEFINITIONS

First we define cognitive radio and ad hoc networks. No specific Radio Regulation either domestically in the US or internationally has defined cognitive radio. Working Party 8A of the ITU-R in a Draft New Report on Software Defined Radio (SDR) tentatively defines cognitive radio as "*A radio or system that senses, and is aware of its operational environment and can be trained to dynamically and autonomously adjust its operating parameters accordingly*".

Mobile ad hoc networks can be defined as self-organizing peer-to-peer networks of mobile stations operating without the control of a centralized access point. Each node independently determines access and the usual access protocol is Carrier Sense Multiple Access (CSMA). Nodes coordinate amongst themselves locally to determine channel access. An important parameter for spectrum management is the number of neighbor nodes a given node must have a connection with for the network to be successful. The number of connected nodes along with frequency reuse distances determines the number of channels needed for the network. In general, it is agreed that ad hoc wireless networks require each node to be connected to 6-8 neighbors. This is a subject of active study [3].

Ad hoc transmitting sources are often times not line-of-sight to their intended receiver. Instead, other nodes relay packets of information in order to send data across a network. A diagram illustrating multi-hop routing in an ad hoc network is shown in Figure 1. The message from A is communicated to B by hopping between nodes C, D, E, F, and G.

[1] The views and opinions of the authors expressed herein do not necessarily state or reflect those of Alion Science and Technology or its clients and sponsors.

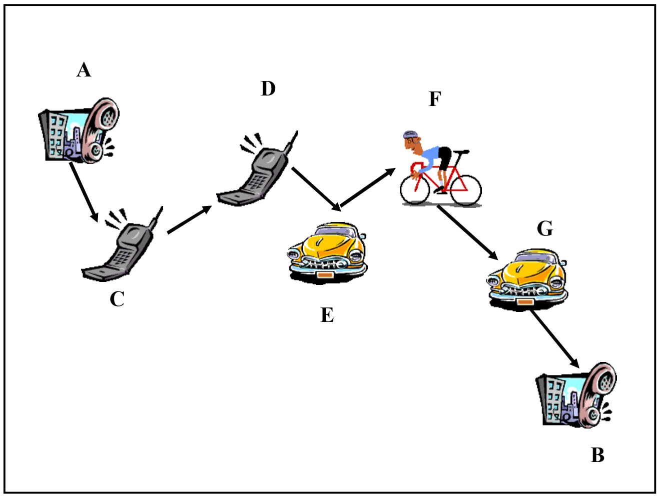


Fig. 1. Packet relay of a message.

III. REQUIREMENTS FOR COGNITIVE RADIO

Currently, the Department of Defense's (DoD's) transformation to network-centric warfare is increasing the need for radio spectrum, and the adoption of new techniques such as cognitive radio and software defined radio (SDR) may provide ways to increase the amount of information that can be communicated over the currently available spectrum. DoD mobile ad hoc networking waveform initiatives such as the Joint Tactical Radio System (JTRS) Wideband Networking Waveform (WNW) and Soldier Radio Waveform (SRW) must coexist and compete with many other military and civil systems for a limited amount of available spectrum. Additionally, the DoD has the requirement for communication in geographical areas (for example urban environments) which already include both cooperative and non-cooperative operations. The DoD organization most involved in developing cognitive radio as a potential solution to these problems is the DARPA XG program [1]. XG hopes to create technologies that can increase by a factor of 10 or more DoD's ability to efficiently utilize spectrum. The vision is for XG-enabled radios to automatically select the spectrum and operating mode that will minimize disruption of existing users while ensuring the successful operation of US systems in a battlefield environment.

After considerable technical studies and planning, the XG program is now entering the prototype demonstration phase and hopes to show through live over-the-air testing that a cognitive radio system can detect other transmitting devices and adapt its behavior to avoid interfering with them [4]. Test locations will include Federal Government-owned range(s) and one or more urban areas. The prototypes will be operated in portions of the 30 MHz to 2 GHz band to evaluate their ability to implement various dynamic spectrum sharing policies and scenarios.

To date, the most ambitious deployment of a cognitive radio system is the Canadian deployment of the Project MILTON [5]. The MILTON Network is a wireless network that continually senses the radio environment and responds to physical, electrical, and regulatory requirements. As early as 1997, Canada had the need for community broadband wireless Internet in remote areas, and the Canadian Research Centre in Ottawa conducted propagation studies of the possible use of the 5 GHz unlicensed spectrum for cognitive radio.

IV. REGULATION CONCEPTUAL OVERVIEW

Regulations for cognitive radio should specify constraints for coexistence between competing systems but not specify the implementation techniques. The implementation should be left to radio and protocol designers to specify. For example, the regulations should not specify a particular waveform to be used. The regulatory constraints should be kept to a minimum to allow flexibility in operation. The ad hoc wireless cognitive radios can operate any way they want as long as they abide by the rules for coexistence. For example, an issue needing study is whether control of the network should include control channels or beacons. This should not be included in the regulations but rather the system implementer should make the decision on how the network is operated. Also, it is necessary that the radios adhere to the regulations (compliance) which should be verified by testing.

V. A CASE STUDY ON REGULATION DEVELOPMENT-5 GHZ RADAR SHARING WITH UNLICENSED WIRELESS ACCESS SYSTEMS

A good precedent for how the regulatory development should proceed for cognitive radio is the WRC-03 allocation of wireless access systems (domestically denoted as unlicensed operations) in the 5 GHz band. At WRC -03 the regulatory community agreed on a method for 5 GHz spectrum sharing of radar and wireless access systems. The basis for the sharing was agreement on the use of Dynamic Frequency Selection in 5230-5350 MHz and 5470-5725 MHz. The specific sharing method is described in Recommendation ITU-R M 1652 Annex 1 and includes the following criteria:

- maximum signal levels and EIRPs for wireless access systems (unlicensed devices)
- listen before transmit (CSMA)
- monitoring for the presence of radars between transmissions
- maximum duration of 10 seconds after which transmission must cease if an occupying radar signal is detected
- minimum of 30 minutes before the wireless access system can attempt to reoccupy the channel vacated by the radar operation

Taking this a bit further, a good model of how specific sharing criteria might be developed for cognitive radio is that approved by the FCC in the US for this 5 GHz sharing situation. The specific sharing criteria are from the FCC rules (47 C.F.R. §15.407) for unlicensed national information infrastructure (U-NII) devices as listed in TABLE I.

TABLE I
US FCC REGULATORY IMPLEMENTATION OF DFS

⋮

(h) Transmit Power Control (TPC) and Dynamic Frequency Selection (DFS).

(1) Transmit power control (TPC). U-NII devices operating in the 5.25–5.35 GHz band and the 5.47–5.725 GHz band shall employ a TPC mechanism. The U-NII device is required to have the capability to operate at least 6 dB below the mean EIRP value of 30 dBm. A TPC mechanism is not required for systems with an e.i.r.p. of less than 500 mW.

(2) Radar Detection Function of Dynamic Frequency Selection (DFS). U-NII devices operating in the 5.25–5.35 GHz and 5.47–5.725 GHz bands shall employ a DFS radar detection mechanism to detect the presence of radar systems and to avoid co-channel operation with radar systems. The minimum DFS detection threshold for devices with a maximum e.i.r.p. of 200 mW to 1 W is −64 dBm. For devices that operate with less than 200 mW e.i.r.p. the minimum detection threshold is −62 dBm. The detection threshold is the received power averaged over 1 microsecond referenced to a 0 dBi antenna. The DFS process shall be required to provide a uniform spreading of the loading over all the available channels.

(i) Operational Modes. The DFS requirement applies to the following operational modes:

(A) The requirement for channel availability check time applies in the master operational mode.

(B) The requirement for channel move time applies in both the master and slave operational modes.

(ii) Channel Availability Check Time. A U-NII device shall check if there is a radar system already operating on the channel before it can initiate a transmission on a channel and when it has to move to a new channel. The U-NII device may start using the channel if no radar signal with a power level greater than the interference threshold values listed in paragraph (h)(2) of this part, is detected within 60 seconds.

(iii) Channel Move Time. After a radar's presence is detected, all transmissions shall cease on the operating channel within 10 seconds. Transmissions during this period shall consist of normal traffic for a maximum of 200 ms after detection of the radar signal. In addition, intermittent management and control signals can be sent during the remaining time to facilitate vacating the operating channel.

(iv) Non-occupancy Period. A channel that has been flagged as containing a radar system, either by a channel availability check or in-service monitoring, is subject to a non-occupancy period of at least 30 minutes. The non-occupancy period starts at the time when the radar system is detected.

Let us now interpret these FCC regulations. Figure 2 indicates that a WLAN node implementing the DFS capability is permitted to send normal traffic for a maximum period of 200 ms after detection of the radar signal. During the remaining time period of 10 seconds, intermittent management and control signals can be sent. This gives the cognitive radio system adequate time to terminate communications on the current channel, identify alternate available spectrum (i.e. use DFS), and move to the new spectrum channel. How it does this is up to the network designer to implement and not contained in the regulations.

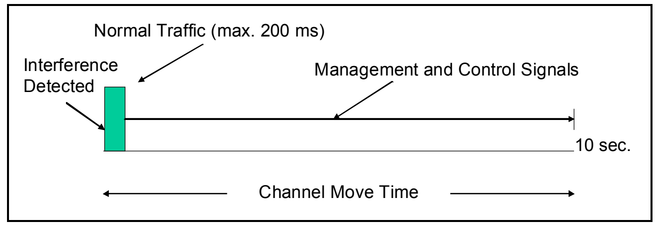


Fig. 2. Dynamic Frequency Selection for radar and unlicensed operations

Note also that the IEEE is even more lenient in its move time requirement, and in its response to FCC 03-287 [6] asserted that specifying the condition for detection of the primary radar signal occupancy is likely to be implementation dependent and need not be codified in the FCC rules. It stated that allowing this flexibility would avoid constraining the future development of innovative approaches that may provide superior performance.

VI. REGULATORY ALTERNATIVES

Below we list with comment regulatory issues needing review for cognitive radios:

A. *Machine Readable Policies*

There is a need for regulatory community acceptance (both national and international) of the principle that mobile transceivers may download dynamic spectrum access policies in a standard machine readable format (both allocations and spectrum sharing technical parameters) and use these downloads as a basis for accomplishing cognitive radio spectrum sharing. These policies may be specific to particular geographic locations, allocated bands, and sharing scenarios. This will require that spectrum regulation is written in such a way that it can be interpreted by the radio and that the radio is able to exploit such regulation [1].

B. *Acceptable Interference*

With cognitive radio it is possible that interference (probably not harmful interference) will occur when the primary system communication commences. This needs to be an agreed assumption in any regulatory scheme for cognitive radio. It is implicit in the already agreed 5 GHz radar unlicensed sharing.

C. *Allocation Footnotes and Rules for Spectrum Use*.

One rule set could specify discrete frequencies and bands that must be avoided under all conditions (e.g. distress and safety channels, Radio Astronomy, Radionavigation). Also, Fixed-Satellite uplink bands should be avoided since cognitive radios may have an unacceptable aggregate interference effect on in-orbit Fixed-Satellite receivers.

D. *Protection from Interference*

Cognitive radios should be afforded no protection from interference in the traditional sense. This is an obvious regulatory implication since cognitive radio can adapt to the spectrum environment.

E. *Flexibility in Service Definitions*.

Current allocations are based on radiocommunications services such as mobile and fixed which make specific distinctions regarding permissible radio operations. This type of distinction between services blurs with multifunction devices that send information (voice, video, data, geolocation etc.) as bits. These traditional service definitions may not be appropriate since cognitive radio devices can share spectrum with like and dissimilar systems. Indeed, with the move to more flexible spectrum use there is a need to reduce the number of service definitions.

F. *Security Requirements*

Security is critical to ensuring the integrity and reliability of cognitive radios. It is needed to protect information controlling radio operating parameters and to protect the communicated information content. The SDR Forum has studied SDR system security and has issued white papers providing an excellent overview of the topic [7] [8].

G. *More Unlicensed Spectrum Allocations*

Increased spectrum allocations are required for unlicensed use both nationally and internationally. Unlicensed spectrum is ideal in some respects for cognitive radio. Note, however, that the unlicensed bands are subject to congestion and laissez-faire operation. Indeed, [9, Section 5.3.3] makes note, "that in the world of dynamic spectrum access (DSA) there may be spectrum hogs who may monopolize the spectrum to the exclusion of others." A further look at the regulations and monitoring of unlicensed spectrum may be needed. Keeping this in mind, cognitive radio for Public Safety and DoD applications may wish and need to operate in licensed bands where there is more orderly use of the spectrum.

H. *Unlicensed Sharing of Unused TV Channels and the "3650 MHz Band"*

Another possible use of cognitive radio may be to identify unused TV channels by spectrum sensing techniques and location technologies such as GPS [10]. This would provide an opportunity for the development and use of new unlicensed wireless communications and more efficient use of the TV spectrum. Another possibility is permitting unlicensed operation in the 3650 MHz band using a DFS-like mechanism which would listen for FSS earth station uplink signals [11].

I. *Specification of Minimum Signal Levels*.

In the FCC rules (47 C.F.R. §15.407 (h)), DFS thresholds are specified for detecting the presence of radar systems and initiating a protective response. Some sensitive systems such as satellite communications, radionavigation, GPS, and radio astronomy may require very low threshold levels for interference protection. Consequently, a single threshold cannot apply to all cases, and the thresholds for dynamic frequency sharing will be scenario dependent and must be established for each sharing situation.

J. *Use of Heteromorphic Waveforms*.

Heteromorphic waveforms (e.g., Orthogonal Frequency Division Multiplexing OFDM) can morph to utilize gaps in the spectrum based on the parameters of time, space, frequency, data rate, and other characteristics to increase spectrum capacity density and thus improve spectrum efficiency [11].

K. *Worst Case Analyses*

Currently, system planning is done based on worst-case analyses to demonstrate non-interference to an incumbent system. Worst-case analyses and scenarios may not be appropriate for cognitive radio and new regulations may need to be written for cognitive radio system planning. In particular for the NTIA Manual Chapter 10 and possibly other regulatory texts such as international ITU-R Recommendations, new text may need to be added for system planning of cognitive radio. Eventually, the ITU Radio Regulations may need additions.

L. *Experimental Allocations*

The allotment of spectrum to support dynamic spectrum access experimentation should be considered to support the test and evaluation of various dynamic access and sharing techniques. This would provide a small amount of spectrum, such as the 10-MHz band segments recommended in the President's Spectrum Policy Initiative, to prove the benefits of cognitive radio dynamic spectrum access [12].

M. *Tiered Spectrum Access Rights*

Cognitive radios employing the most sophisticated spectrum access techniques will be able to most efficiently and effectively utilize the spectrum. One method of encouraging radio developers and users to employ the best technology is to provide these users and systems with preferential spectrum access opportunities as compared to users having less capable radios. Examples of features that enable improved spectrum access include geolocation capability (via GPS or Galileo), high sensitivity RF sensors, group behaviors to increase detectability of hidden nodes, morphable spectrum-agile waveforms, smart antennas, advanced artificial-intelligence reasoning capability, and adaptive power control.

VII. TESTING OF UNLICENSED COGNITIVE RADIO SYSTEM INTERFERENCE ON METEORLOGICAL RADAR

The latest WRC-2003 allocation of the 5250-5350 MHz and 5470-5725 MHz bands for wireless access systems is being considered by the Canadian government for unlicensed “last mile” solutions in rural and remote areas as a form of broadband Internet access. Applications may include real-time transfer of video files, CD files, digitized voice, and high data content information.

To determine the optimum regulatory implementation of this ITU WRC-2003 allocation, Canada recently performed testing of the impact of wireless access DFS systems on Meteorological Radars [13]. In confirming the viability of dynamic spectrum access, these tests concluded that: (a) the DFS would detect the radar sooner than the wireless access system would corrupt the radar and (b) the degradation to the radar is related to the mean amount of interference power and a mean power of –79 dBm is the threshold for reflectivity degradation.

VIII. INTERNATIONAL PERSPECTIVE

Currently, changes to the International Radio Regulations occur roughly with a lead time of 5-10 years. Thus changes to the International Radio Regulations tend to be evolutionary rather than revolutionary. The topic of modernization of spectrum management is not on the agenda for WRC-07. This is unfortunate for new allocation concepts such as cognitive radio which fit in this topic area. The WRC process needs to more rapidly respond to new technologies such as cognitive radio but changes will be difficult to implement since regulation changes are based on member nations reaching consensus.

There exists a current international regulation, Article 4.4 that provides some support for cognitive radio. Article 4.4 states: Administrations of the Member States shall not assign to a station any frequency in derogation of either the Table of Frequency Allocations in this Chapter or the other provisions of these Regulations, “*except on the express condition that such a station, when using such a frequency assignment, shall not cause harmful interference to, and shall not claim protection from harmful interference caused by, a station operating in accordance with the provisions of the Constitution, the Convention and these Regulations*.”

The ITU-R Recommendation ITU-R M 1652 Annex 1 discussed earlier permits wireless access systems to stay on a channel for ten seconds before moving off. This implicitly assumes that during this period, the wireless access system is not causing harmful interference and thus can be interpreted as an action in accordance with Article 4.4 of the International Regulations.

IX. CONCLUSIONS

Regulatory action to establish DFS as a recommended method for dynamically sharing spectrum between unlicensed wireless access devices and radiolocation systems in the 5-GHz band created a regulatory precedent for cognitive radio. Extending this very narrowly defined precedent to multi-band and multi-function cognitive radios will be challenging since the elements of a general-purpose cognitive radio are not sufficiently defined to adopt universal spectrum management regulations. The US DoD is currently undertaking the proof of concept [1] utilizing prototypes and demos that may help point to the appropriate regulatory approach.

Specific regulations can be adopted after a successful demonstration of cognitive radio prototypes. Within the US, NTIA and FCC should encourage this process through activities of The President’s Spectrum Policy Initiative [12]. Specific candidate scenarios for dynamic spectrum sharing should be identified, following the DFS example, and radio-independent standards should be established for sharing/coexistence. After proof of the cognitive radio concept is established, the NTIA and FCC should institute domestic regulations and support international adoption of cognitive radio techniques. In the US federal sector, the DoD should leverage the results of the XG demonstration program to support this regulatory evolution to dynamic spectrum access. It is hoped that the European spectrum management community will also conduct proof-of-concept demos and afterwards adopt spectrum regulations in support of cognitive radio.

REFERENCES

[1] XG Working Group, *The XG Vision*, Request for Comments, Version 2.0, Fall 2004

[2] *ITU Radio Regulations (WRC-03)*, “Allocation of 5.47-5.725 GHz to the mobile service for implementation of wireless access systems (WAS) including RLANs, subject to the provisions of Resolution 229”

[3] Xue, Feng and P.R. Kumar, “The Number of Neighbors Needed for Connectivity of Wireless Networks”, *Wireless Networks*, 10, 2004, pp. 169-181, Kuwer Academic Publishing

[4] *Proposer Information Pamphlet (PIP)*, BAA 05-05 for Defense Advanced Research Projects Agency (DARPA) Advanced Technology Office (ATO)

[5] Sydor, John, “Interference Resolution and Control in High Frequency Reuse Environments using Cognitive Radio”, Berkeley Wireless Research

Center Workshop on Cognitive Radio, November 2004
[6] IEEE 802.18 RR-TAG Comments to FCC 03-287 (2003)
[7] *Report on Issues and Activity in the Area of Security for Software Defined Radio*, SDRF-02-A-0003-V0.00, SDR Forum Security and Architecture Working Group, September 2002
[8] *SDR System Security*, SDRF-02-A-006-V0.00, SDR Forum Security and Architecture Working Group, November 2002
[9] Berggren et al., *Dynamic Spectrum Access*, Phase 1 Report TRITA-S3-RST-0407, September 23, 2004
[10] FCC NPRM, *Unlicensed Operation in the TV Broadcast Bands*, ET Docket No. 04-186, May, 2004
[11] FCC NPRM, *Unlicensed Operation in the Band 3650-3700 MHz*, Docket No. 04-151, April 2004
[12] NTIA (2004), *Spectrum Policy for the 21ST Century – The President's Spectrum Policy Initiative: Report 1, Recommendations of the Federal Government Spectrum Task Force*, June 2004
[13] Brandao, Andre, John Sydor and Wayne Brett, "5 GHz RLAN Interference on Active Meteorological Radars", IEEE VTC2005-Spring 2005

Wireless Access Standards and Spectrum

José M. Costa
NORTEL
3500 Carling Avenue
Ottawa, Ontario CANADA K2H 8E9
Tel.: +1 613 763-7574
costa@nortel.com

Abstract

The purpose of this presentation is to describe the standardization and spectrum activities on Wireless Access, in particular Broadband Wireless Access (BWA), in the ITU. This includes the recent collaboration of ITU with IEEE and ETSI for the development of broadband wireless standards, as well as other relevant developments in ITU such as RLANs, IMT-2000 and IMT-Advanced. The presentation includes a description of the ITU organization, relevant standards and spectrum Recommendations produced to date, and work plans for future activities. Emphasis is made on the peculiarities of the use of the radiofrequency spectrum and the additional flexibility that it provides to enable fixed, mobile and nomadic wireless access applications.

Outline

- Broadband Wireless Access (BWA) activities in ITU
- Wireless Metropolitan Area Networks (WMAN)
- Broadband Radio Local Area Networks (RLAN)
- International Mobile Telecommunications - 2000 (IMT-2000) and systems beyond IMT-2000 (IMT-Advanced)
- Summary

International Telecommunications Union Broadband Activities

ITU-D	ITU-T	ITU-R
Development Sector	**Telecommunication Standardization Sector**	**Radiocommunication Sector**
• Assisting developing countries	• NGNs • Mobile telecomm networks • Broadband cable networks	• WMANs • RLANs • IMT-2000 and beyond
Study Group 2	Study Groups 9, 13 and 19	Study Groups 8 and 9

3 ISART 7-9 March 2006

Significant activities in ITU-R

ITU-R – **Radiocommunication Sector**

Radiocommunication Advisory Group (RAG)

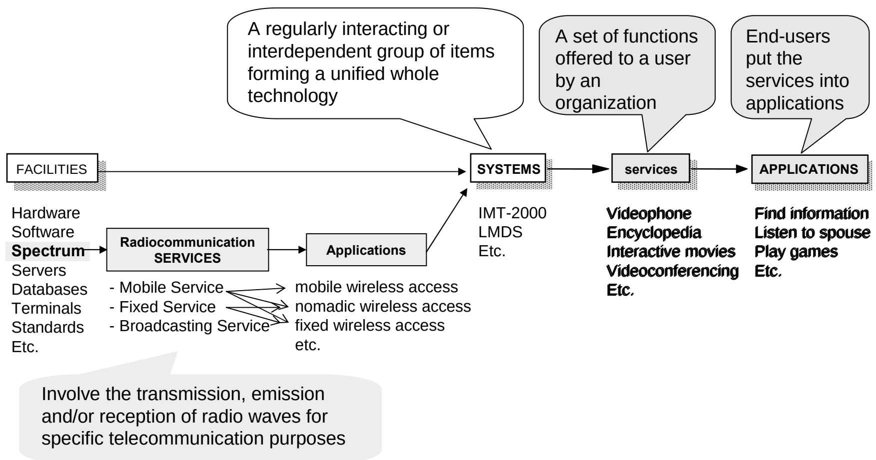
- **Spectrum Management (SG 1)**
  - **UWB (TG 1/8)** (finished)
- **Fixed Satellite (SG 4)**
- **Broadcasting (SG 6)**
- **Mobile Radio (SG 8)**
  - **2500-2690 MHz sharing sat./terr. (JTG 6-8-9)**
  - **Land Mobile (WP 8A)**
  - **IMT-2000 & IMT-Advanced (WP 8F)**
- **Fixed Radio (SG 9)**
  - **Fixed systems (WP 9B)**

WRC-07: Agenda Item 1.9 (JTG 6-8-9); Agenda Item 1.4 (WP 8F)

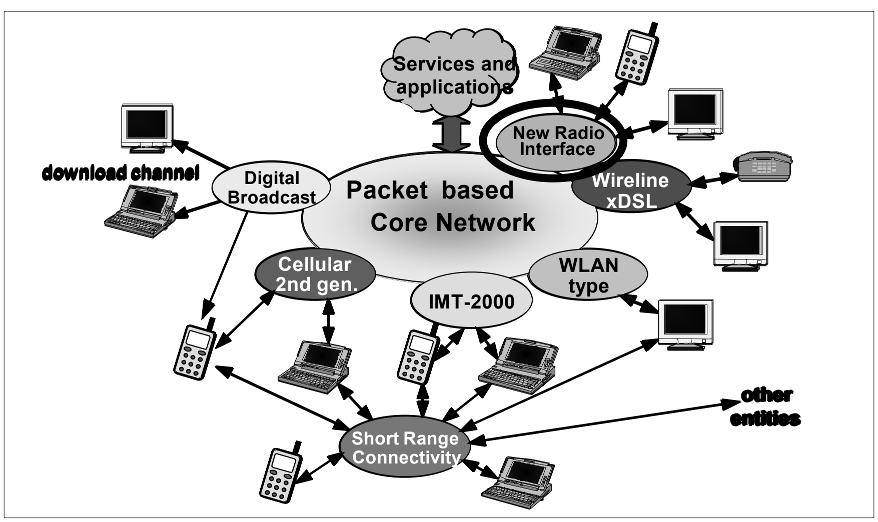
WRC-07	World Radiocommunication Conference in 2007
SG	Study Group
WP	Working Party
(J)TG	(Joint) Task Group

4 ISART 7-9 March 2006

The need for spectrum: Radiocommunication Services enable wireless telecommunication services

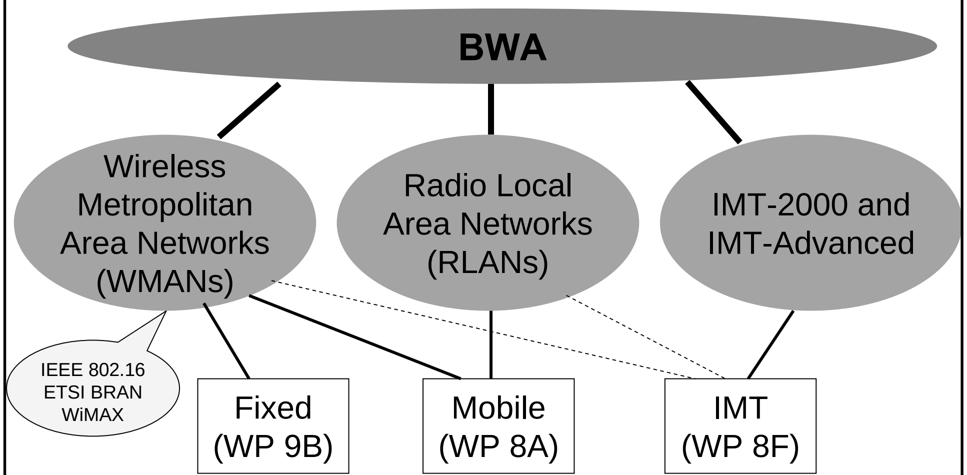


Future network of systems with a variety of access systems



Reference: Recommendation ITU-R M.1645

Broadband Wireless Access (BWA) Systems and Standards in ITU-R



7 ISART 7-9 March 2006

Wireless Metropolitan Area Networks (WMAN)

> Ongoing relationship between ITU, IEEE and ETSI to incorporate the IEEE 802.16 and ETSI BRAN BWA standards in ITU Recommendations.

- ITU-D requested assistance from the ITU-R Joint Rapporteur Group 8A-9B on access technologies for broadband communications.
- Draft new Recommendation(s) for WMANs originally developed in the Joint Rapporteur Group 8A-9B and now being continued in ITU-R Working Party 9B for the Fixed Service and ITU-R Working Party 8A for the Mobile Service.

8 ISART 7-9 March 2006

ITU-R WP 9B (Fixed Service)

> Draft New Recommendation ITU-R F.[9B/BWA] (Doc. F.9BL19):

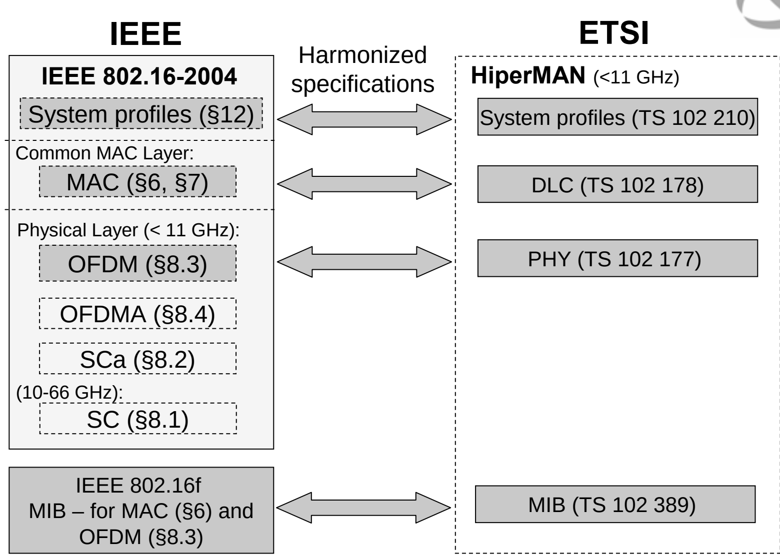
- "**Radio interface standards for broadband wireless access systems in the fixed service operating below 66 GHz**"
- Includes the harmonized IEEE WirelessMAN standards (IEEE 802.16) and ETSI HiperMAN standards (ETSI BRAN).
- The draft new recommendation was adopted by SG 9 in Dec 2005.

> In addition, a working document on technical and operational requirements is under development (Annex 9 to Doc. 9B/167).

> Next meeting of WP 9B: 27 June – 5 July 2006, Japan.

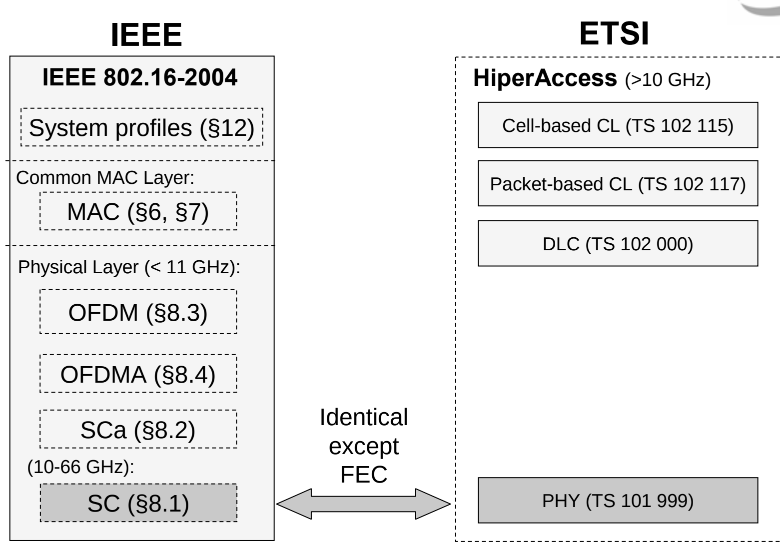
9 ISART 7-9 March 2006

Harmonized standards for below 11 GHz



10 ISART 7-9 March 2006

Standards for above 10 GHz



ITU-R WP 8A (Land Mobile Service)

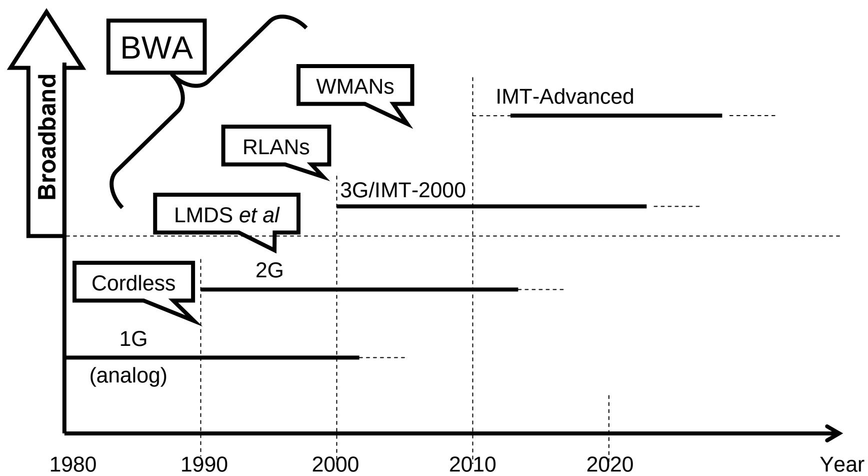
> Broadband Radio Local Area Networks (RLANs)
- Standards: Recommendation ITU-R M.1450 (further information)
- Spectrum: 83.5 MHz at 2.4 GHz and 455 MHz at 5 GHz

> Proposed draft Recommendation on "**A broadband wireless metropolitan area network radio interface standard[s] for nomadic access systems in the mobile service operating below 6 GHz**" (Annex 12 to Doc. 8A/277)

> Proposed draft Recommendation on "**Radio interface standards for broadband wireless access systems in the mobile service operating below 6 GHz**" (Annex 13 to Doc. 8A/277)

> Next meeting of WP 8A: 21-30 March 2006, Geneva

Generations of mobile wireless systems plus other radio systems



13 ISART 7-9 March 2006

Evolving Capabilities of IMT-2000 and Systems Beyond

> Goal: anytime, anywhere, anyone – the deployment of IMT-2000 systems started in the year 2000

> IMT-2000 original minimum requirements for radio technology evaluation:

- 144 kbit/s (for vehicular high speed),
- 384 kbit/s (for medium speed), and
- 2048 kbit/s (for indoor, low speed)

> Currently the standard supports up to 10 Mbit/s, further enhancements are being developed.

> Research targets for systems beyond IMT-2000 include: 100 Mbit/s for high mobility and 1 Gbit/s for low mobility, for deployment after 2010.

14 ISART 7-9 March 2006

IMT-2000 frequency spectrum requirements

> For IMT-2000, 749 MHz of spectrum have been identified:
- 806 - 960 MHz
- 1 710 - 2 025 MHz
- 2 110 - 2 200 MHz
- 2 500 - 2 690 MHz

> More spectrum may be needed for systems beyond IMT-2000 from the year 2010 onwards; this will be addressed at WRC-07 and preparations are underway in ITU-R WP 8F.

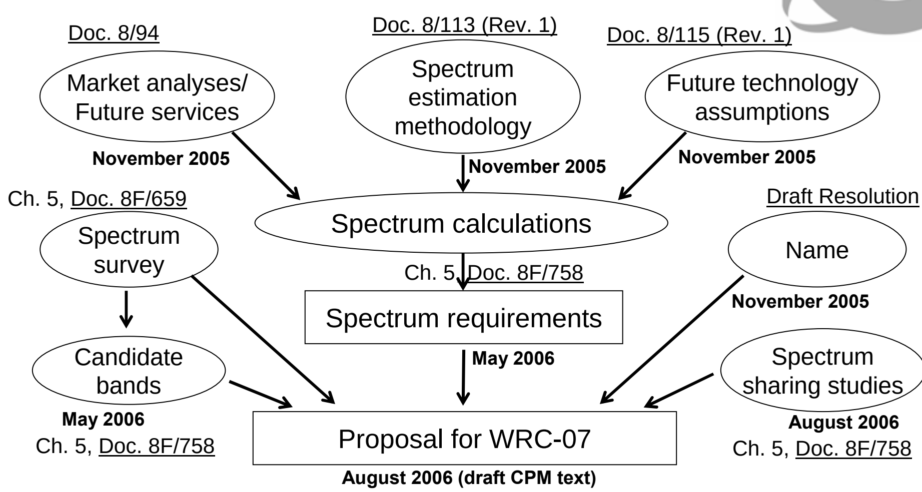
> Spectrum may need to be shared with other Services and applications, and might not all be available everywhere.

Potential candidate bands in WP 8F

> The additional potential candidate bands identified in WP 8F include:
- 410 – 430 MHz
- 450 – 470 MHz
- 470 – 806/862 MHz
- 2 300 – 2 400 MHz
- 3 400 – 3 600 MHz
- 3 600 – 4 200 MHz
- 4 400 – 5 000 MHz

> Next meeting of WP 8F: 3-10 May 2006, Biarritz, France

ITU-R WP 8F Work Plan



2007-2010: Development of standards for IMT-Advanced

17 ISART 7-9 March 2006

In conclusion…

> Broadband wireless metropolitan area networks, such as those based on IEEE and ETSI standards, together with the ongoing developments on RLANs, IMT-2000 and systems beyond IMT-2000, will lead to ubiquitous broadband wireless access.

> ITU global spectrum allocations and Recommended standards will enable integrated global systems for fixed, mobile, and nomadic broadband applications.

18 ISART 7-9 March 2006

Summary

> Have shown the organization of ITU-R with emphasis on the most significant groups in support of wireless standards and spectrum.

> Have described the spectrum activities in ITU-R, in particular the regulatory aspects of the use of the spectrum and the ongoing work to assess the spectrum requirements for IMT-2000 and IMT-Advanced.

> Have described the standardization activities in ITU-R, in particular those leading to wireless metropolitan area networks and the ongoing development of IMT-2000 and IMT-Advanced.

19

ISART 7-9 March 2006

References

> ITU Radio Regulations, 2004.
http://www.itu.int/publications/folderdetails.aspx?lang=e&folder=R-REG-RR-2004&menu=categories

> ITU Internet Reports 2003: Birth of Broadband
http://www.itu.int/osg/spu/publications/sales/birthofbroadband/

> ITU-R Wireless Access Systems Portal
http://www.itu.int/ITU-R/study-groups/was/index.html

> ITU-R Handbook on “Fixed Wireless Access”, 2001
http://www.itu.int/publications/productslist.aspx?lang=e&CategoryID=R-HDB&product=R-HDB-25

> Recommendation ITU-R M.1457, “Detailed specifications of the radio interfaces of IMT-2000”, 2003.
http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=R-REC-M.1457

> Recommendation ITU-R M.1645, “Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000”, 2003.
http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=R-REC-M.1645

> ITU Handbook on “Deployment of IMT-2000 Systems”, 2003.
http://www.itu.int/itudoc/gs/imt2000/84207.html

Contact Information:	José M. Costa NORTEL 3500 Carling Avenue Ottawa, Ontario CANADA K2H 8E9	Tel.: E-mail:	+1 613 763-7574 costa@nortel.com

20

ISART 7-9 March 2006

Broadband Spectral Sensing for Dynamic Spectrum Allocation

Robert Newgard, Principal Engineering Manager
Rockwell Collins
Christina Conway, Principal Program Manger
Rockwell Collins
Thomas Journot, Principal Electrical Engineer
Rockwell Collins
Joseph Jiacinto, Senior Electrical Engineer
Rockwell Collins
Rockwell Collins, 400 Collins Rd NE, Cedar Rapids, IA
(319) 295-5314 office, (319) 295-3751 fax ranewgar@rockwellcollins.com
(319) 295-9484 office, (319) 295-8823 fax cmconway@rockwellcollins.com
(319) 295-2790 office, (319) 295-3751 fax tjjourno@rockwellcollins.com
(319) 295-6186 office, (319) 295-5928 fax jfjiacin@rockwellcollins.com

Abstract: The focus of this work is the development of a high speed, low power, broadband spectrum sensor that will enable future military and commercial communication systems to monitor the frequency spectrum and determine where and when spectrum is unused. The long term objective is to incorporate spectral sensing into DoD systems enabling warfighters to deploy quickly anywhere in the world and have access to available spectrum without time consuming spectrum management and allocation processes. Through the use of advanced radio architectures and novel digital signal processing techniques, Rockwell Collins has developed a low power (<2.5 Watts), broadband (20 – 2500 MHz), high speed (18 GHz/sec) spectrum sensor.

I. Introduction

All one has to do is look at the National Telecommunications and Information Administration's Frequency Allocation [1] chart for the United States to quickly discern that the majority of the spectrum has been allocated. The demands for RF spectrum allocations are high, but it has been determined via field measurements that the utilization of the spectrum in some bands can be very low. As the U.S. military drives toward network centric warfare, major challenges are how to use shrinking bandwidth efficiently and how to access under-utilized spectrum. Due to the growing demands of commercial and government users for bandwidth, the military allocations for domestic and foreign uses will become strained. The spectrum is there and Rockwell Collins has developed a way to tap into it.

The first step in dynamic spectrum allocation (DSA) is to sense the RF spectrum and determine the utilization or availability of the bands of interest. In order to maximize the efficiency of the radio system, a separate sensing channel needs to be provided in the system. The key attribute of this sensor is that it must have sufficient dynamic range that the probability of detection is high enough not to interfere with other users. The sensor also must be able to scan a wide frequency band at a very high rate in order to maximize the timely determination of available spectrum. Thirdly, to enable widespread proliferation, the sensor cannot be a burden to size, weight, power and cost of any system.

Through the DARPA NeXt Generation (XG) Communications System program, Rockwell Collins was funded to develop a high performance, low power sensor.

II. Sensor Architecture

The XG Communication System sensor scans any frequency band in the 30 MHz to 2.5 GHz range and enables the XG System to communicate using available RF frequencies that have no activity.

The sensor provides high speed spectral scans for applications requiring low power and small form factor. The sensor is capable of scanning frequencies from 30 MHz to 2500 MHz at a rate up to 18 GHz/sec. The novel low power design is based on a hybrid combination of super-heterodyne and digital sampling receiver architectures. The sensor includes significant signal processing capability in the form of a high speed programmable digital signal processor (DSP), which provides the ability to download a wide range of spectral processing algorithms. The standard output of the sensor is a Fast Fourier Transform (FFT) of the RF spectrum with an instantaneous bandwidth of 16 MHz. It can also output time domain samples for further post processing in the XG system. The sensor utilizes a novel dynamically reconfigurable frequency architecture in order to eliminate internally generated spurious signals from the output.

The sensor is a digital sampling receiver which:

- Converts RF to Digital data
- Performs FFT on the digital data to determine power in distinct spectral segments (bins)
- Transfers the spectral power data to the XG Communication System
- Utilizes Ethernet 100BaseT for control and data handling

III. RF Frequency Plan

A trade-off between frequency coverage, power consumption and scan rates was necessary in order to maximize performance. The frequency plan utilizes a traditional super-heterodyne, double conversion approach, with a digital sampling receiver at the final Intermediate Frequency (IF), as shown in Figure 1. A dual IF approach was necessary to minimize spurious crossovers and maximize RF performance over the broad input frequency range (30-2500 MHz). Tradeoffs were made in the areas of spurious, size, power, commercial off the shelf parts availability, and RF performance. The architecture utilizes miniature MMICs and SAW filters, providing miniaturization of the broadband RF circuitry.

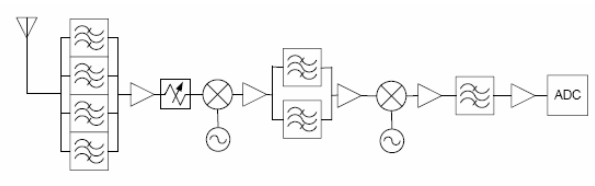


Figure 1 Sensor Block Diagram

The input form the antenna is fed into a bank of preselector filters. The main purpose of these filters is to prevent local oscillator (LO) leakage and image mixer products from corrupting the desired signal in the IF. They are wideband roofing filters, having no narrowband channelization capabilities. Note that, in special cases where it is desired to further reduce intermodulation, a narrowband preselector can be added to the sensor in order to minimize the interface from nearby strong signals (e.g., broadcast FM, TV, etc.). A wideband low noise amplifier located directly after the preselector filters minimize the impact to receiver noise figure. Following the LNA is a variable attenuator (i.e., 0 – 31 dB), allowing the sensor to be configured for optimum signal performance across the band. The signal from the attenuator is then fed into the first mixer which translates the signal to one of two IF signals. All low band signals (i.e., < 1 GHz) are up converted and all high band signals (i.e., >1 GHz) are down converted. IF signals are then fed to the second mixer and down converted to a frequency that can be directly under-sampled by the analog to digital converter (ADC). The under-sampling technique allows a low power ADC (i.e., clocked at low frequency) to digitize the RF signal, without incurring penalties of high power and expensive of high clock rate ADC's. The instantaneous bandwidth at the input to the ADC is 16 MHz.

The synthesizers for the two LOs are traditional integer N phase locked loops (PLLs). The unique feature of the second LO PLL is that it utilizes a monolithic IC which includes both the digital PLL circuitry along with the voltage controlled oscillator (VCO), enabling a small form factor and very low power. The PLLs are both designed with 1 MHz channel spacing enabling both loops to settle in less than 80 µsec, enabling very aggressive utilization of power management. The sensor architecture and signal path components were deliberately selected to balance the need for high intercept, low noise figure, and low dc power consumption.

IV. Sensor Performance

The sensor was tested in various configurations to demonstrate its capabilities to operate in numerous RF environments. The measured data, given in Figures 2 – 5, shows that the sensor performance lends itself to a broad range of military and commercial applications. A distinguishing feature of the RF performance is that it was achieved with very low power consumption (<2.5 W, with power management). The RF components were selected to maximize the performance across a broad frequency range without high levels of DC power consumption. The objective was to develop a sensor with a high enough intercept point in order to determine whether or not wide bands of spectrum are available. The overall objective was to develop a sensor capable of very low power consumption in a small form factor. The following figures illustrate the measured performance of gain and noise figure across frequency as well as IP3, and SFDR vs attenuation settings.

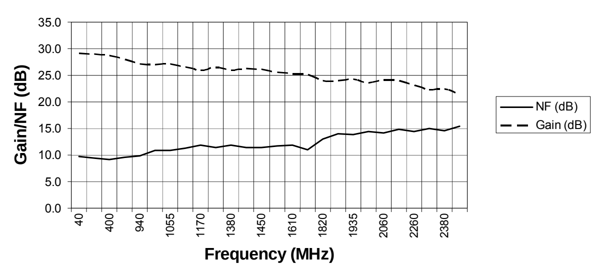


Figure 2, Gain and Noise Figure versus Frequency

Figure 2 shows front end gain of 26-28 dB, with a nominal noise figure of 10 dB across 30-1000 MHz. Above 1000 MHz the gain drops and noise figure rises due to increased front end losses in the switch matrix. Lower noise figures can be achieved with an external preamp.

Figure 3 shows the increase in noise figure vs. attenuation settings. The data was taken at 300MHz, 1050MHz and 2000MHz. The noise figure does not follow a one-to-one relationship with attenuation setting due to the fact that the attenuator follows the first amplifier.

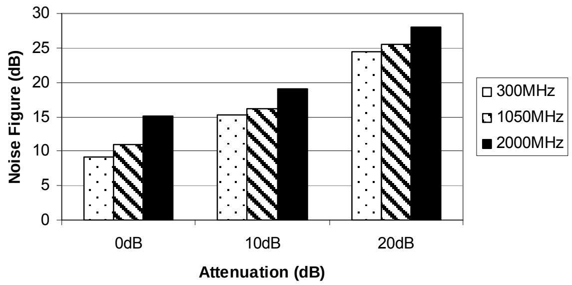


Figure 3 Noise Figure versus Attenuation Settings

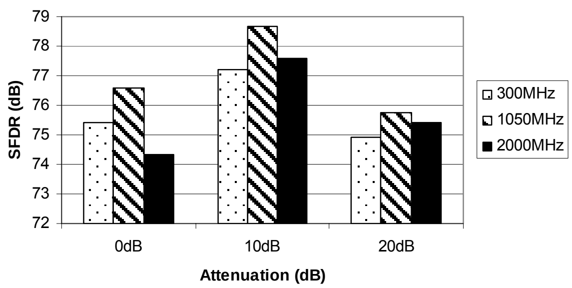


Figure 4 shows the Spurious Free Dynamic Range (SFDR) of the Sensor with the spurious reduction algorithm operating. Note that the dynamic range is not greatly affected by the attenuation setting.

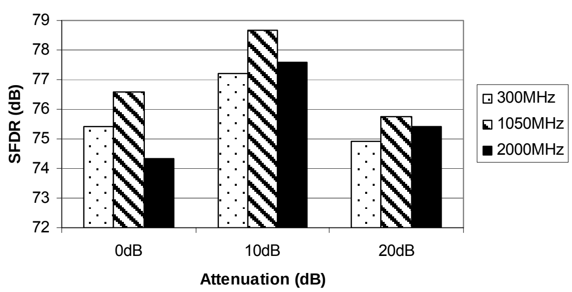


Figure 4, Spurious Free Dynamic Range

Figure 5 shows the input third order intercept (IIP3) point of the sensor for various attenuation settings. The sensor trades-off low power operation for moderate IIP3. In crowded frequency bands or in the presence of strong interferers, the higher attenuation settings are more likely to be used and those settings provided improved IIP3 performance.

The sensor provides the capability of scanning at 18GHz/sec for full band sweeps with 100 kHz resolution.

When scanning in a 25 kHz resolution bandwidth, the sensor scans at 4.8GHz/sec. The ability to scan in resolution bandwidths up to 200 kHz allows for extremely fast detection of large bands of unused spectrum. The sensor provides better than 70dB of

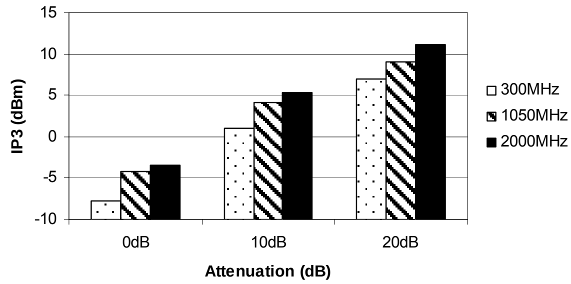


Figure 5 Input Third Order Intercept Point

SFDR relative to third order intercept and noise figure. The RCI XG Sensor offers extensive capability in the control and DSP functions. The unit provides features such as automatic gain control (AGC), DSP-based averaging, selectable spurious reduction algorithms, and advanced power management.

V. Power Management

Advanced power management is a key driver in the design of the XG sensor circuitry. This technique, referred to as "just-in-time power", has resulted in power savings as much as 60% depending on the scan speeds and resolution bandwidth settings. The DSP is configured such that it can fully characterize the turn-on and turn-off characteristics of each individual sub group (i.e., synthesizer, LNA, ADC, etc). Each element is powered only when it is needed. In this manner, synchronization of the various circuits is adaptive for changing conditions.

VI. Comparison to Commercial Spectrum Analyzers

The RCI XG sensor is compared to some commercially available spectrum analyzers in this comparison are some of the small low cost units that are intended for portable applications as well as larger lab bench test equipment. In general, the sensor provides much faster frequency scanning, at much lower power consumption than commercial spectrum analyzers. The key RF performance parameters (dynamic range, bandwidth, etc.) are very similar. Note that the sensor has a much lower noise floor, allowing it to detect very weak signals

VII. Conclusion

The sensor provides features and capabilities far superior to any commercially available spectrum analyzer. The sensor offers frequency coverage and spectral detection performance in addition to expanded DSP intelligence at a very low DC power consumption level. It can serve as a valuable asset in a large variety of commercial and military applications.

The sensor provides for selectable power settings depending on the scan speed and resolution required. The Ethernet control enables multiple users to control the one sensor, thereby enabling broad utilization of sensing assets. This combination provides for

Table 1. Included in Table 1 RCI XG Sensor versus Commercial Spectrum Analyzers

	Rockwell Collins XG	**Rohde & Schwarz FSH3**	**Anritsu MS 2711**	**Hewlett Packard 8560EC**
Sweep time (25kHz RBW)	14 msec (w/ 60 MHz span)	20ms-1000s No info on speed/RBW	6.5s Full span No info on speed/RBW	~200 msec (w/ 60 MHz span)
Frequency Coverage	30 – 2500 MHz	0.1 – 3000MHz	0.1 – 3000MHz	30 Hz – 2.9 GHz
Noise Floor (25kHz)	-114 to -117dBm	-75dBm	-	-107 dBm
Dynamic Range (P1dB-NF)	100 to 107dB	-	-	103 dB
SFDR (IP3)	72 to 75dB (IP3 = -5)	60dBm (IP3=+13)	-	65dB (IP3 = +11)
Volume	13.6 in^3	35.8 in^3	168 in^3	1773 in^3
Power	<2.5W	< 7W	13.75W	180W

exceptional overall capability that is extensible to application beyond dynamic spectrum allocation (e.g., situational awareness, spectrum mapping, low power ELINT, etc).

VIII. Acknowledgement

This work was performed under the funding of DARPA ATO office (FA8750-04-C-0216). The authors would like to thank DARPA PM Preston Marshal for his guidance and vision of the future of DSA and the XG System. The authors also acknowledge the contributions of Rockwell Collins Sensor Team.

IX. Bibliography

[1] NTSA
http://www.ntia.doc.gov/osmhome/allochrt.html

Metrics-based Regulation of Effectiveness and Efficiency in Dynamic Spectrum Access Systems; the Art and Science of Dealing with Radio Complexity

Kalle R. Kontson
Alion Science and Technology
Phone: 301-918-1531 Fax: 301-918-1544 email: kkontson@alionscience.com
J. Michael O'Hehir
Alion Science and Technology
Phone: 410-573-7467 Fax: 410-573-7137 email: mohehir@alionscience.com

Abstract: Potential regulatory structures for the introduction of dynamic spectrum access (DSA) in advanced radio technologies are examined to support an argument for a new way of evaluating systems. The qualities of effectiveness and efficiency in spectrum access are decomposed into multiple technical features affecting spectrum access. These features can be used to associate and aggregate the results of separate evaluations of each feature to produce overall scores for advanced radio technologies. Termed the "Spectrum Scorecard", this approach is being researched and developed at Alion to provide a method for addressing the complexity and interdependence of multiple features in advanced radio technologies employing DSA, and to serve as an adjunct to regulatory approaches where system qualities determine spectrum access rights.

1. Introduction

The diversity of emerging spectrum technologies has confused the concept of spectrum efficiency and effectiveness. With the increased use of digital modulations, multiple access schemes, and mobile ad hoc networks (MANETs), spectrum efficiency has become a complex notion. Dynamic Spectrum Access (DSA) systems have been suggested as a fundamental, technology-enabled method to make more effective and efficient use of scarce available spectrum. However, practical methods for designing, developing and managing DSA remain in the formative stages.

The basic intent of efficient spectrum use is captured in the National Telecommunications and Information Administration's (NTIA) Manual of Regulations and Procedures for Federal Radio Frequency Management[1], stating that Federal users should "*... employ up-to-date spectrum conserving techniques as a matter of normal procedure ...".* The same principle applies to commercial, state and local use regulated by the Federal Communications Commission (FCC).

Technologies such as adaptive array antennas, multiple-input-multiple-output, advanced modulation techniques, and power control actually improve spectrum efficiency and reuse. These features enhance the portfolio of options that empower DSA devices. Such emerging technology features also introduce a significant level of complexity. Some Department of Defense (DoD) military requirements inherently increase the amount of spectrum required. To add to the complexity, spectrum management must balance operational requirements (e.g., anti-jamming, low probability of intercept) with spectrum efficiency to meet national interests. Evaluation and regulation of spectrum efficiency and effectiveness can no longer be effectively accomplished by existing spectrum certification and license management processes.

2. A Futurist View of Regulation

In 2002, a concept for an Intelligent Wireless Device Bill of Rights was introduced at the FCC Technological Advisory Council.[2] It was introduced as a hypothetical, top-level, legal framework for accommodating future dynamic spectrum access, where open access to the spectrum on an as-needed basis is the dominant mode of spectrum access protected by the law of the land. The draft of this Wireless Bill of Rights (WBoR) is presented below.

Article 1: *Any intelligent wireless device may, on a non-interference basis, use any frequency, frequencies or bandwidth, at any time, to perform its function.*

- **Article 1, Tenet 1:** *To exercise rights under this Article, intelligent devices must be mentally competent to accurately determine the possibility of interference that may result from their use of the spectrum, and have the moral character to not do so if that possibility might infringe on the rights of other users.*
- **Article 1, Tenet 2:** *To exercise rights under this Article, intelligent devices must actively use the wireless spectrum within the minimum time, spatial and bandwidth constraints necessary to accomplish the function. Squatting on spectrum is strictly prohibited.*

Article 2: *All users of the spectrum shall have the right to operate without harmful electromagnetic interference from other users.*

- **Article 2, Tenet 1:** *Priority of rights under this Article may be determined by the proper* *authorities only in cases of National emergency, safety of life or situations of extreme public interest.*

- **Article 2, Tenet 2:** *Rights under this Article may be exercised only when the systems exercising the rights are designed, as determined by the state of the practice, to be reasonably resistant in interference.*

Article 3: *All licensing, auctioning, selling or otherwise disposition of the rights to frequencies and spectrum usage shall be subordinate to, and controlled by Articles 1 and 2, above.*

Since its original publication, the draft of the WBoR has been discussed, analyzed and evaluated[3]. It suggests an unlicensed, but highly regulated spectrum management regime wherein the right to access spectrum on an as-needed basis is balanced with the right to be protected from interference by other users. These competing rights are predicated on what might be termed a standard for "reasonable radio behavior".

An in-depth discussion of the sweeping policy and economic issues implied by the WBoR is not addressed in this paper. However, there has been sufficient interest in the community to suggest that this model for spectrum regulation may serve as a long-term objective, where the future end-state is envisioned to be dominated by a regulatory model for self-managing, cooperating systems that have been deemed to be worthy of open spectrum access as a result of having been certified for good "Spectrum Citizenship"; i.e., meeting a reasonable radio behavior standard. Command & control models would be in the minority, operating as a special-case exception to the general law of the land. The dominant rule would be similar to open access to highways enjoyed by citizens, where the burden of assuring safety and minimizing collisions is placed on regulation of mechanical qualities of vehicles and competence of operators. Once good highway citizenship is certified and access authorized for both the vehicle and operator, drivers may drive on any road at any time, with few exceptions.

Entering this brave new world defined by the WBoR is not going to be an instantaneous change that can be introduced by flipping a regulatory switch. Maintaining stability in the processes, technologies, regulations, business interests and security interests of spectrum management for both the commercial sector and the government sector (especially public safety and DoD) is a compelling argument for resisting change, or moving toward change at a slow and measured pace. However, the argument for change, in particular toward a more open "spectrum sharing" environment, is supported by a growing constituency. The perceived payoff is far better effectiveness and efficiency in utilizing the spectrum to satisfy our rapidly growing appetite for constant and pervasive broadband connectivity with our surroundings, whether through communicating, sensing, or both, using the RF spectrum. The work of the FCC Spectrum Policy Task Force[4] (SPTF) was a leading effort to define that change. Other efforts have begun to unravel the legal and public policy implications of undergoing that change[5]. The first IEEE conference dedicated to dynamic spectrum access networks, IEEE DySPAN 2005, was held in November of last year, and was a resounding success.

2.1 Structuring a Good Spectrum Citizenship Incentive Program for Spectrum-dependent Systems

A common thread in the WBoR and other concepts advocating migration to an open access structure is the dependence on technologically advanced systems to automate and distribute the frequency/channel selection process and make judicious use of the spectrum. In essence, these principles advocate empowering radio devices with the authority and responsibility to administer spectrum management. Furthermore, these devices are presupposed to meet certain certifiable standards for intelligence; or at least reasonable behavior, if not intelligence.

Designing and manufacturing such devices is not easy or cheap. In addition, some of the more sophisticated behaviors envisioned for the future involve environmentally aware and cooperative systems. That can lead to the need for some sort of standards, but such standards should not stifle innovation and flexibility in implementation technology. Regulations and standards that are contemplated to encourage and accelerate our march toward the WBoR must have two qualities: they must provide incentives for investing in advanced technology that enables good Spectrum Citizenship, and they must provide a framework for consistent evaluation of certifiable attributes without dictating use of specific technologies.

In July, 2005, the FCC TAC organized and entertained a session on a concept called tiered spectrum access rights[6]. This concept advocates the use of increased access to spectrum as an incentive to produce and deploy more sophisticated, spectrum-friendly equipment. Simply stated, the idea is to implement a regulatory model that provides proportionately higher tiers of spectrum usage privileges to those systems that can be certified to meet correspondingly higher performance requirements for spectrum-friendly behaviors. If applied widely, the assertion is that such a model offers the following advantages:

- Promotes superior technologies by offering incentives for use
- Expands capacity of the limited resource
- Begins an industry-driven migration toward new regulatory models, such as the WBoR

There is precedent for such regulations; the idea is not new. For example, exceptions to restrictions on High Occupancy Vehicle (HOV) lane access for hybrid and electric cars is offered in many states. FCC Part 15 allows for higher EIRP for directional antennas in some bands. And, closest to being on point, FCC Part 15 offers expanded spectrum access for systems with Transmitter Power Control (TPC) and Dynamic Frequency Selection (DFS).

However, the examples above are the exceptions, rather than the rule. The tiered spectrum access regulatory model would, as a matter of policy, successively apply this principle as a general rule to virtually all spectrum being reallocated, or licensed, or auctioned. The goal would be to convert each band, band by band, to operate under these tiered access principles, as a mechanism for motivating and sustaining a migration toward the WBoR while constantly applying pressure with incentives for advancing spectrum efficient technologies.

Successful application of such a tiered spectrum access rights model requires some regulatory intervention and oversight. To implement this scheme, three difficult issues need to be addressed:

- **Defining variable spectrum rights:** What additional rights can be offered in response to well-designed wireless systems? Access to additional bands? Permission for wider bandwidths? More power? Expanded geographic coverage?

- **Reward-worthy equipment characteristics:** What wireless system characteristics are sufficiently beneficial, measurable and predictable to justify expanded spectrum access rights? What are the measurable earmarks of good Spectrum Citizenship?

- **Implementation:** How do we measure spectrum qualities? How do we manage and adjudicate coexistence? What regulations need to change to implement such a model? Equipment authorization? Certification? Licensing?

The three categories of issues are discussed in the remainder of the paper. The intent is to focus mainly on defining metrics and measures of good Spectrum Citizenship, and how those metrics and methods of measuring relate to implementation. The treatment is intended to apply to both commercial and public safety systems regulated by the FCC and Federal government systems regulated by the NTIA, in particular the DoD. The underlying principle that has guided the FCC rulemaking in recent past has been to apply non-prescriptive, technology-agnostic rules to provide incentives to the market, and couple success to financials. The DoD and public service/safety perspective is different; the goal is to guarantee effective performance when needed, which is usually only a fraction of the time (fortunately).

3. Defining Tiered Spectrum Access Rights

In order to apply tiers of spectrum access rights as an incentive, the incrementally increased rights at each tier must be manageable, and be perceived to add sufficient value to users to justify investment in technologies needed to qualify for that tier.

As a start, the three commonly accepted degrees of freedom (or dimensions) in spectrum management can be defined: frequency, time and space. These three, plus an angular directional dimension, were recently defined by Bob Matheson as comprising the "electrospace" [7]. Thus, spectrum access rights can be viewed as dealing with the right to occupy a smaller or larger volume of the electrospace. This view of the trade space of spectrum usage was also captured by the FCC Spectrum Policy Task Force:

> "The Task Force also analyzed the benefits of parceling out spectrum using variations in frequency, space, power, and time to maximize the use of spectrum. In the past, the Commission has recognized and licensed spectrum primarily by defining spectrum rights in terms of the first three dimensions. The Task Force found that new technological developments are changing the way in which each of these spectrum dimensions is used. In addition, new technology now permit(s) the Commission to increasingly consider the use of time, in combination with frequency, power, and space, as an added dimension that could permit more dynamic allocation and assignment of spectrum usage rights." [Ref 4, page 19]

As discussed in Reference 7, there are practical problems with defining and dealing with a finite number of possibilities in the electrospace. To expand the trade space to include those technical characteristics that most interest and influence spectrum users, a subset of one or more of the electrospace dimensions must be put on the bargaining table. For example, power limits, contiguous bandwidth limits, and overall non-contiguous bandwidth limits will have a direct bearing on the perceived value of rights at a spectrum access tier. Other performance-related features that have a more subtle connection to spectrum access rights may be of paramount importance to make or break the business case for a user. Further, rights must be enforceable to have practical value. These subtleties were addressed by Dale Hatfield and Vanu Bose in the FCC TAC meeting of July, 2005 (see Reference 6). For example, the expanded rights to use angular segments may be counter to the need to minimize weight, size and power consumption in portable devices that cannot practically be designed for use with directional antennas.

To attract commitment to better technologies at higher tiers, the rights offered as incentive must be simply and undeniably more attractive and valuable than those of lower tiers. So, perhaps we end up exactly where we started; that is, as a first cut the tiers of rights should be comprised of simple, physically defined rights: additional spectrum, increased authorized power, additional contiguous bandwidth, and expanded geographic coverage. In addition, when reallocating spectrum where the existing regulations restrict operations to specific services (such as TV), the defined rights in each tier should also include expanded service rules contingent on deployment of spectrum friendly technologies.

4. The Spectrum Scorecard – Quantifying Effective and Efficient Spectrum Access

In the old regime, regulations were driven by "service rules"; allocation and authorization for transmission were predicated on limiting transmission to only one kind of service, such as FM radio broadcast. Technical limitations of analog transmission played a major role in these regulatory trappings by limiting flexibility of analog transmissions to those services for which a particular hardware configuration was designed. TV transmitters could only send TV signals and thus provided only TV services. It is not hard to imagine that regulations and service rules naturally evolved to mirror the constraints of such old technology.

Under the inflexible and compartmentalized technical and regulatory structures of our fathers' radio world, addressing spectrum efficiency was simpler. A specific service matched with standardized modulation and license-based geographic coverage. Addressing spectrum utilization was like counting and managing use of bricks; they were standard bricks and they did not change in size, weight and shape. One could determine the capacity of a truck hauling bricks with a very simple calculation. The truck-load of bricks did not suddenly become aware of a change in the mission and endeavor to become cinderblocks.

Now when we compare the relative simplicity of the good old days with what is being presented at conferences during the past few years, a far more complex scenario presents itself. The topics presented at last year's ISART[8] and DySPAN[9] Conferences send a completely different picture.

Under work being done for the DoD, Alion Science and Technology has begun developing a structured process for defining metrics and measures of performance for complex future radio systems. The initiative has been named the Spectrum Scorecard and attempts to provide a relative measure of "Spectrum Citizenship" for spectrum-dependent devices. Spectrum Citizenship, simply put, is an indication of how effectively a system can access and leverage spectrum to successfully fulfill a given set of objectives without using more than its fair share of the spectrum resource, or infringing on other's rights. The Spectrum Scorecard is still under development at the time this paper was assembled. The focus of the work has been on communications systems, although the concept can apply to any RF devices.

The initial scorecard work evolved from a conceptual treatment of multiple features of emerging radio technology. We tried to put certain attributes on a common graphical scale giving visual representations of aggregate, system-level "spectrum goodness", while looking at the relative contributions of each attribute. Figure 1 shows a first instantiation.

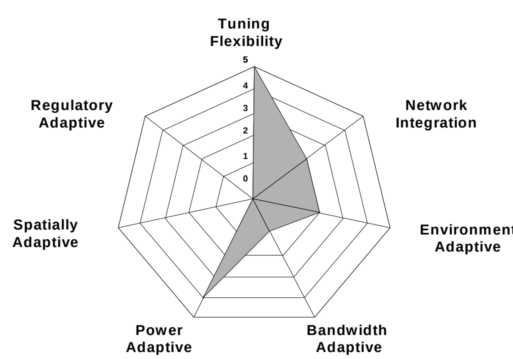


Figure 1 – Initial Concept of the Spectrum Scorecard

The figure above was a first attempt at expressing the complexity and interrelated dimensions of Spectrum Citizenship. The first cut shown in Figure 1 suffered simultaneously from insufficient detail, mixing different types of attributes, and being too complex. Our challenge was to try to reduce the set of metrics to a manageable number of things that really matter to spectrum use, and still preserve the ability to represent the effect of a complex set of features. We refined and redefined our view of spectrum-relevant technology trends in four major categories:

- RF characteristics (physical) domain
- Bits and information domain
- Network management domain
- Cognitive domain

This categorization is one of many possible views, and is not offered here as the recommended exclusive view, but is used to add some manageable structure to the analysis of complexity. The role these domains play in a manageable structure for Spectrum Citizenship is shown the bar graph in Figure 2.

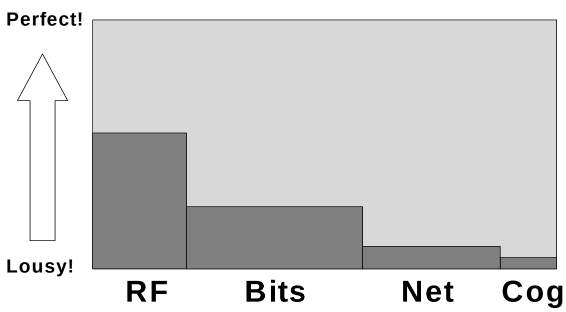


Figure 2 – Notional Spectrum Citizenship Scorecard

The vertical scale is meant to be a normalized measure of spectrum effectiveness/efficiency. Our objective is to define domains that are able to be scored in some way that allows an easy, comparative assessment of a system's overall Spectrum Citizenship. It also supports a first glance assessment of where system improvements are possible by isolating strong and weak aspects. Further, the depiction in the figure allows for different weighting of the impact of each domain by allowing the width of each bar to be different, and defining the overall Spectrum Citizenship score in terms of a percentage of the surface of the graph's pallet that is covered, in aggregate, by the bars. Thus, the less total open space above the bars, the better the overall score.

The domains are each comprised of a number of features. For example, the RF characteristics domain includes familiar items like modulation, power, receiver selectivity, antenna performance. These features are not necessarily exclusive to only one domain, but may appear as players in more than one domain. Thus, there are interrelationships between domains and features.

Conceptually, this seems an appealing depiction. If there was such an agreeable standard method for depicting Spectrum Citizenship score, the ability to adjudicate tiered spectrum access rights may be plausibly enabled. But, the devil is in the details. What does the vertical scale really mean? How can it be normalized? What are the features and metrics that comprise each domain and how are they scored? Are these the right domain definitions? Is this the right method for depicting designs worthy of reward?

Alion's ongoing efforts are focused on answering these and other questions; if they can be answered at all. It is still a work in progress. Our definitions of domains and features are still evolving. In the following sections we describe the domains and illustrate some of their features in terms of Spectrum Citizenship.

4.1 RF Characteristics (Physical) Domain

The concept of a spectrum space similar to that defined by Bob Matheson (Reference 7) is helpful in the RF physical domain. However, to actually score systems, specific characteristics need to be addressed.

A device with a high Spectrum Citizenship score should have sufficient flexibility to adapt to any mix of challenges and provide the best possible solution. Flexibility is dependent on both the availability and usability of the RF trade space. Thus, flexibility in features is a key component in spectrum usage effectiveness. Multiple features contribute to a system's Spectrum Citizenship, including frequency agility, bandwidth management, modulation, power management, and antenna management.

4.1.1 Frequency Agility

Frequency agility improves the supportability of a system. Flexibility in frequency tuning range will allow a system to be accommodated in available allocated bands. Additionally, using unused frequencies from a frequency resource list improves the overall spectrum efficiency of a system. A frequency agility metric for systems must measure the ability to operate over a wide range of frequencies. To score well, a system should also readily adapt to as many available allocated bands as possible. Emerging systems will have the capability to operate in a very broad range of tuned frequencies, covering more than an octave or even more than a decade of frequency range. As an objective for achieving the highest score, systems should operate contiguously over the broadest frequency RF range possible.

Spectrum-dependent devices that share common frequency resource pools can increase the overall duty cycle of spectrum use. This is accomplished best by sensing or in some other way assessing the environment and selecting the instantaneous frequency of use from the common frequency resource pool. Examples of systems exhibiting this attribute include

High-Frequency - Automatic Link Establishment (HF-ALE), trunked radio systems, cellular telephones, and IEEE 802.11 devices. To score minimally in frequency agility, systems must also be able select the optimal operating frequencies from a common frequency resource pool. To score well, the objective systems will select optimal operating frequencies based on the local electromagnetic environment and the prevailing regulatory environment. The Defense Advanced Research Project Agency (DARPA) neXt generation (XG) project is experimenting with instantaneous frequency selection based on sensing and policy-based regulatory rule sets. XG provides a benchmark for a high-scoring system in frequency agility.

4.1.2 Bandwidth Management

A spectrum-dependent system should adjust its bandwidth based on deployment conditions. Conditions that could require adjustment in bandwidth might include LPI/LPD, dynamic channel assignment, power restrictions, information rate, interference, etc. Bandwidth management can reduce the need for frequency resources when conditions allow and thereby increase the opportunities for reuse. Fixed bandwidth systems that do not scale RF bandwidth according to prevailing conditions would score low. To score acceptably, spectrum-dependent systems should have some method to reconfigure bandwidth. In addition, acceptable devices must conform to limits on unintentional emissions (spurious/harmonic) so that the systems do not inadvertently deny bandwidth to other uses. To score well in this feature spectrum-dependent systems will dynamically adapt their RF bandwidth based on prevailing conditions.

4.1.3 Modulation Management

Modulation management includes both the traditional consideration of modulation efficiency and the trend to support adaptive capability. Modulation is important in the determination of spectrum requirements and efficiency. Numerous modulation strategies exist for communications systems with varying efficiencies. To score acceptably, systems should exhibit reasonably efficient modulation in terms of bits/sec/Hz. In this case, numeric values for acceptable and superior performance may be appropriate. For example, 0.8 bits/sec/Hz may be a threshold for acceptable scoring. More advanced capabilities using multi-state digital modulations may get better scores, albeit at the cost of additional power. An objective target would be set for high-scoring systems to exhibit a high degree of modulation efficiency, such as 4 or more bits/sec/Hz.

Adaptive modulation offers another degree of flexibility. For example, current research in the use of non-contiguous modulations may produce systems with large virtual bandwidths, composed of multiple narrow bandwidths. While few current technologies will provide the ability to use non-contiguous bandwidth, some scoring credit should be assigned to systems that include non-contiguous bandwidth modulation capabilities or other adaptive modulation technology beneficial to spectrum efficiency.

A previous study has identified over 140 attributes that are regulated by spectrum management worldwide. Not all attributes are regulated for all allocated frequency bands. The ability of future waveforms to adapt to specific regulatory constraints is essential to improve accommodation by national administrations. Some scoring of the ability to flexibly adapt to these regulatory requirements for modulation standards should be applied. To be acceptable, systems will be reconfigurable to adapt to regulatory requirements. To score well, objective systems will automatically adapt their modulation and waveform characteristics to prevailing local regulatory constraints.

4.1.4 Power Management

In the mobile tactical environment, radio link distances vary on a minute-to-minute basis. Using the minimum power required to maintain an operationally acceptable link will reduce the effective spectrum footprint and promote frequency reuse within the operational environment. To score well, systems will have power management capability to minimize their spectrum footprint consistent with operational requirements.

4.1.5 Antenna Management

Advances in antenna technologies offer many operational benefits such as cancellation of interference, elimination of multi-path effects, increase in frequency reuse, and reduction of probability of detection. Adaptive antennas can be used to form null patterns in the direction of interfering signals. This capability is a significant improvement in overall spectrum reuse and operational effectiveness. To be acceptable, a RF device will incorporate adaptive antennas to reduce the effect of interference. Adaptive (diversity processing) array antennas can also be designed to exhibit another degree of flexibility relevant to Spectrum Citizenship. The same adaptive array processing may be used to either minimize the effect of a multi-path environment or exploit a multi-path environment to implement multiple spatial channels on the same frequency; i.e., multiple-input,-multiple-output (MIMO). Newer variants of the 802.11 standards-based WLANs are already available that do some of this. There is very high potential value of such technology in terms of spectrum efficiency and effectiveness. Thus, to score well, spectrum-dependent systems will incorporate adaptive array antennas to improve channel capacity performance in a multi-path environment as well as eliminate interference. Here, as with modulation, a numeric value can be assigned in terms of overall gain in capacity, or spectrum reuse.

4.2 Bits and Information Domain

Some features are a direct result of convergence of services made possible by reducing all forms of communications to digitized bits and packets. These are essentially the lowest order life-forms of information sent through communications channels, and they can be used to build any higher form of life. Just as bricks and cinderblocks form houses, bridges, wells, office buildings, fortification walls and fences that ultimately comprise our towns and cities, so do bits and packets build voice, video, and data that further explode to become infinitely variable forms of services in the application domain. The relevance to spectrum is tied to the protocols and technologies that determine how bits and packets are sent.

Decisions made at the application layer of a network have huge impact on spectrum use. For example, do I send full motion video or does a still shot serve the function for providing positive ID on a terror suspect? Or, is Voice Over IP (VOIP) more efficient than analog voice or the codec-based digitized voice transmission techniques perfected in the cellular industry? The comparison of these alternatives is relevant when reduced to simple functional performance; e.g., words successfully sent and understood for every kHz of RF transmission. The convergence of services and digitization of our information puts that choice in play in a much more relevant manner than in a circuit-switched, analog system, where one choice of function matches up to only one choice for channel use.

At this time, we are in the early stages of relating features in this domain to a Spectrum Citizenship score. Features of this domain are such interrelated things as information transfer, payload structure, data/symbol structure, protocols and applications. Choices for features such as error correction, IP header compression, and application-level bandwidth management can influence the spectrum requirements, and would influence the score.

4.3 Network Management Domain

This domain is closely related to, and overlapping with the preceding domain. Networking efficiency can have a substantial impact on the amount of spectrum a system requires. Over 90% of the total bandwidth can be consumed by networking overhead information. Alternative protocols can improve this situation by a factor of ten thereby decreasing the amount of spectrum required by a similar factor. Additionally, a tighter integration between network and spectrum management systems is required to improve flexibility and responsiveness to operational requirements.

4.3.1 Network Overhead

The migration from wired networks to MANETS exposes shortcomings and additional requirements relative to traditional wired network protocols. For example, packet collisions in wireless networks occur when two or more nodes transmit simultaneously. Various media access control (MAC) protocols attempt to resolve this problem in networking. Failure to have an effective MAC protocol will cause reduced throughput and the retransmission of unacknowledged packets. To achieve an acceptable Spectrum Citizenship score, systems will have an effective and efficient MAC protocol strategy.

The Transport Control Protocol (TCP) guarantees end-to-end packet delivery. This requires an acknowledgement from the destination node for every packet sent. Without that acknowledgement, the packet is retransmitted. Additionally, if packets are acknowledged out of order, the packets will be retransmitted. The efficiency of basic TCP in a wireless multi-hop environment has been shown to be less than 25%. TCP strategy studies[10] have shown alternative TCP protocols to be more than twice as efficient for the same WAN environment. To achieve an acceptable Spectrum Citizenship score, systems will have an efficient transport control strategy.

4.3.2 Topology Management

Maintaining network connectivity in a self-forming, self-healing MANET requires robust routing strategies. The network overhead and resulting spectrum bandwidth required to accomplish MANET routing varies by protocol strategy and needs to be minimized.

Wireless networks exhibit less connection stability than wired networks. This puts an additional strain on the routing protocols used to maintain routing tables (for a table driven strategy) or discovery of routes (in an ad hoc strategy). Overhead differences of various routing strategies can be significant. The control packets to perform routing functions should be minimized to reduce the overall spectrum footprint. Systems will minimize the overhead associated with network routing functions to achieve acceptable scores.

However, investment in certain overhead intensive additional routing complexity can have overwhelmingly beneficial effects on spectrum efficiency. For example, the integration of adaptive array antennas into the objective environment will require close integration with an optimal network routing strategy. While this increases the amount of protocol complexity, the improvement in throughput is significant and can be used to reduce the amount of spectrum required. So, objective spectrum-dependent systems that integrate directional antenna spatial domain processing into an optimal network routing and MAC strategies may score higher, notwithstanding increased overhead demands.

4.3.3 Bandwidth Availability Management

The availability of spectrum to support the objective environment will impose bandwidth limitations. These limitations must be integrated with access priority and quality of service strategies. End-to-end quality of service (QoS) will have to consider spectrum bandwidth limitations of each traversed link in determining guaranteed delivery. Alternatives to the open shortest path first (OSPF) protocol may be required due to bandwidth limitations. The OSPF protocol is one form of QoS strategy determined by the number of nodes traversed. In wireless networks, the shortest path may not have the effective bandwidth to support the required QoS due to link interference. So high scoring systems should integrate bandwidth availability into QoS strategies.

The interface between network management and spectrum management support systems must provide a seamless provisioning of frequency resources. Ultimately, fully decentralized, autonomous spectrum operations will support determination of spectrum access based on a combination of integrated network management capabilities that simultaneously consider sensed local electromagnetic environment, coordinated knowledge gathered from other cooperating spectrum citizens and prevailing policy/regulatory constraints. These concepts are portrayed in the Global Electromagnetic Spectrum Information System (GEMSIS) Initial Capabilities Document[11] (ICD) and Concept of Operations[12] being pursued by the DoD.

The message here is that the integration of emerging spectrum management processes and technologies into the network management of a wireless communication system is a prerequisite to achieving good Spectrum Citizenship. To score at an acceptable level, systems must have network management systems that provide seamless transfer of spectrum access requirements and frequency resources. To score well, systems will determine spectrum access based on "spectrum situational awareness": sensed local electromagnetic environment, coordinated exchange of spectrum knowledge with other spectrum citizens, and automated rules based on policy/regulatory constraints.

4.4 Cognitive Domain:

The attributes of this domain are extremely difficult to measure, and each feature is dependent on other system features described in the preceding domains. For example, some knowledge and awareness of the surrounding spectrum use is dependent on receiver sensing capability and/or some form of network-supported spectrum information exchange. While there are many attributes of the "smart radio" or "cognitive radio" that may prove very beneficial to effective spectrum use, it is not a prerequisite to cognitive domain functions that the devices are cognitive radios. Much of the advantage for cognitive spectrum use can be implemented by having a smart network management system controlling some pretty dumb radios. The functions of network management and individual radio behaviors are intertwined. Spectrum Citizenship in the cognitive domain is discussed in terms of two features: geographic (spatial) awareness, and environmental awareness.

4.4.1 Geographical Awareness

Geographical awareness will be important in determining the prevailing regulatory environment and optimizing network performance through the exchange of location information. Knowledge of the node's location is essential in determining the prevailing regulatory environment as well as for directional networking. Within recent years, regulatory constraints have moved beyond uniform allocation within national borders to specification of allocation to specific geographical areas within national borders. Also, as is discussed by the FCC in recent rulemaking proceedings involving cognitive radio, in receive-only operations where systems do not transmit a signal, location technology may be an appropriate method of avoiding interference because sensing technology would not be able to identify the locations of nearby receivers. To achieve an acceptable score, a system will have knowledge of its own current location.

The exchange of location information between cooperating nodes supports the use of directional antennas and reduced network routing overhead. Both these attributes reduce the amount of spectrum required to support operations. Use of directional antennas requires knowing in what direction to form the beams or nulls. With the implementation of MANETS the exchange of location information will reduce the setup time to form the beams and nulls. Knowledge acquired from cooperating nodes will also improve network routing strategies. Thus, systems that can exchange location information will score well. Such systems would be required to integrate location information into antenna, frequency assignment, and network protocols to optimize spectrum effectiveness and minimize the spectrum footprint imposed on the spectrum neighborhood.

4.4.2 Environmental Awareness

To improve frequency reuse and efficiency, some existing systems access a shared set of frequencies by monitoring the traffic on specific frequencies and accessing those channels currently not in use. Examples of this include HF-ALE, cellular and trunked radios. To achieve acceptable Spectrum Citizenship scores, spectrum-dependent systems will monitor frequencies in common frequency resource lists. To accomplish autonomous operations, centralized allotment and distribution of resources must be replaced with completely decentralized, ad hoc access to the spectrum. To score well, objective spectrum-dependent systems will monitor their entire range of operating frequencies to determine unused spectrum.

In the objective future environment, not all spectrum-dependent devices in the spectrum neighborhood will be collaborating nodes. A potentially advantageous cognitive function is the ability to detect and categorize the non-collaborating systems. This capability will aid in determining potential frequency sharing opportunities. The identification of the modulation associated with observed signals enables the determination of potential frequency sharing opportunities and potentially increasing frequency reuse. Emerging cognitive technology in this area includes algorithm-based modulation identification techniques. Thus, objective spectrum-dependent systems should be able to determine the modulation of observed signals and invoke optimal Spectrum Citizenship decision-making algorithms in response.

Sensing at only one node does not mitigate the risk of interference to undetected receivers (the hidden node). The exchange of observed spectrum use between nodes, including receivers, reduces the potential to cause interference within the environment. This ad hoc collaborative behavior will demand establishment of certain standards to govern information exchange formats and content. However, the potential payoff may contribute greatly to Spectrum Citizenship. Objective spectrum-dependent systems will exchange environmental information to reduce the potential for interference to score well. Such coordinated spectrum awareness among citizens within a common spectrum neighborhood is envisioned and described in the FCC Rulemaking proceedings on Cognitive Radio.

5. Implementation

The Spectrum Scorecard is suggested as a method for evaluating new systems against standard Spectrum Citizenship metrics. Its viability will depend on the ability to evaluate programs in terms of their overall spectrum efficiency and effectiveness, and the implementation of policy changes necessary to mandate spectrum consideration in implementing a tiered spectrum access rights regulatory scheme.

To date, over 30 features related to communications have been identified by Alion. Obviously, not all have been discussed here. Each affects the spectrum required to operate a system, amount of frequency reuse that can be obtained by a system, or the overall efficiency and effectiveness of the system. The Spectrum Scorecard features can be evaluated and optimized individually, however there are interrelationships that must also be considered. For example, the adoption of directional antennas will improve a system's potential frequency reuse, and also affects network routing strategy by increasing the overhead burden with directional routing.

To achieve the relatively simple roll-up of scoring depicted in Figure 2, the detailed features, attributes and complex interrelationships for each domain (discussed in section 4) must be evaluated and aggregated to provide a domain score. A suggested method is to use the same graphical representation to score and aggregate the scoring impact for each feature within a domain, as depicted in Figure 3. Scores for the features are derived by defining a Figure of Merit and Measure of Performance scores (between 0 and 5) for each. Specific rules for these two steps are being developed for each feature/attribute.

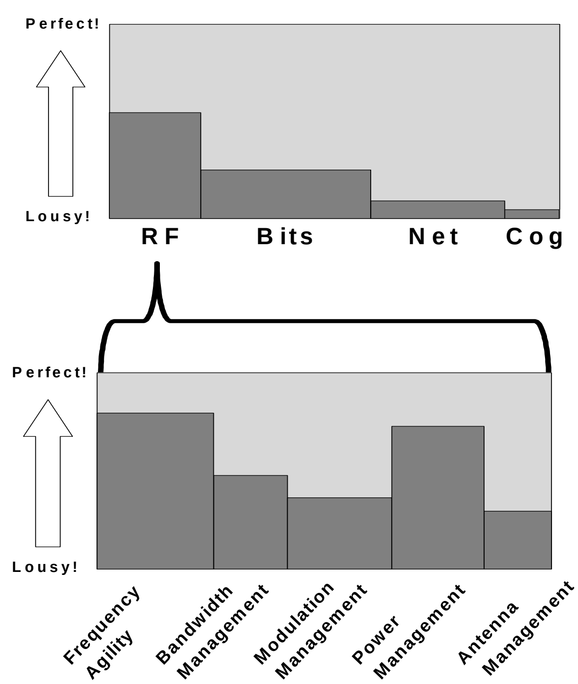


Figure 3. Aggregation of RF Features to Yield a RF Characteristics Domain Score

Note that we have shown the Network Management and Cognitive domains as having virtually no scores at all. With few exceptions we would expect current and near-term systems to score at acceptable levels in the first domain, but have low scores for the last two domains. The last two domains represent the largest area for growth in enabling technical features. It is where systems designers can make the most significant progress in Spectrum Citizenship. Thus, it is the area where the largest carrot for increased spectrum access rights should be dangling.

6. Conclusion – A "Guidance System" for Spectrum-Friendly Radio Technology

The Spectrum Scorecard is presented as means to evaluate Spectrum Citizenship within a tiered spectrum access rights model. Historically, DoD has relied on the spectrum certification process to ensure spectrum-dependent system supportability. This process often occurs too late in the system's design, cost, performance, and schedule identification cycle and is not structured to ensure optimized use of the electromagnetic spectrum, and does not provide pre-design goals and guidance to emphasize Spectrum Citizenship during the development and acquisition processes. This coupled with the transformation to network-centric operations and the need for bandwidth on demand necessitates a different approach to optimize DoD spectrum use.

In the commercial and public safety sectors, the FCC has shown interest in market-driven technology evolution. The Spectrum Policy Task Force and other activities lean toward regulatory structures that encourage sharing and highly automated, embedded spectrum management methods that leverage advances in technology to enable these new ways of granting spectrum access rights.

In both cases, the motivation to provide the enabling technology must have a strong presence in the design of systems that will reap the benefits of new spectrum access paradigms. The tiered spectrum access rights model is suggested as providing such motivation to build systems that display good Spectrum Citizenship.

This paper presents an initial Spectrum Scorecard that identifies and grades the complex ensemble of attributes of a spectrum-dependent system which impact the amount of spectrum required by the system and the efficiency with which the system uses the spectrum. A Spectrum Scorecard can provide a structured method of imposing pressure for good Spectrum Citizenship in emerging system design to earn desired levels of spectrum access rights.

7. References

[1] "*Manual of Regulations and Procedures for Federal Radio Frequency Management*", Washington, DC: US Department of Commerce, NTIA, May 2003 (revision May 2004).

[2] K. Kontson, *In Pursuit of a Wireless Device Bill of Rights*, FCC Technological Advisory Council, Sept 18, 2002.

[3] P. Ryan, *Wireless Communications and Computing at a Crossroads: New Paradigms and Their Impact on Theories Governing the Public's Right to Spectrum Access,* JOURNAL ON TELECOMM. & HIGH TECH. LAW, Vol 3, April, 2004.

[4] Spectrum Policy Task Force Report, FCC Docket No. 02-135, November, 2002.

[5] E. Goodman, *Spectrum Rights in the Telecosm to Come*, San Diego Law Review, Vol 4, April, 2004.

[6] *Technology-driven, Tiered Spectrum Access Rights*, presentations by multiple authors, FCC Technological Advisory Council, July 28, 2005

[7] R. J. Matheson, *Flexible Spectrum Use Rights*, (157–168), Proc. of the International Symposium on Advanced Radio Technologies (ISART), NTIA Special Publication SP-05-418, March 2005.

[8] Proceedings of the International Symposium on Advanced Radio Technologies (ISART), NTIA Special Publication SP-05-418, March 2005.

[9] Conference Record, 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005), 8-11 Nov 2005.

[10] Hari Balakrishnan, et al., *A Comparison of Mechanisms for Improving TCP Performance over Wireless Links*, IEEE/ACM Transactions on Networking, Vol 5, No. 6, December 1997.

[11] Global Electromagnetic Spectrum Information System (GEMSIS) Initial Capabilities Documents (Flag Level Review), Joint Staff, 10 December 2004.

[12] Global Electromagnetic Spectrum Information System (GEMSIS) Concept of Operation (draft), Joint Staff, 15 December 2003.

Conformity Assessment
of
Policy-Based Adaptive Radio Systems

Stephen Berger
TEM Consulting
Phone – (512) 864-3365
FAX – (512) 869-8709
E-Mail – stephen.berger@ieee.org

Abstract

This paper evaluates the requirements and attending challenges faced by regulatory conformity assessment systems when qualifying and then managing policy-based adaptive radio systems. Conformity assessment is the process by which a product or system is confirmed as meeting requirements. Regulatory agencies are required to protect some public interest. They protect the public interest by assuring that a product or service meets a set of defined requirements. Regulatory agencies utilize conformity assessment systems to carry out their mandates. Typically these systems are implemented using relevant ISO guidelines, with appropriate adaptations to the needs of the agency.

The technological innovation and flexibility of policy-based adaptive radio systems present particular challenges. Due to their dynamic and adaptive nature design qualification, often called type acceptance, will be a based on a probability of compliance rather than a fixed and proven determination. There are simply too many variables and potential operating states to measure the entire matrix of input conditions to output performance. The full matrix of possibilities will have to be reduced to a sparse but testable matrix. Further new tools for evaluation, field surveillance and field management will be required by these new technologies. This paper reviews the system, per the ISO guidelines, identifies the requirements with their attending challenges and suggests possible approaches for a robust conformity Assessment System.

Introduction

All regulators and other authorities responsible for the proper operation of a system implement some kind of conformity assessment system (CAS). Typically these systems follow ISO guidelines, making appropriate modifications for the needs of the agency, industry and technology being dealt with. This paper begins by reviewing the elements of a conformity assessment system for dynamic, adaptive radio systems, particularly policy-based adaptive radio (PBAR) systems.[1]

After a general outline of a CAS is described the similarities and differences between current systems and a CAS for PBAR are reviewed and challenges are identified.

Where appropriate, use of existing methods is assumed. The discussion here will highlight areas that are new or relatively unique and particularly those where fundamental shifts in approach may be required to implement a successful system.

Terms

For brevity the following abbreviations will be used in this paper:

[1] In the current draft of IEEE P1900.1 "Policy-Based Adaptive Radio" is defined as:

A radio that is governed by a predetermined sets of rules for behavior that are independent of the radio implementation regardless of whether the implementation is in hardware or software and both senses and adapts to its environment. The rules define the operating limits of such a radio. The definition and implementation of these rules can be:

- during manufacture or reconfiguration;
- during configuration of a device by the user or service provider;
- during over-the-air provisioning; and/or
- by over-the-air or other real-time control.

CAS **Conformity Assessment System**
A system designed to provide assurance that a product or service complies with specified standards and normative specifications.

PBAR **Policy-Based Adaptive Radio**
A radio that is governed by a predetermined sets of rules for behavior that are independent of the radio implementation regardless of whether the implementation is in hardware or software and both senses and adapts to its environment. The rules define the operating limits of such a radio. The definition and implementation of these rules can be:

- during manufacture or reconfiguration;
- during configuration of a device by the user or service provider;
- during over-the-air provisioning; and/or
- by over-the-air or other real-time control.

(IEEE P1900.1 draft)

Dynamic Spectrum Allocation
The near-real-time reallocation of spectrum resources in response to changing circumstances, including changes of the radio's state (operational mode, battery life, location, etc.), changes in environmental/external constraints (spectrum, propagation, operational policies, etc.), and/or in response to a received command.
(IEEE P1900.1 draft)

Adaptive Radio
A radio that adjusts its operation on a near-real-time basis to meet application needs in accordance with changing circumstances, including changes of the radio's state (operational mode, battery life, location, etc.), changes in environmental/external constraints (spectrum, propagation, operational policies, etc.), and/or in response to a received command. (IEEE P1900.1 draft)

DASM **Dynamic, Adaptive Spectrum Management**
A system of spectrum management that implements dynamic spectrum allocation using adaptive radios.

Conformity Assessment Systems

Most regulatory structures conform to the guidelines of ISO 17011[2] in designing conformity assessment systems.[3] Therefore this paper assumes that policy-based adaptive radio systems will be required to comply with regulations structured under the guidance provided in ISO 17011[4] and its companion documents. Hence, to receive regulatory approve some key questions must be satisfactorily answered before these systems will be permitted. Among these questions are:

1. What are the requirements for a minimal acceptable system and how is that assessment to be made?
2. Are the testing lab/testers/lab assessors qualified to effectively evaluate designs?
3. Will the vendor deliver units within manufacturing tolerances to those evaluated? What process must be implemented in their quality and change control systems for vendors to have adequate control of their production?
4. How will regulatory officials know if non-compliant units are deployed and what

[2] ISO/IEC 17011:2004 replaced three sets of overlapping requirements for the same attributes: ISO/IEC Guide 58:1993 (laboratories), ISO/IEC Guide 61:1996 (certification bodies) and ISO/IEC TR 17010:1998 (inspection bodies).
[3] Conformity assessment systems check that products, materials, services, systems or people measure up to the specifications laid out in a relevant standard. A lack of confidence in their competence to perform these tasks may result in redundant, costly and time-consuming assessments by different accreditation bodies in different countries. Such costs could be drastically reduced if a conformity assessment body, supervising such a system, can be assessed once and the results accepted globally.
[4] ISO 17011 aims to harmonize requirements worldwide for organizations that assess the competence of "conformity assessment" bodies.
It will provide a global benchmark for "accreditations bodies" to ensure that they operate in a consistent, comparable and reliable manner. It sets out a uniform set of requirements for bodies that verify the activities of conformity assessment bodies - from testing, inspection, management system certification to personnel certification, product certification and calibration.

corrective actions can be taken if interference problems arise?

5. Are there adequate safeguards that the systems will be used as intended?

In order to provide satisfactory answers to these questions most CAS contain the following elements:

- Type Testing / Design Evaluation
- Assessment of the Supplier's Quality and Change Management System
- Field Surveillance
- Field Management / Enforcement
- Training of Personnel & ongoing communication with stakeholders
- User Outreach

In order to assist in the design and management of CAS systems a number of international standards and guidelines have been developed. These documents are intended to encourage a degree of uniformity of different systems. There are many good reasons for this to be done including harmonization of regulatory requirements, allowing single qualification for multiple markets. There are economies to be gained when the same or a similar compliance evaluation can be used by multiple agencies in different countries. For these reasons there is significant similarity in CAS around the world. However, forces toward differentiation assure that there are also significant differences. The forces pressing for harmonization of requirements and CAS are in ongoing tension with those moving for differentiation of these same systems. Nevertheless, especially for mobile devices, the trend for harmonized requirements is strong.

One of the primary tools utilized to promote harmonization is the international consensus standards process. Various bodies develop standards for new technology areas in an open, consensus process. Regulators then adopt those standards, creating harmonization of regulations internationally. The IEEE 1900 series of standards has been initiated to serve this function for PBAR systems.

Several types of certification systems exist:

- Some comprise type testing only. These systems assume that if the design is compliant, as represented by a typical device submitted for evaluation, then the deployed units will also be in compliance.
- Other systems include initial testing and field surveillance. In these systems regulatory resources are focused on initial design evaluation supported by enforcement actions where non-compliance is identified. Quality and change control are left to the discretion of the manufacturer.
- Still other systems include initial testing of a product and assessment of its suppliers' quality systems, followed by routine audits, that takes into account the factory quality system and the testing of samples from the factory and the open market.

No single system is right for all product types and all purposes. Each application will adopt a specific implementation based on the history, needs, and resources of the regulatory authority along with its assessment of the consequences of non-conformity. On the last point, an agency is likely to implement a very stringent system if the consequences of non-conformity could be catastrophic. Alternately where consequences are less severe, especially if addressing non-compliance over time is acceptable, then more permissive systems may be employed.

Design Verification

Most radio transmitters are currently qualified for regulatory certification based on a set of complex but definitive engineering tests of various parameters. Such an approach is ineffective when evaluating a PBAR for compliance in a DASM system. The device is simply capable to too many operating states to test them all. Further, its evaluation will be based on a complete device, however the power of PBAR devices is that they combine three components that may be independently updated or modified. A PBAR has a hardware radio, controlled by software which is in turn operated under the constraints of a policy set. Any testing of a final device will have a specific combination of hardware, software and policies. However, the design of PBAR systems is dependent on being able to update these components, particularly the policy set, independently. New approaches to design verification would seem to be needed.

Within the IEEE 1900 series of standards IEEE Project 1900.3 seeks to establish a means of qualifying software modules intended for use in PBAR devices, independent of the hardware. IEEE P1900.3 proposes to develop an evaluation regiment which would demonstrate that the candidate software module would not command its host device into a disallowed state. IEEE P1900.3 does not propose that such testing is sufficient to qualify a final device. However, should a software module fail an evaluation and be found to, under some condition, instruct a host to transmit in a disallowed manner, then there is no reason to go further. Clearly unless the software module can pass an evaluation that module in an actual radio could not pass a qualification test.

A particular challenge of testing the software is the need of providing a means for proving the independence and completeness of the software evaluated. PBAR devices, like other wireless devices are certainly going to be hosts to a broad arrange of applications and features. Certainly it would be counterproductive if a device had to be re-certified every time one of the games on it was modified. However, to allow freedom to modify other applications and features will require strong assurances and safeguards that the radio control software is complete, independent and protected from all other software on the host device.

Some radios even utilize operating systems with application modules performing diverse functions from radio control to user applications, even games and entertainment. In such an architecture even knowing what function potentially influence the radio operation is a significant challenge. Even more daunting is proving that an application or module will and can never influence the radio operation.

Another issue is how to test the many possible scenarios that a PBAR may encounter as it senses its RF environment and adopts its operation to its environment. Physical creating and testing a device in the many different RF environments is both difficult and time consuming. New types of instrumentation do make the challenge somewhat less challenging but the root challenge remains.

One proposal is to require that when a candidate device is submitted for evaluation a simulation model of the device also be submitted. Using modern simulation tools the model of the device may be stressed under a wide variety of environments. Physical testing could be reduced to a smaller set of tests, with a primarily being validation of the accuracy of the model.

Possibly the best qualification regiment for these devices would be a carefully constructed combination of physical device testing, software qualification and simulation.

In reality all evaluations are probabilities. Every test has a measurement uncertainty and degree of test-to-test and lab-to-lab repeatability. However, the degree of flexibility and adaptability of PBAR devices will materially heighten the degree to which the final evaluation is, in the end, a probability of compliance.

How much certainty is sufficient to issue an equipment grant? The problem is very similar to evaluating product line quality. The only way to have 100% certainty is to test 100% of the production line. Even then it may be argued that 100% certainty is not provided as test-to-test repeatability may allow a passing device to fail on retest. For production line testing samples that provide statistically satisfying qualification of a production line has been long accepted. In the EMC arena international standards give guidance that a sample sufficient to demonstrate that 80% of a production line is compliant 80% of the time is sufficient.[5]

The probabilistic nature of PBAR evaluations is increased by the need to move from avoiding interference to managing interference within acceptable levels. In days gone by it was possible to build in enough margin so that there was not significant change of interference from a compliant system. However, as spectrum becomes increasingly scares and DASM systems are used to increase efficiency they will be designed to manage systems to an acceptable level of interference. Whether the systems are ultimately designed to give 80%, 90% or 99.999% confidence that interference will not occur, the number will not be 100% confidence. Hence the final conclusion of PBAR qualification is likely to be, “The evaluation of this system concludes that 80% of the units produced will cause less than 2% interference 80% of the time.”

In a parallel fashion qualification of PBAR devices will be based upon a sampling of their possible operating states. These states can and should be intelligently selection. Perhaps a combination of test engineering analysis aided by simulation results could narrow the actual testing to a manageable number of tests of most vulnerable or most critical operating states.

Properly constructed an evaluation combining component testing of hardware and software, simulation and physical device testing could deliver high confidence that the device is in compliance with the requirements of the DASM system.

In reality the probabilistic nature of an evaluation is not new. However, the degree to which PBAR evaluations are probabilities will take some getting used to.

---

[5] Typically between 3-10 devices must be tested to provide this level of confidence. This number of tests on a large production volume is not considered overly burdensome.

Qualification Regiment

New technologies call for new test methodology. With PBAR systems physical testing presents particular challenges due to the dynamic and adaptive nature of these systems. However, standardized test methods must be available to provide adequate evidence that system designs meet specifications and can be relied upon to operate within regulatory boundaries. It may be argued that regulators will be prevented from allowing such systems without such a qualification regiment because they will not be able to adequately prove such systems will operate as specified.

Key characteristics of a qualification regiment for PBAR systems are:

- **Abstracted Evaluation** – The interaction of real systems to must be abstracted into an evaluation regiment that has a satisfactory predictive correlation to the interactions between real systems. Trials using real systems will be essential in research and for proof of concept trials of PBAR systems. However engineering development and regulatory qualification will require an abstracted evaluation methodology.
- **Diverse Spectral Presentations** - Because PBAR systems are designed to sense and respond to their RF environment it is necessary to present the systems a diverse set of spectral environments and evaluate the PBAR system response. While it will be necessary to develop a full matrix of RF environments and scenarios, such a matrix will be far too full to test completely. A sparse matrix will be necessary. That sparse matrix must be defensible as adequately surveying the full population of possible operating environments PBAR devices will face when operating in a DASM spectrum scheme.
- **Multiple Variables** – The qualification regiment for PBAR device must be capable of manipulating the variables used by the DASM system. As all DASM systems manage for both time and frequency these qualification regiments must be able to present scenarios controlled for both time and frequency. They must also monitor the PBAR device's response in time and frequency.
- **Statistical Results** – DASM systems are inherently statistical in nature. Accordingly evaluation of PBAR devices will be inherently statistical. Regulatory testing for compliance with DASM system requirements will seek to provide assurance that the candidate device will never enter a disallowed state. However, given the infinite number of spectral presentations that may exercise different logic paths in the device, complete testing, using all possibilities, is not possible. Sufficient testing so as to provide a high degree of statistical confidence that a candidate system will behave appropriately is possible.

  In optimizing system design statistical evaluation will be even more important. It is unlikely that competing implementation possibilities will present uniform comparisons. Each implementation is likely to have its own strengths and weaknesses. Choosing the best system will be a process of selecting the system that presents the best overall performance. Having evaluation regiments that help quantify the relative strengths of competing implementations will be essential to the development and refinement of DASM systems and the PBAR devices that operate in them.
- **Repeatability and Uncertainty** – Given the dynamic and multi-variable nature of DASM systems developing evaluation regiments with acceptable repeatability and measurement uncertainty will present a particular challenge. Nevertheless it is necessary that these evaluation regiments be repeatable within a lab and between different evaluation labs. It is also necessary that the measurement uncertainty be small enough to allow meaningful comparison of results between one system and another.

Physical Testing

Physical testing of PBAR systems provides its own significant challenges. In this section several approaches are explored for the physical testing of PBAR devices. Each offers its own strength and benefits.

Conducted Testing

Conducted testing is often selected as the preferred test method because RF environments and measurements may be established more reliably. Similarly it may be assumed that many DASM qualification tests will be best performed by connecting candidate systems, interference sources and test equipment through conducted paths, which can be controlled for relative path loss.

Development of conducted system tests will undoubtedly be a significant component of PBAR device qualification. Proper and reliable implementation of such system test beds will need careful development and specification if repeatable tests are to be provided. Test bed qualification regimens will also be needed to assure that different labs perform the same evaluation of a candidate product.

Radiated Test Environments

Most device qualification regiments find it necessary to also provide radiated tests, it is likely that DASM systems will also require a radiated version of system testing. Establishing stable and repeatable test environments for radiated testing will present particular challenges. In considering the environment to be used in such testing two RF test methods recommend themselves, Gigahertz Transverse ElectroMagnetic Cells (GTEM) and Reverberation Chambers.

GTEM Cells

A GTEM puts the test object in a far field environment created by nature of the GTEM cell. GTEM's do not use antennas but the cell itself is the transducer, transforming a conducted signal into a transverse 377 Ohm wave. The test object is then exposed to the same kind of RF environment it would experience in the far field. The input to the cell and monitor instrumentation can be arranged very much like a conducted test, with the required RF environment being created by as many instruments or transmitting objects as are necessary to create the RF test environment. Similarly, using directional couplers the test object's emissions may be drawn from the cell and monitors for EUT operation under various test conditions.

A particular challenge is how to synchronize the test object function with the test environment. In order to demonstrate compliance with various timing tests it is necessary to either be able to trigger an attempt to transmit by the test object or to have a warning signal as it prepares to transmit. Such synchronization signals are then used to test the objects ability to defer a transmission within a set time should another transmission occur. It is exceeding difficult to synchronize the test object with the test scenario presentations. The GTEM has the advantage of creating a far field environment with little physical distance from the surrounding laboratory. If the test object software can make available appropriate synchronization signals then a GTEM based test allows a relatively convenient way of providing these to the test control software.

GTEM cell are likely to fill a significant role in PBAR device testing. GTEM cells are inherently wideband devices, they easily adapt to a wide variety of instrumentation, provide a stable RF environment and allow convenient access to the object under test.

A wide variety of RF scenarios can be presented to a test object using automated test software. The EUT response would then be recorded and the results compiled. Once test software has been written, automated testing would support testing large numbers of scenarios. This approach makes extensive exercising of a test object under a wide variety of conditions feasible.

Reverberation Chambers

A relatively new tool coming into increasing use is the reverberation chamber. Most test environments attempt to prevent or damp reflections, so as to create a stable, well controlled RF environment. In exact contrast a reverberation chamber utilizes reflections and actually encourages them. In reverberation chambers paddles are inserted into highly reflective cells and moved, mixing the reflected RF signals into a wide variety of combinations. The test object is subjected to a large number of RF presentations by the mixing actions of the stirring mechanism.

Reverberation chambers create extreme multi-path and fading environments. Because they arbitrarily mix the signals inside the cell it is arguable that they present a test object with a complete sampling of all possible test environments. Because DASM systems require exposure to a wide variety of environments there would seem to be a natural match between the characteristics of the reverberation chamber and the evaluation needs of DASM systems.

Reverberation chambers do have frequency limits that are size depended. It takes relatively large chambers to support and stir lower frequency environments. GTEM cells by contrast start working at DC and continue to be usable up to the point at which the performance of connectors and absorber components begin to degrade. Generally GTEM's are considered usable from DC to several GHz, with some manufacturers claiming 18 GHz performance. With a reverberation chamber there is not a natural upper frequency limit. Any frequency that can be radiated can be reverberated. However, the size of the cell creates a lower frequency boundary.

For testing of RF transmitting devices establishing stable connections in highly reflective environments is a problem and at times is not possible. The multi-path inside of reverberation chambers is simply too great to allow sustained link connection. To overcome this challenge it is likely that for DASM testing modified reverberation chambers would be necessary. Such chambers would introduce enough absorptive material to allow stable connections to be established, while simultaneously retaining significant reflective components with the stirring mechanism, to allow extensive mixing of the RF environment.

In summary there is a significant and exciting challenge to explore the possible methods for presenting a test object with a wide variety of RF scenarios and monitoring the EUT response. Conducting testing, GTEM and reverberation chambers may all come to find a role in full system evaluation. Alternately it may come about that only a subset of these is necessary to fully evaluate a DASM system and demonstrate its compliance.

Test Scenarios

Qualification testing will require the preparation of a diverse set of test scenarios to present to a DASM system. Happily this is not totally unplowed ground.

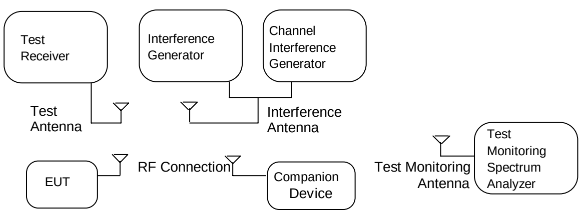


Figure 1 - Equipment configuration for a typical dynamic access test

The testing regiments, of necessity, become as complex as the DASM systems being implement. If the DASM system manages time and frequency as independent variables then its qualification testing must vary test scenarios with amplitude, time and frequency as independent variables. If a DASM system becomes more sophisticated, managing spectrum using time, frequency, coding and MIMO for signal separation, then test scenarios will be required that independently manage these variables in divergent but controlled combinations.

The complexity of the test requirements be determined by the candidate systems complexity. Systems designed to support duplex connections, delivering real time services will require stressing those connections. The following illustration from ANSI C63.17-2006 shows how interference with controlled amplitudes, relative time and frequency will be required to test the appropriate response of the candidate system.

Similarly complex will be the test monitoring system, which must become RF analogs to logic analyzers, monitoring a number of events simultaneously so that their relative timing and amplitude and other variables may be studied.

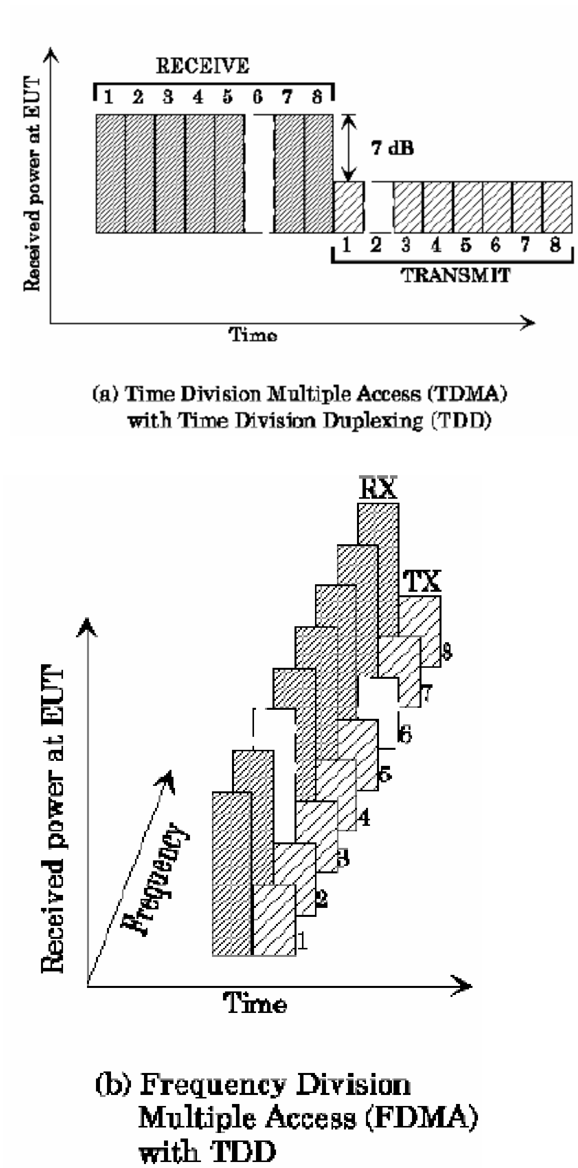


Figure 2 - Testing of TDMA and FDMA duplex systems requires simultaneous and coordinated control of frequency, time and amplitude

Qualification Paradigms

A very significant companion task for qualification testing is clearly defining the purpose of the qualification testing. Is this testing being performed to demonstrate proper system function or only efficient use of spectrum? These issues become extremely important when qualifying a product. The issues that arise often legitimately impinge on multiple aspects of a product. However, to have clear and objective qualification regiments consensus solutions must be developed and promulgated.

For cell phones such a division has been developed in which a mobile handset must receive an FCC equipment grant then it must receive CTIA approval before most carriers will accept the device into their network. The FCC equipment grant is limited to issues of spectrum use while the CTIA certification is focused on operational performance in a cellular network.

The advantage of this dual system is that it allow the regulatory requirements to be managed exclusively by the FCC, knowing that other issues, while very important will be managed by the CTIA certification. Such a division of responsibility supports a consistency of scope, which may not be possible otherwise.

There are a number of issues, some of which can be identified now. Others will certainly arise as specific DASM implementations are proposed and as these implementations become increasingly complex, using more variables to increase spectral efficiency and achieve other benefits.

It is important to identify these issues as soon as possible. For each issue its potential impact to spectrum efficiency or system operation must be explored and understood. Then, a consensus must be developed to assign this to an appropriate sphere of qualification. Some issues will appropriately be assigned as a spectrum management concern. Implementations that fall outside of certain parameters should not be allowed to operate. Other issues will not degrade spectrum efficiency or create increased interference directly but may affect system performance. While important these issues may be assigned as system qualification issues. The authority responsible for approving system components may be willing to accept degraded performance because of other benefits being delivered by a particular implementation.

However, seldom will issues fall completely into one area or another. An issue that degrades system throughput may result in increased system traffic as more retransmissions are required to successfully deliver the required data. The retransmissions will utilize more spectrum, degrading spectrum efficiency. So, as a secondary effect, system efficiency will affect spectrum efficiency. Such arguments can become increasingly complex, going to 3rd and 4th order effects. Qualification paradigms, representing consensus agreements, are required to support system qualification and avoid delays due to different qualification philosophies.

Clear boundaries and qualification parameters will be essential to the advancements of DASM systems. Without such definition and support DASM systems will be hindered due to conflict over these issues and a perception that the systems are not sufficiently matured to be trusted in full deployment.[6]

Probability of detection

Particularly difficult qualification issues arise when system performance depends on multiple variables. An example arises when considering a system with blind slots being tested for proper channel selection in a least interference channel system. The issue is, does a system qualify if its probability of detecting a spectrum opportunity falls below 100%? If so, at what probability of detection does it fail to meet requirements? In this example, the technique being tested requires that a device monitor a large number of channels and select the channel that has the least interference, below a certain threshold.

To test least interfered channel requires that all available channels have interference introduced on them at defined levels. One channel is left interference free. To pass the test a device must show that it will properly select the open channel.

A complexity arises when evaluating compliance with a system with blind slots. Using the system described above with, blind slots created by it reaction time, the ability of the system to identify an interference free channel will be limited by these blind slots. The illustration below assumes a system in which a dummy bearer is used to synchronize operation. It shows that for some relative assignment of a timeslot for the dummy bearer and the available channel the system will pass but for other relative assignments it will fail. The question then becomes; does this system pass or fail the qualification test?

In the illustration, the system will properly identify the available, non-interfered channel in 3 of 6 timeslots but fail to identify it in the other 3 timeslots.

Perhaps the core question is, "What probability of detection is required for a system to be judged to have successfully implemented a technique?"

Test system capabilities

The ability to test is built upon the capabilities of available test systems. In the illustration below interference is introduced to test channel access in a fixed frame system. As can be clearly seen the

[6] These considerations may recommend that with in the IEEE 1900 series a test standard be developed for spectral efficiency, perhaps 1900.4 and a different evaluation be developed for system operational efficiency, perhaps 1900.5?

interference generator has a defined bandwidth. A threshold is defined but the interference, while it will cross that threshold at a given point, defined to be at the channel edge in this illustration, has only slightly different energy levels on either side of the boundary.

This physical requirement of real test systems and test uncertainty created by the precision with this the system can control the amplitude or frequency of the interference creates a limit on the ability to test the candidate system.

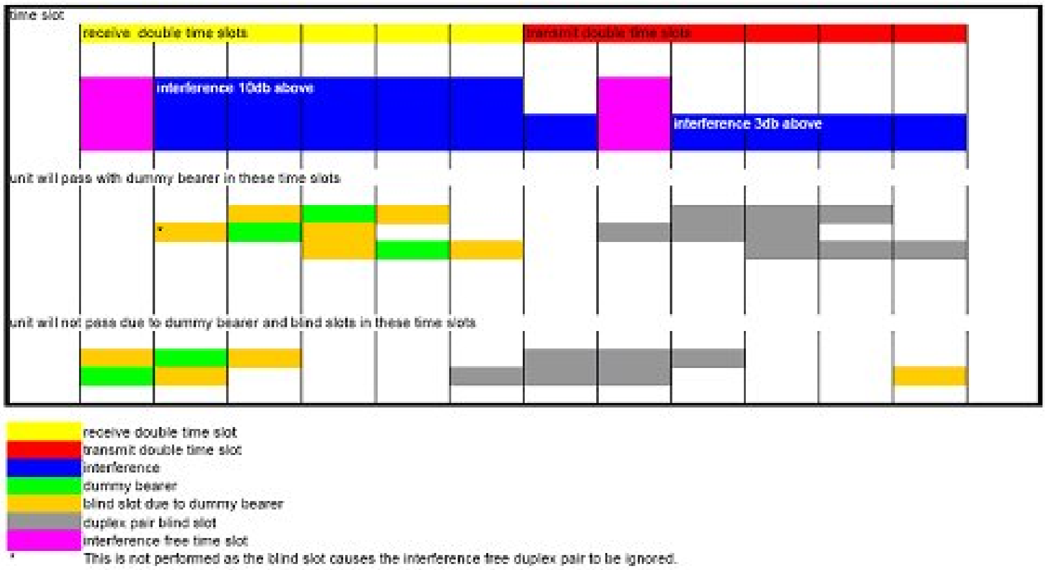


Figure 3 – Blind slots create non-detectable transmission opportunities

While the issue illustrated is reasonably straightforward, other test system dependencies are not. It is not uncommon in complex tests to have divergent results that when analyzed in detail reflects not differences between candidate systems but differences in test system capability.

End result correlation

Ultimately a qualification regiment is an abstraction that is intended to predict a desired real-world result. Because the qualification regiment is an abstraction its efficacy in predicting the desired result is always questionable and mostly like imperfect. The degree to which a qualification regiment correlates to the desired result is typically a subject of ongoing debate. Finding an acceptable level of correlation will be a consensus building process in the end. A balance is necessary between the compliance cost and the degree of confidence delivered that specifications are met.

It is not unusual, especially with new technologies to observe many suggestions for 'improvements' to the qualification regiment being offered. In some cases, when good review and adoption processes are not in place, effective adoption of refinements and new innovations to a qualification regiment is missed. Effective adoption would be a process that would see the adoption of innovations that deliver solid benefits at reasonable cost and would reject those that fall short.

With a system as complex as DASM systems promise to be the qualification regiment could easily become quite expansive. It must be efficient for DASM systems to advance.

Much will be gained if the first qualification regiments have solid justification in correlation to the desired end result. Further, if the rational and process of adoption or rejection of candidate elements of the regiment are clear, a process will be established by which to judge and select among future candidate tests. Effectively correlating the qualification regiment to real-world performance will require the difficult but very important task of having good real-world performance information to compare with the results of the qualification regiment or to candidate elements for that system.

Test repeatability and uncertainty

For complex measurements and evaluations the issues of repeatability and uncertainty are significant. Certainly for an evaluation as complex as those required of a DASM system there are significant challenges to achieving repeatability and uncertainty within acceptable bounds.

Given a complex test there is a significant chance that two different laboratories will arrive at different results when evaluating the same system. Even the same lab may arrive at different results at different times, evaluating the same system. How test repeatability and uncertainty will be dealt with are formidable challenges in designing a CAS for DASM systems.

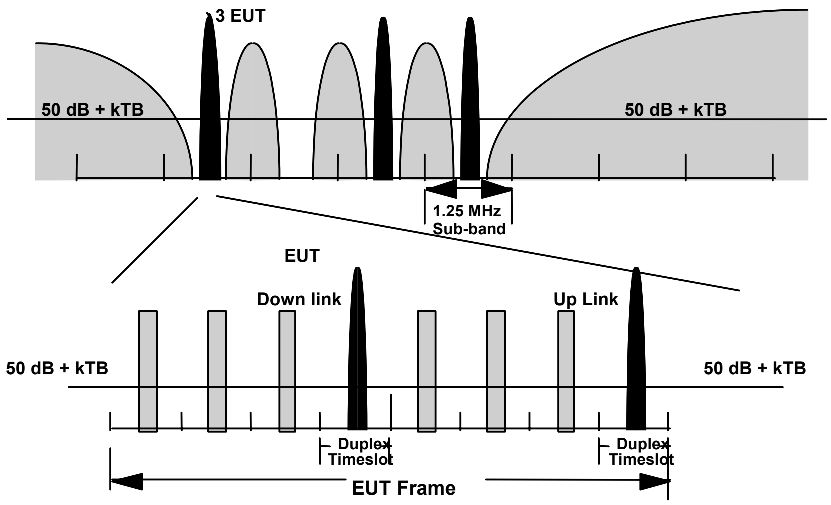


Figure 4 - Testing channel access in a TDMA fixed frame system

Conclusions

This paper has presented the elements of a generic CAS system and explored the particular challenges presented in designing a CAS for DASM systems. These qualifications may be a designed mixture of software and hardware evaluation with simulations and analysis. New methods for performing conducted and radiated tests will be required. Having clear and technically mature test standards is essential if these requirements are to be met. It is this challenge that the IEEE 1900 series of standards seeks to meet.

Until satisfactory qualification regiments are available regulatory adoption of DASM systems may not be possible.

References

ANSI C63.17-1998, *American National Standard for Methods of Measurement of the Electromagnetic and Operational Compatibility of Unlicensed Personal Communications Services (UPCS) Devices*

ISO 9001:2000 Ed. 3 *Quality management systems -- Requirements*[7]

ISO 9004:2000 *Quality management systems — Guidelines for performance improvements*

ISO 10007:2003 *Quality management systems — Guidelines for configuration management*

ISO 10012:2003 *Measurement management systems — Requirements for measurement processes and measuring equipment*

ISO/TR 10013:2001 *Guidelines for quality management system documentation*

ISO 10015:1999 *Quality management — Guidelines for training*

ISO/TR 10017:2003 *Guidance on statistical techniques for ISO 9001:2000*

ISO/IEC 17011:2004 Ed. 1, *Conformity assessment -- General requirements for accreditation bodies accrediting conformity assessment bodies*

ISO/IEC 17024:2003 Ed. 1 *Conformity assessment -- General requirements for bodies operating certification of persons*

ISO/IEC 17025:2005 Ed. 2 *General requirements for the competence of testing and calibration laboratories*

ISO/IEC 90003:2004 Ed. 1 *Software engineering -- Guidelines for the application of ISO 9001:2000 to computer software*

[7] The current version of ISO 9001 replaced the previous version of ISO 9001 as well as ISO 9002 and ISO 9003.

Structure and Authentication of Policy Loads for Policy Defined Radio Systems

Stephen Berger
TEM Consulting
512-864-3365
stephen.berger@ieee.org

John Chapin
Vanu, Inc.
617-864-1711
jchapin@vanu.com

Abstract

A policy defined radio is a device whose operating parameters are constrained by downloaded policy rather than hardware limitations. Correct operation of these devices depends both on the contents of the policy and on the authentication mechanisms that assure only approved policies will be accepted. This paper explores the new requirements created by policy-defined radios and suggests a structure for policy loads that naturally fits those requirements. In addition a means of satisfying these requirements by adopting tools being used by law enforcement for computer forensics and offered by the NIST National Software Reference Library (NSRL) is explored.

Introduction

The core concept of policy defined radio is to break the link between the capabilities of a radio device and its operating limits. Under the policy defined radio concept, a device may have a great deal of capability but be limited in its operation to a set of policies, which are downloadable to the device.

Policy defined radio significantly improves both regulatory control and operational flexibility of radio systems. Regulatory control is improved because policies can depend on factors such as location and time of day, and policies can be changed over time as justified by changing market, societal or technical conditions. Operational flexibility is improved because network managers can exploit a single hardware device for a variety of different functions requiring different capabilities.

However, the flexibility and adaptability of policy defined radio presents new challenges. This paper describes and analyzes these challenges and suggests a policy structure to help solve them. The use of cryptographically signed files to authenticate downloaded policies is considered in detail.

Background

What is a policy

A policy is an artifact that provides a yes/no answer to the question "is it acceptable to perform a specified transmission." A transmission is specified in part by the traditional radio parameters of center frequency, power, and modulation. There can also be nontraditional policy inputs, including time and date, geographic location of the radio, output of a spectrum sensor, and information provided by other collaborating radios.

A policy is represented as a data set that can be downloaded to the radio device, typically as a file or group of files. Within this data set there can be data (tables, strings) to be interpreted by a policy engine within the radio, or software for direct execution.

In implementing a policy-defined radio, the policy engine or software need not run before each transmission. If a series of transmissions will use the same center frequency, power, and modulation, for example, the policies need only be consulted once. For more sophisticated policies, the device may need to monitor information like time of day or geographic location against certain boundaries provided by the policy, to know when the policy must be checked again.

Comparison to SDR and other radio types

A critical attribute of a true policy-defined radio is the ability to download multiple policies from independent sources. A transmission is only permitted if all active policies permit it. This is called *policy mixing*. The importance of policy mixing will become clear later in the paper.

A Software Defined Radio (SDR) is a device where the signal processing functions of the radio are implemented in software rather than in hardware. This enables a single hardware device to support multiple radio standards.

An SDR is not necessarily policy-defined. The purpose of SDR software is to generate and receive signals rather than to permit or reject transmission. If multiple SDR software loads are downloaded from different sources, the single load that is running at any given time normally has complete control over the transmitted signal. Therefore a traditional SDR does not perform policy mixing. Of course, a radio that uses SDR to implement its signal processing can also use policy-defined radio techniques for management.

A radio with Dynamic Access Spectrum Management (DASM) is a device which, instead of requiring a static frequency reservation, implements access rules that permit transmission under specific conditions. A policy-defined radio does not necessarily support DASM. For example, only policies that represent static frequency reservations may be allowed on the radio. Similarly, a radio with DASM is not necessarily policy-defined. The DASM rules can be implemented in local software rather than being downloadable or mixable. The term Policy-Based Adaptive Radio (PBAR) has been identified in the current draft of IEEE 1900.1 to identify implementations that are both policy-defined and adaptive.

A Cognitive Radio uses autonomous goal-directed reasoning, also known as artifical intelligence, for local control. This enables the radio to learn from events and thereby adapt more effectively to user needs and environmental challenges. A radio can be cognitive without being policy-defined and vice versa. However, cognitive radio techniques are widely considered to be a promising approach for implementing sophisticated policy engines for policy-defined radio, so there is a close link between the two approaches. However, the goal-directed reasoning is bounded such that regulatory authorities may confirm that the adaptability will never configure the radio for a disallowed transmission state.

Conformance assessment

Most regulatory structures follow the general requirements of ISO 17011 in designing conformity assessment systems. It is assumed that policy defined radio systems will be required to comply with regulations and other requirements generally structured under the guidance provided in ISO 17011 and its companion documents. Hence, to receive regulatory approve some key questions must be satisfactorily answered before these systems will be permitted. Among these questions are:

1. What are the requirements for a minimal acceptable system?
2. Are the testing lab/testers/lab assessors qualified to effectively evaluate designs?
3. Will the vendor deliver units within manufacturing tolerances to those evaluated?
4. How will regulatory officials know if non-compliant units are delivered and what corrective actions can it take?
5. Are there adequate safeguards that the systems will be used as intended?

Within question 1 of this general framework, policy-defined radio creates a challenging set of new problems for conformance assessment, especially when compared to traditional radios.

- *Correctness of a policy artifact:* Does the artifact actually represent a legal and desirable policy? Does the artifact comply with local regulatory requirements?
- *Correctness of a policy engine:* If the policy artifact is data rather than code, does the policy engine in the device behave correctly when interpreting that data?
- *Correctness of policy support in the device:* Does the device provide correct inputs, e.g. do all sensors behave as required. Also, does the device correctly respond to policy decisions, e.g. does it halt transmissions when rejected by the policy?
- *Correctness of policy use by the device:* Does the device check the active policies at the required operational points?
- *Device integrity:* Does the device sufficiently defend against a malicious user or third party who seeks to corrupt or bypass policy processing?
- *Policy authentication:* Does the device sufficiently defend against attempts to download unapproved or corrupted policies?

This paper only considers the last requirement, assuring the radio operates only with approved policies. The other conformance requirements are equally critical and are the subject of active work by multiple researchers, for example in the DARPA XG program.

Structure of policy loads

Rationale

Regulators have different levels of concern for different aspects of device behavior. Some requirements, such as those that relate to operator safety from RF exposure, require high levels of assurance. Other requirements, while still important, require lower levels of assurance. In particular if effective means of field management are provided some risk of violation of this class of lower level parameters may be allowed.

An example of a less stringent requirement might be international unintentional emission requirements under the IEC standards. When considering the issue of manufacturing tolerance of devices the IEC standards give guidance for statistical sampling of production lines. This guidance finds that a statistical assurance that 80% of a production line will be compliant with the unintentional emission requirements 80% of the time is considered satisfactory compliance. The issue in this case is to keep the required sample testing to reasonable limits. If higher levels of assurance were required large numbers of devices would require testing simply to prove compliance. Such additional testing may have arguably little value. The point being made is that there are different levels of assurance required by regulators for different requirements.

Structure

As the requirements of policy loads for software defined radio are considered there would seem to be three levels of policies:

1. Some policies are critical and require strong assurance that they will never be changed in deployed devices. Examples might be that a handheld device will never exceed the RF safety limits. Another might be that a device intended for use on planes will never transmit on radio navigation frequencies. For these policies, called in this paper Critical Policies, there must be very strong safeguards. A characteristic of some and perhaps all of these policies is that they apply widely, even globally. As examples the RF safety and unintentional emission requirements are widely implemented using international standards.
2. Of only slightly less concern are what may be termed Regulatory Policies. These policies contain the basic operating constraints established by regulatory authorities. These also require significant safeguards and the ability to verify that deployed devices in fact are operating ONLY with approved policy sets.
3. The third group of policies is called here Management Policies. These policies are available to network managers for use in optimizing the systems under their supervision.

A logical priority is apparent in this organization. A Regulatory or Management policy must never override a Critical policy. Similarly a Management policy must always operate within the bounds of both Critical and Regulatory policies.

Policy Level	Scope	Assurance	Flexibility
Critical	**Global or Regional**	**Very High**	**Low**
Regulatory	**National**	**High**	**Moderate**
Management	**Local**	**Moderate**	**High**

Policy mixing

One of the goals of this structure is to allow great flexibility to network managers while giving regulators the assurance that devices will only operate within prescribed parameters.

In contemplating this goal it becomes clear that there are tendencies to separate these levels of policies by their scope of application. Critical policies tend to have very wide, even global scope. As an example, all nations require that RF exposure limits be observed and despite national differences, there are international recommendations on those limits. In contrast Management policies would normally only apply to local systems and be modified between systems to optimize performance.

This organization suggests the possibility of levels of control and flexibility. Critical policies will require the highest level of control and assurance while Management policies may require only moderate control and offer great flexibility. If there is good confidence that Critical and Regulatory policies can be relied upon to properly restrict device behavior then great flexibility and freedom may be given at the level of management policies.

The policy mixing mechanism defined earlier becomes valuable in this context. Critical, regulatory and management policies will be authored independently and loaded into the device at different times. Policy mixing assures that each policy load has the independent ability to consider and potentially reject each specified transmission.

Authentication of policy loads

Threat model for policy authentication

Policy authentication in a policy-defined radio defends the radio against download of corrupt or unapproved policies. As a basis for assessing policy authentication mechanisms, we categorize the threats that could lead to incorrect download attempts.

- *Accident:* The owner or operator of a device accidentally presents a policy to the wrong device, or there is a configuration management error somewhere.
- *Cheating by user:* The user seeks to improve performance or access capabilities beyond what is permitted by approved policies.
- *Cheating by manufacturer:* The manufacturer seeks to improve performance or access capabilities beyond what is permitted by approved policies.
- *Local attack:* The user or other person with physical access to the device seeks to cause it to behave in a harmful manner.
- *Remote attack:* A person without physical access to the device seeks to cause it to behave in a harmful manner.

Clarifying the threat model enables considering which attacks should be considered and which ignored when designing the policy authentication mechanisms.

For example, a cheating manufacturer could design the device to internally modify or ignore a downloaded policy. Possibilities like these mean that cheating by manufacturer is better addressed through legal deterrents than through any technical mechanism. As a result this threat is not considered in the policy authentication mechanism.

While a local attack will be rare, cheating by the user is fundamentally the same case and is likely to be a significant problem. Therefore the policy authentication mechanism must defend against this.

In particular, the traditional authentication mechanism where a policy is accepted if and only if it is presented across a hardwired local connection (rather than delivered remotely across the network) is not an effective mechanism for policy-defined radios.

The most common threats are likely to be accidents and remote attacks, both of which need to be considered.

Public key cryptography

Public key cryptography offer basic mechanisms that can be leveraged for policy authentication. With public key cryptography, the approved policy author encrypts the policy using their secret key. Their public key is well known, for example stored in the radio device, and successful decryption using that public key provides proof that the policy could only have come from the approved policy author.

A public key approach could be incorporated into a regulatory or certification scheme whereby only a highly trusted source could encrypt a policy, especially a critical or regulatory policy. It is possible that the only method for having a policy properly encrypted for transmission would be at the end of a regulatory approval or certification process. Used in this way the regulatory or certifying authority would have assurances that disallowed policies could not be deployed.

Interrogation of device policies

In addition to authentication of policy downloads by the devices receiving them, cryptographic mechanisms can also be used to learn externally what policies a device has loaded. This supports a number of user requirements.

Cooperating devices will require a means for validating that other devices in their local system are using identical or compatible policy loads.

System or network managers will require a means for knowing what devices are operating on their system and what policy loads they are using. When system operation must be modified the manager will require a means for confirming that updates have been successfully received and implemented.

Digital signatures

The basic mechanism supporting policy interrogation is the use of digital signatures produced by HASHing algorithms. A digital signature is a short number, typically 128 to 1024 bits, computed from an original file, with the following characteristics:

1. The original file cannot be reproduced from the digital signature. Thus the original file is secure and remains confidential even though the signature may be made widely available.
2. It is computationally infeasible to create a new file that computes to the same digital signature as an existing file. These algorithms truly produce a unique signature or “fingerprint” of the file.

Policy interrogation based on signatures

By use of digital signatures policy sets can be uniquely identified with high confidence. To achieve this, a list of signatures of validated policy sets must be available from a trusted source.

This could be arranged by having regulatory authorities escrowed policy files as part of the approval process and issue signatures on the escrowed files. Further, policy defined radios can be required required to have a capability to HASH their policy sets and transmit those codes (either over the air or over a local connection) when requested.

This would provide a means of verifying a radio’s policy set without knowing the details of the policy set. Regulators and network managers would be able to check deployed systems and gain assurance that they were operating with approved policy sets.

If devices transmitted with HASH codes when given an authorized command to do so then regulators and network managers would be able to verify the systems operating in their vicinity and use this tool when dealing with network or regulatory issues.

Parallel of the NIST NSRL

An important model to be studied for application to policy defined radio is the NIST National Software Reference Library (NSRL). The NSRL is a project supported by the U.S. Department of Justice's National Institute of Justice (NIJ), the FBI and other federal, state, and local law enforcement agencies, with the National Institute of Standards and Technology (NIST) to use computer technology in the investigation of crimes involving computers.

The NSRL is designed to collect software, produce profiles of those files and incorporate the profiles computed from this software into a Reference Data Set (RDS). The RDS can be used by law enforcement, government, and industry organizations to review files on a computer by matching file profiles in the RDS. This information is used by law enforcement when analyzing computers or file systems that have been seized as part of criminal investigations. Standard program files and other files that are known and unaltered can be eliminated, leaving only files that may contain evidence.

The RDS is a collection of digital signatures of known, traceable software applications. There are application hash values in the hash set which may be considered malicious, i.e. steganography tools and hacking scripts. Hence the mechanism may be used to separate known trusted files, from known malicious files and unknown files.

Reference Data Set version 2.10 was released in September 2005 containing 10,663,650 unique digital signatures for 33,860,009 files. The signatures are produced using SHA-1, MD5 and CRC32 algorithms.

HASH Examples of Different Microsoft Notepad Versions[1]

Version	Bytes	SHA-1
NT4\ALPHA	68368	F1F284D5D757039DEC1C44A05AC148B9D204E467
NT4\I386	45328	3C4E15A29014358C61548A981A4AC8573167BE37
NT4\MIPS	66832	33309956E4DBBA665E86962308FE5E1378998E69
NT4\PPC	68880	47BB7AF0E4DD565ED75DEB492D8C17B1BFD3FB23
WINNT31.WKS\I386	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67
WINNT31.SRV\I386	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67

At this time NIST believes that the SHA-1 has not been broken, however there are known MD5 collisions and weaknesses. The NSRL data provides an MD5 to SHA-1[2] mapping to facilitate the migration away from MD5. The SHA-1 algorithm will be superseded in 2010 by FIPS 180-2, Secure Hash Standard, which contains SHA-224, 256, 384 and 512. The NSRL plans to provide a SHA-1 to SHA-256 mapping.

---

[1] From Douglas White’s White, Douglas, Presentation to the EAC TGDC, July 9, 2004,

[2] Secure Hash Algorithm (SHA-1) is specified in FIPS 180-1. It is a 160-bit hashing algorithm which performs 1045 combinations of 160-bit values to produce a unique digital signature or “fingerprint”.

Currently the NSRL is used by ISPs to track application sharing on servers. System administrators are also using this tool to confirm valid operating system file states on machines in their network. These applications have many similar characteristics to the needs of policy defined radio systems and suggest the possibility of a positive adoption of a modified version for the needs of policy defined radio systems.

Conclusions

Potential for standardization

If a standard in the IEEE 1900 series were to be developed to support this approach to the structure and authentication of policy loads, it might be outlined as follows:

1. Policy categories and contents of each category.
2. Protections required of each category of policy.
3. Public key cryptographic algorithms to be used for authentication
4. Digital signature algorithms to be used.
4. Format for requesting transmission of HASH codes and format for their transmission.
5. Informative annexes outlining how this functionality might be utilized by regulators, manufacturers and network managers to facilitate their work and assure the integrity of the system.

Closing Summary

This paper has reviewed the requirements for conformance of policy defined radio systems, with a particular focus on authentication of policy downloads. A hierarchy of policies is proposed containing three categories of radio policies. Parallels with the needs of law enforcement in computer forensics, as addressed in the NIST NSRL, are explored for application to the needs of policy defined radio. Mechanisms for providing the required security, verification and independence of various policy levels and components of the verification system are then discussed. Taken together, the hierarchy, use of cryptographic mechanisms and implementation protections are proposed as a possible means of addressing the requirements of policy provisioning systems.

References

White, Douglas, "NIST National Software Reference Library (NSRL)", Presented at the Mid-Atlantic Chapter HTCIA Meeting , September 28, 2005, VA[3]

White, Douglas, "National Software Reference Library (NSRL)", Presentation to the EAC TGDC, July 9, 2004, Washington D.C.

[3] NIST NSRL presentations are available at: http://www.nsrl.nist.gov/Presentations.htm

Multi-Band, Multi-Location Spectrum Occupancy Measurements

Mark A. McHenry
Dan McCloskey
Shared Spectrum Company, 1595 Spring Hill Road, Suite 110, Vienna, Virginia, USA
Tele: 703-761-2818x103, Fax: 703-761-2817
mmchenry@sharedspectrum.com

Abstract

This paper describes multi-band (30 MHz to 3000 MHz), multi-location (six urban and rural locations) spectrum occupancy measurements performed from January 2004 through August 2005. The project consisted of building a high dynamic range spectrum measurement system, developing a data collection and processing system and conducting spectrum occupancy measurements at six locations. The project goal was to measure the spectrum occupancy in all bands 30 MHz to 3,000 MHz. These measurements are critical to determine what bands have low utilization and to develop cognitive radio algorithms related to dynamic spectrum sharing (DSS). The low spectrum occupancy indicates that a DSS radio system could access a huge amount of "prime" spectrum. The unoccupied, large contiguous spectrum blocks shows that the DSS radio can use conventional contiguous waveforms and that high temporal agility is not required.

1. Introduction

Access to radio spectrum is at a crossroads. More and more technological alternatives are becoming available and demand from both public and private sectors is increasing rapidly, if not exponentially. Increasingly, there is recognition that the root of the problem is that most of the spectrum is actually unused, and the present system of spectral regulation is grossly inefficient. Current spectral regulation is based upon the premise that slices of the spectrum, representing uses within specified upper and lower frequency bounds, must be treated as exclusive domains of single entities who are the recipients of exclusive licenses to use specific frequency bands.

1.1 Dynamic Spectrum Sharing Technology Development

The U.S. Government is investing significant research and development funding to develop spectrum sharing technology. Programs include the DARPA XG Program and the NSF NeTS-Pro-WiN Program. These investments may be wasted unless they address the true spectrum use situation. It is critical to immediately conduct more spectrum measurement and analysis described here to insure that the government's R&D investments target the correct technical problems.

1.2 Importance of Spectrum Measurements

Spectrum measurements are critical to government policy makers and to NSF (and other) researchers in the development of new spectrum access technologies. Specifically, spectrum occupancy studies identify what spectrum bands have low or no active utilization and thus may be appropriate for spectrum sharing. They the provide information on the signal characteristics within these bands, which is needed to design spectrum sharing algorithms. The most important use of the data is to convince senior US government officials that RF spectrum is being used very inefficiently, and that they should make R&D investments and policy changes to support the development of dynamic spectrum sharing radios.

What is needed from spectrum measurements is a band-by-band analysis of the:

- Spectrum occupancy (peak and average)
- Detected signal parameters (transmission gap statistics, transmitter mobility, number of transmitters, the signal bandwidths, and other parameters)
- FCC/NTIA rules to determine reasons why signals were not detected (i.e. not present, very low duty cycle, or too weak to detect)

This information needs to be collected over a wide range of locations to assess the variations of spectrum usage in environments that have different spectrum users present and population densities (urban, suburban, and rural). The results from the different measurements must be consistently analyzed and plotted because of the complex nature of the spectrum use problem and because of the contentious points of view related to spectrum issues.

There have been several previous broadband spectrum surveys [1][2], but they did not provide temporal spectrum use information and could not be used to provide spectrum "white-space" estimates.

1.3 Project Goals

The project goal was to measure the spectrum occupancy in all bands from 30 MHz from 3,000 MHz. This provides information on:

- What bands have low utilization,
- How the spectrum is being used (what types of equipment, where, when, mobile or fixed, …),
- The existing user's equipment parameters (signal bandwidth, modulation, power levels, etc),
- The spectrum occupancy gap width and duration statistics,
- The number of transmitters in each band, and
- The background noise level.

These parameters are critical to developing cognitive radio algorithms related to dynamic spectrum sharing. Some of the above parameters come directly from the spectrum data. Other parameters need to be interpreted using models and hypothesis, which is an ongoing investigating and is not included in this project.

1.4 The National Radio Network Research Testbed (NRNRT)

Measurements contained in this paper are part of the National Radio Network Research Testbed (NRNRT) project. The NRNRT is a National Science Foundation (NSF) project that supports research and development of new radio devices, services, and architectures, providing a valuable facility for use by the research and development community in testing and evaluating their systems.

The NRNRT consists of:

(1) a field measurement and evaluation system for long-term radio frequency data collection, and an experimental facility for testing and evaluation of new radios;

(2) an accurate emulation/simulation system that incorporates long-term field measurement, for use in evaluating new wireless network architectures, policies, and network protocols; and

(3) innovative experimentation with wireless networks that integrate analysis, emulation/simulation, and field measurements.

2. Measurement Locations

Table 1 shows the measurement locations in this study. The locations include outdoor urban and rural locations, and an indoor location. Presumably, New York City has the nearly the highest and the National Radio Astronomy Observatory has the lowest spectrum occupancy of any locations in the US. Most of the measurement locations were highly elevated and had excellent line-of-sight to the surround area (thus, maximizing the detection probability).

3. Measurement Equipment

The equipment used for measurement in this study consisted of a spectrum analyzer, an SSC-designed pre-selector, an omni-directional discone antenna, a small log periodic array (LPA) for frequencies greater than 1000 MHz, and a laptop computer. The antennas were connected to a high-linearity pre-selector. A long RF cable, a pre-selector control cable, and a pre-selector power cable connected the Pre-Selector box to a shielded Faraday cage enclosure (Figure 1). The shielded enclosure contained a 3 GHz spectrum analyzer, a laptop computer, and power supplies. Power was provided to the equipment using an extension cord plugged into a 120 volt AC outlet or via a gasoline powered generator.

Location	Dates	Purpose
Inside Shared Spectrum Company offices	2/4/2004 2/9/2004 10/28/2004	Test equipment
Outside in Shared Spectrum parking lot	4/6/2004	Urban location
Riverbend Park in Northern Virginia	4/7/2004	Rural location
Tysons Corner shopping center parking lot in Vienna, Virginia	4/9/2004	Urban location
National Science Foundation (NSF) building roof in Arlington, Virginia	4/16/2004	Elevated, urban location
New York City	8/5/2004 8/30/2004	Elevated, urban location
National Radio Astronomy Observatory, Green Bank, West Virginia	10/4/2004	Very quiet, rural location
Shared Spectrum office roof	12/15/2004 -6/9/2005	Elevated, urban location

Table 1: Measurement Locations

Figure 1: The RF shielded enclosure in which all the data collection equipment is placed.

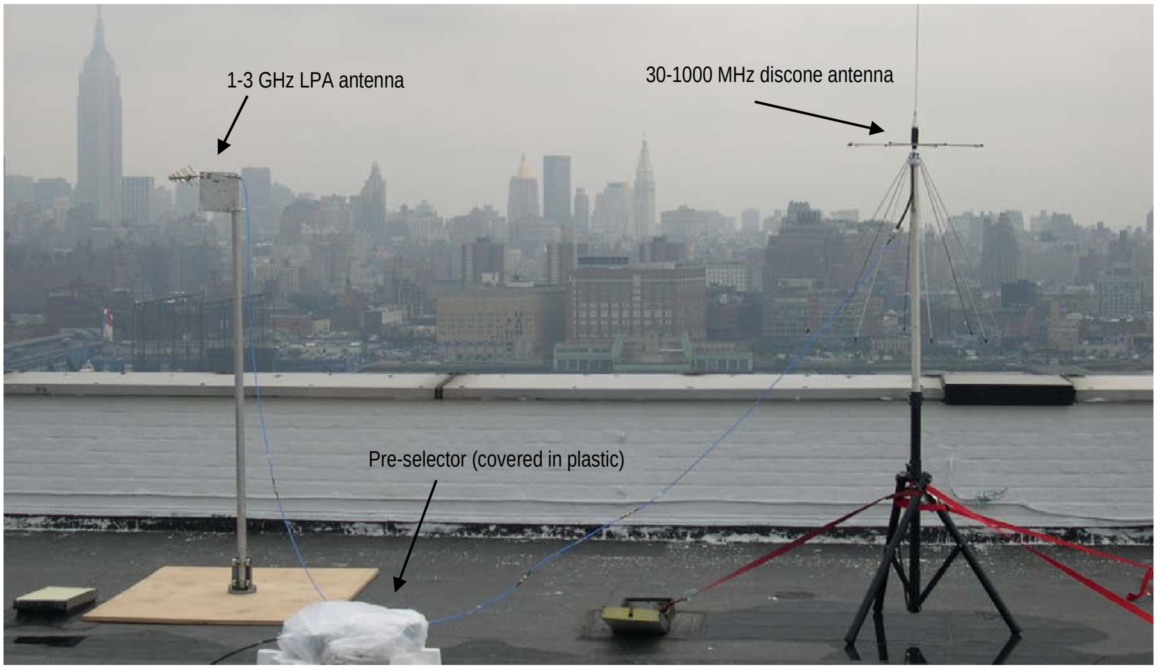


Figure 2: Antenna and Pre-Selector on roof in New York City.

3.1 Pre-Selector Description

A Pre-selector was used to improve the measurement sensitivity and dynamic range. The Pre-Selector configuration is illustrated in the block diagram shown in Figure 2. It consists of filters, RF switches, pre-amplifiers and programmable RF attenuators. The upper path is used for signals from 30 MHz to 1000 MHz, and the lower path is used for 1000 MHz to 3000 MHz. The Pre-selector was located on the roof a short distance from the antennas. It was installed in a weather-proof enclosure.

3.2 Data Collection

At each location, prior to measurement, the pre-selector attenuation values, filters and spectrum analyzer RF attenuation values were manually adjusted in each band. The settings that obtained the lowest distortion and noise for each band were used in the input file that controlled the equipment. During the measurements, separate files were created for each spectrum analyzer trace.

3.3 Data Calibration

The plotted spectrum data are calibrated to the power level at the antenna input using the following procedure:

- The recorded power levels measured by the spectrum analyzer are provided in dBm relative to the analyzer input.
- The difference between the power level at the analyzer input and the power level at the antenna due to the losses of the RF cables, filters, and the gain associated with the Pre-selector versus frequency is recorded in a calibration file.
- The calibration values (in dB) where then added to the measured values (via an interpolation process) when plotting the spectrum data in this report.

Thus, the plotted power level values are the absolute value in dBm at the antenna input.

4. Spectrum Measurements

The section provides sample spectrum occupancy measurements and summary occupancy statistics.

4.1 Spectrum Occupancy Plots

Each of the spectrum occupancy plots in Figure 4 has three spectrum occupancy sub-plots. The upper sub-plot is the maximum power value versus frequency measured during the period. The power values are corrected for cable losses, filters, amplifier gain and attenuators (as described in the previous calibration section), and represent the received power level at the antenna terminals. The time in the plot title is the measurement start time.

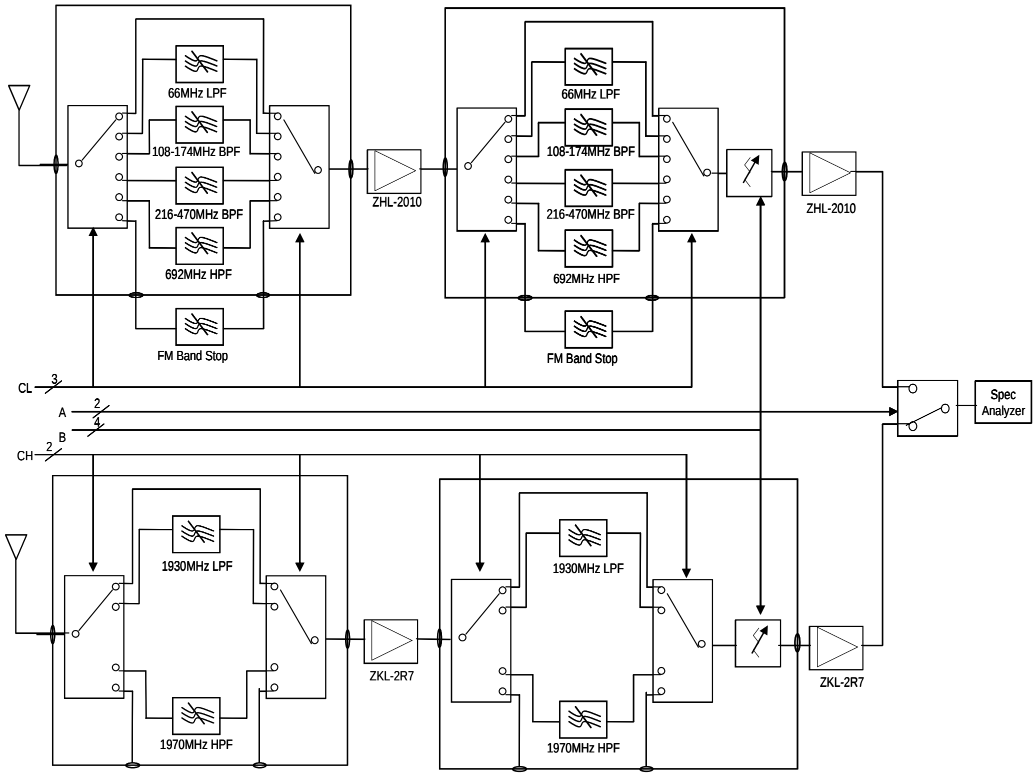


Figure 3: Pre-Selector block diagram.

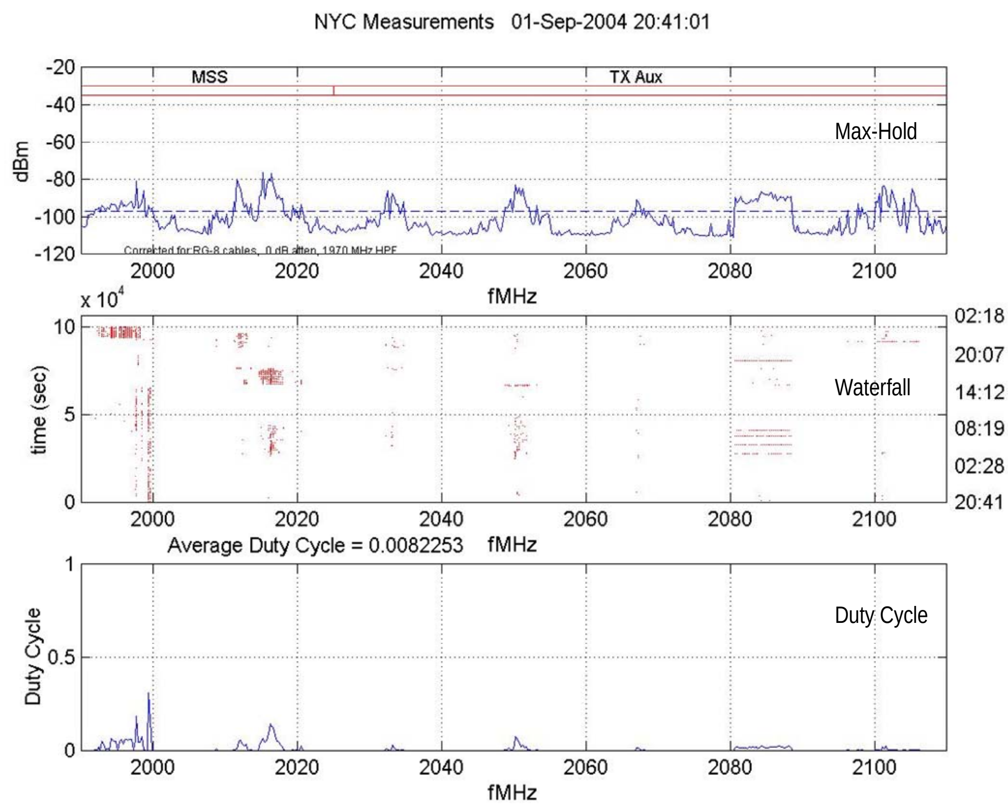


Figure 4: Spectrum occupancy in NYC on September 1 to September 3, 1990 MHz – 2110 MHz.

The middle sub-plot of Figure 4 is a waterfall-type plot with occupancy plotted versus time and frequency. Occupancy is determined when the power level exceeds a threshold. One overall threshold value was intentionally selected for each run, and is shown as a dotted line on the upper plot. Note that, in some cases, the noise level exceeds the threshold, causing inflated occupancy levels. This was not corrected because it would have been necessary to manually select the threshold for each plot. The time shown on the right side vertical axis of each figure is the measurement time.

The last sub-plot of Figure 4 is the fraction of time the signal is above the threshold versus frequency. A fraction of time value of "1" means that the signal was measured above the threshold for the entire duration of the measurement period. The title of this third plot contains the average duty cycle for the entire period.

Figure 4 shows the spectrum occupancy from 1990 MHz to 2110 MHz measured in New York City plot during a 36-hour period from 8:30 pm, September 1, 2004 to 2:18 am, September 3, 2004. It is notable that the Presidential Address given to the Republican National Convention by George W. Bush occurred during the second collection period at approximately 10:00 pm, September 2, 2004. This band is the TV Auxiliary Broadcast band and is used for sending video from remote locations to TV studios. The average duty cycle in this band was very low (0.82%).

Figure 5 shows the spectrum occupancy from 698 MHz to 806 MHz measured in New York City. This band is used by analog (NTSC) TV and digital (ATSC) TV signals. This band has a high average duty cycle (31%). Periods when the TV transmitters are turned off at night and when signal ducting occurs are evident. The off-periods are opportunities for a dynamic spectrum sharing system to utilize the spectrum. The ducting periods indicate that the propagation loses to distant regions are low and a dynamic spectrum sharing system would have to reduce its transmit power level to avoid causing interference to distant regions.

Figure 6 shows the spectrum occupancy from 698 MHz to 806 MHz measured at Tysons Corner, Virginia for a one hour period. The average duty cycle at this location (17%) was less that the NYC value. This was due to fewer TV transmitters in the Tysons Corner area, and it is due to the measurement system's antenna being only 3 meter high at the Tysons Corner instead of being on top of a tall building as was done in the NYC measurement.

4.2 Spectrum Occupancy Summary

The bar graph in Figure 7 provides the average of the occupancy in each location. The average occupancy over all of the locations is 5.2%. The maximum occupancy is 13.1% (New York City) and the minimum occupancy is at the National Radio Astronomy Observatory (1%). It is surprising that the occupancy ratio between the densely populated NYC area and a radio quiet zone (NRAO) is only thirteen. This is a clear indication that the present spectrum allocation process is inefficient. The low occupancy levels at all of the locations show that there is significant spectrum for a Dynamic Spectrum Sharing Radio (DSS) to use. In rural areas, there is enough unused spectrum for a DSS Radio to provide ten times the capacity of all existing wireless devices together.

The bar graph in Figure 8 provides the average of the occupancy in each band. The TV bands, especially the UHF 512 MHz to 806 MHz bands have the highest occupancy values. This is due to the high transmitter power level used (even weak stations that are not strong enough for reception are counted as occupied in this study) and it is due to the 100% transmit duty cycle. Many of the bands with 100's of MHz of spectrum have negligible spectrum occupancy Examples include the 1710 MHz to 1850 MHz band and the 1240 MHz to 1300 MHz band. Some of these bands contain GPS and other satellite downlink and parts of the bands may not suitable for DSS radios. We believe that this is rare and that a significant fraction of these bands can be shared. The 1990 MHz to 2100 MHz TV Aux band and other bands have low occupancy (a few percent) and also could be utilized by DSS radios.

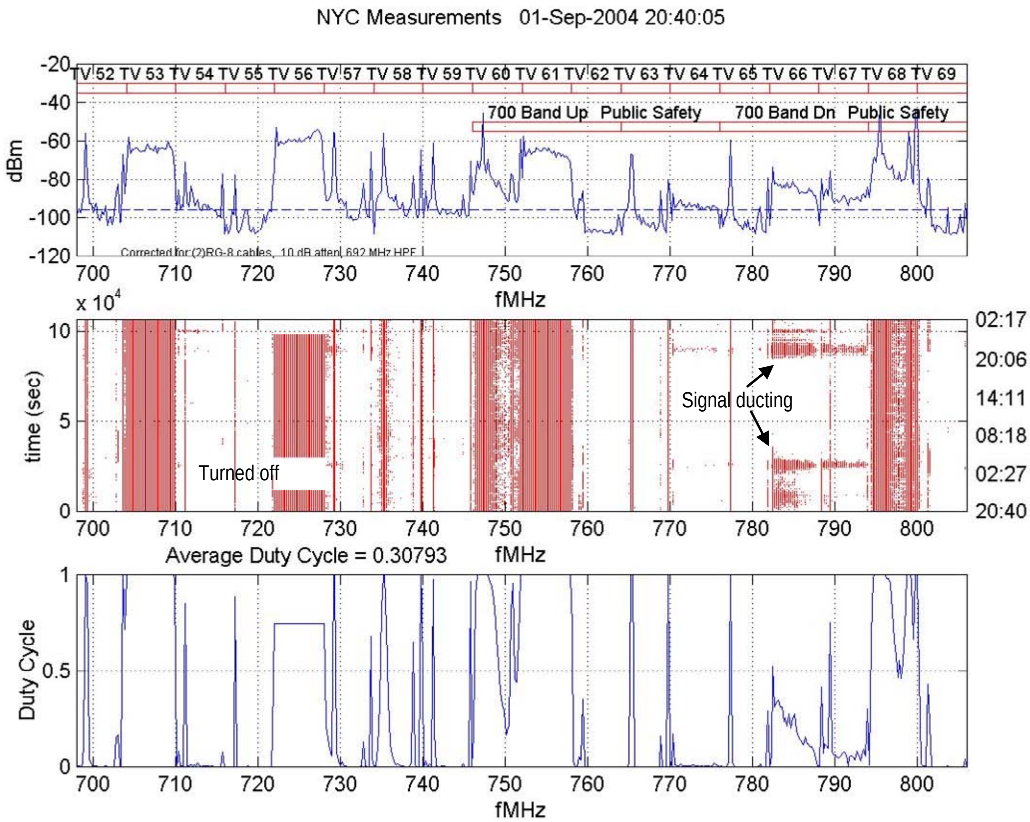


Figure 5: Spectrum occupancy in NYC on September 1 to September 3, 698 MHz – 806 MHz.

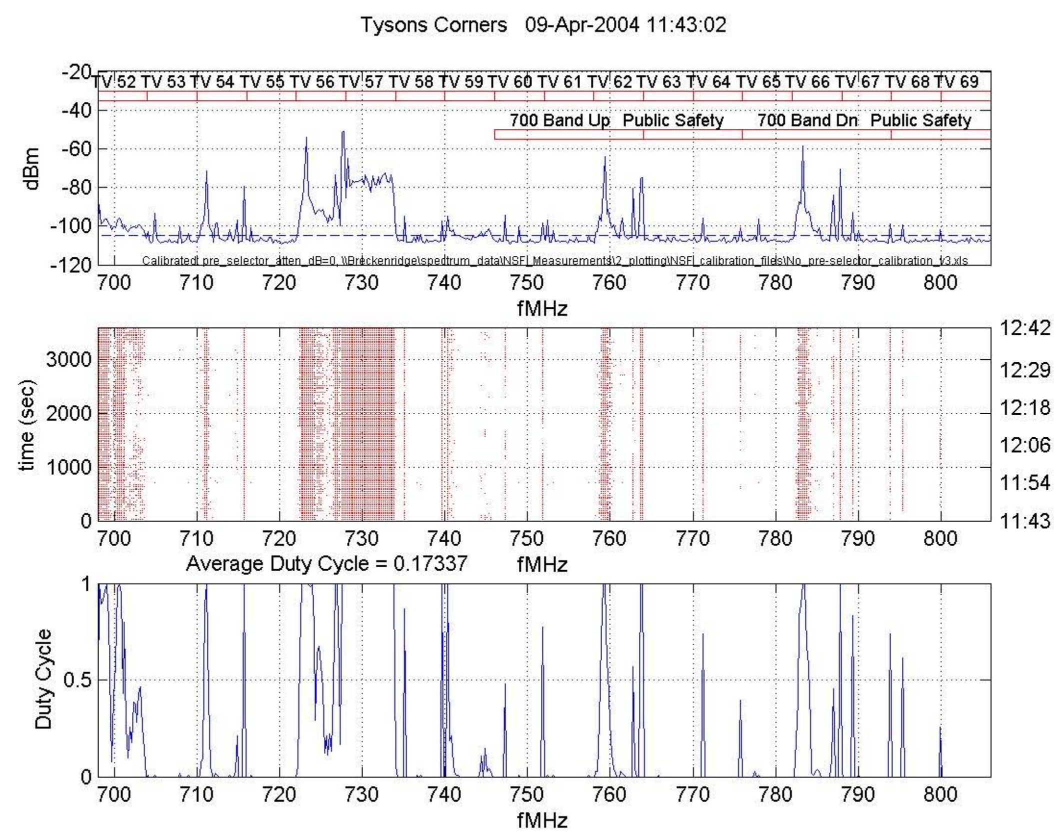


Figure 6: Spectrum occupancy in Tysons Corner, Virginia on April 9, 2004, 698 MHz – 806 MHz.

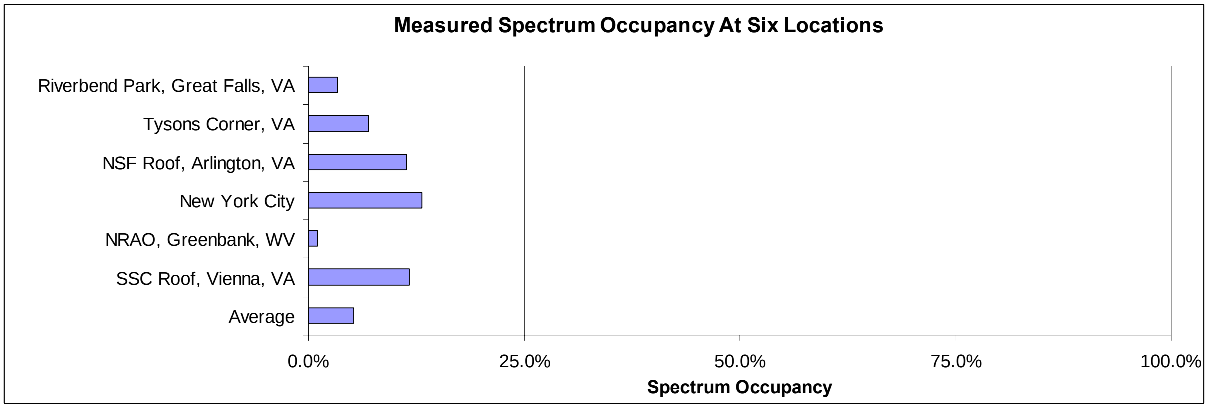


Figure 7: Average spectrum occupancy in each band.

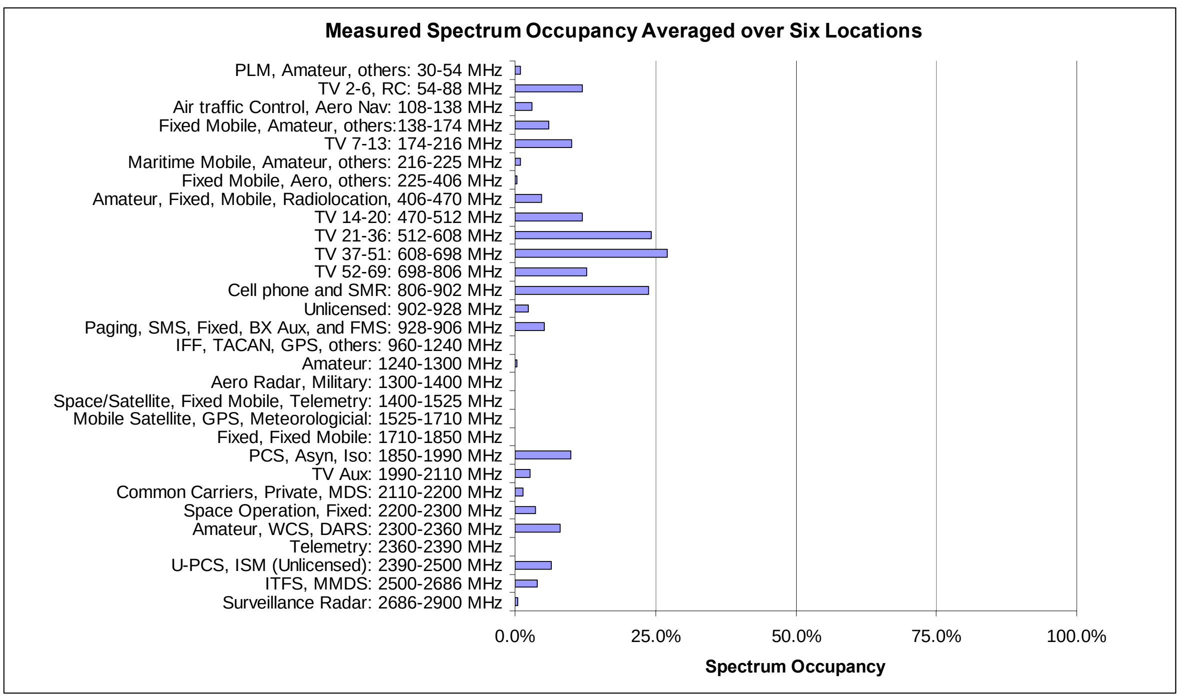


Figure 8: Spectrum occupancy in each band averaged over six locations.

4.3 Measurement Reports

A subset of the data from six locations was documented in individual reports [3-8] (www.sharedspectrum.com) . These reports include plots of the occupancy in each band and a detailed description of the equipment. These reports provide both absolute and comparative estimates of the measured spectrum occupancy in a variety of locations. Electronic copies of the measurements are available (contact the author).

Table 2 shows a summary of the occupancy from all six locations. The average occupancy over all of the locations is 5.2%. The maximum occupancy is 13.1% (New York City) and the minimum occupancy is at the National Radio Astronomy Observatory (1%).

Spectrum usage varies significantly band to band. This is especially true in urban areas when comparing the high occupancy bands (TV, Cell phone, and PCS) to the low occupancy bands (Amateur 1240-1300 MHz, Aero Radar/Military 1300-1400 MHz, and TV Aux 1900-2110). This non-uniform usage is ideal for DSS radios, because it simplifies the spectrum access problem. The spectrum available for sharing is in large, contiguous blocks. It is not scattered in unoccupied, short-term available channels that require a highly dynamic DSS radio to exploit.

Table 2: Spectrum occupancy in each band at each location

Start Freq (MHz)	Stop Freq (MHz)	Bandwidth (MHz)	Spectrum Band Allocation	Riverbend Spectrum Fraction Used	Riverbend Occupied Spectrum (MHz)	Tysons Spectrum Fraction Used	Tysons Occupied Spectrum (MHz)	NSF Roof Spectrum Fraction Used	NSF Roof Occupied Spectrum (MHz)	NYC Day 1 Spectrum Fraction Used	NYC Day 2 Spectrum Fraction Used	NYC Avg Spectrum Fraction Used	NYC Occupied Spectrum (MHz)	NRAO Spectrum Fraction Used	NRAO Occupied Spectrum (MHz)	SSC Roof Spectrum Fraction Used	SSC Roof Occupied Spectrum (MHz)	Average Occupied Spectrum (MHz)	Average Percent Occupied
30	54	24	PLM, Amateur, others	0.03895	0.93	0.00763	0.18	0.00217	0.05	0.04300	0.06250	0.05275	1.27	0.00045	0.01	0.00400	0.10	0.22	0.9%
54	88	34	TV 2 -6, RC	0.10593	3.60	0.11799	4.01	0.36654	12.46	0.52830	0.52080	0.52455	17.83	0.11056	3.76	0.10900	3.71	4.07	12.0%
108	138	30	Air traffic Control, Aero Nav	0.00744	0.22	0.02768	0.83	0.04066	1.22	0.05270	0.04030	0.04650	1.40	0.15485	4.65	0.10000	3.00	0.91	3.0%
138	174	36	Fixed Mobile, amateur, others	0.03372	1.21	0.07692	2.77	0.16865	6.07	0.17080	0.16980	0.17030	6.13	0.02745	0.99	0.07300	2.63	2.15	6.0%
174	216	42	TV 7-13	0.10339	4.34	0.11652	4.89	0.18890	7.93	0.77730	0.77950	0.77840	32.69	0.00220	0.09	0.18100	7.60	4.26	10.1%
216	225	9	Maritime Mobile, Amateur, others	0.00486	0.04	0.00842	0.08	0.01129	0.10	0.05860	0.05950	0.05905	0.53	0.00556	0.05	0.02300	0.21	0.08	0.9%
225	406	181	Fixed Mobile, Aero, others	0.00002	0.00	0.00371	0.67	0.00576	1.04	0.00530	0.00370	0.00450	0.81	0.01842	3.33	0.01300	2.35	0.68	0.4%
406	470	64	Amateur, Radio Geolocation, Fixed, Mobile, Radiolocation	0.02745	1.76	0.07243	4.64	0.10469	6.70	0.16610	0.14750	0.15680	10.04	0.00379	0.24	0.08100	5.18	3.07	4.8%
470	512	42	TV 14-20	0.13313	5.59	0.12160	5.11	0.29794	12.51	0.21140	0.21000	0.21070	8.85	0.00379	0.16	0.15700	6.59	5.00	11.9%
512	608	96	TV 21-36	0.26616	25.55	0.32736	31.43	0.49667	47.68	0.35520	0.34270	0.34895	33.50	0.04283	4.11	0.36400	34.94	23.33	24.3%
608	698	90	TV 37-51	0.23484	21.14	0.39980	35.98	0.47044	42.34	0.46160	0.46090	0.46125	41.51	0.00156	0.14	0.51300	46.17	24.35	27.1%
698	806	108	TV 52-69	0.07627	8.24	0.17337	18.72	0.20048	21.65	0.29580	0.30790	0.30185	32.60	0.00113	0.12	0.31300	33.80	13.79	12.8%
806	902	96	Cell phone and SMR	0.14260	13.69	0.41188	39.54	0.46398	44.54	0.46190	0.46450	0.46320	44.47	0.00017	0.02	0.40000	38.40	22.77	23.7%
902	928	26	Unlicensed	0.00000	0.00	0.03915	1.02	0.08687	2.26	0.22270	0.23460	0.22865	5.94	0.00004	0.00	0.01100	0.29	0.63	2.4%
928	960	32	Paging, SMS, Fixed, BX Aux, and FMS	0.03460	1.11	0.06708	2.15	0.10438	3.34	0.23640	0.24370	0.24005	7.68	0.02459	0.79	0.10000	3.20	1.68	5.2%
960	1240	280	IFF, TACAN, GPS, others							0.03560	0.04080	0.03820	10.70		0.00	0.00000	0.00	0.01	0.0%
1240	1300	60	Amateur	0.00139	0.08	0.00335	0.20	0.01509	0.91	0.00030	0.00010	0.00020	0.01	0.00012	0.01	0.00000	0.00	0.20	0.3%
1300	1400	100	Aero Radar, military	0.00022	0.02	0.00562	0.56	0.00718	0.72	0.02160	0.00130	0.01145	1.15	0.00000	0.00	0.00000	0.00	0.22	0.2%
1400	1525	125	Space/Satellite, Fixed Mobile, Telemetry	0.00000	0.00	0.00000	0.00	0.00083	0.10	0.01520	0.00050	0.00785	0.98	0.00000	0.00	0.00000	0.00	0.02	0.0%
1525	1710	185	Mobile Satellite, GPS L1, Mobile Satellite, Meteorologicial	0.00000	0.00	0.00000	0.00	0.00220	0.41	0.00240	0.00130	0.00185	0.34	0.00082	0.15	0.00000	0.00	0.07	0.0%
1710	1850	140	Fixed, Fixed Mobile	0.00000	0.00	0.00000	0.00	0.00137	0.19	0.02350	0.02540	0.02445	3.42	0.00000	0.00	0.00000	0.00	0.04	0.0%
1850	1990	140	PCS, Asyn, Iso	0.00044	0.06	0.12690	17.77	0.27102	37.94	0.33090	0.34430	0.33760	47.26	0.00001	0.00	0.19300	27.02	13.85	9.9%
1990	2110	120	TV Aux	0.00000	0.00	0.00000	0.00	0.00005	0.01	0.01910	0.00820	0.01365	1.64	0.00009	0.01	0.15900	19.08	3.18	2.7%
2110	2200	90	Common Carriers, Private Companies, MDS	0.00000	0.00	0.00000	0.00	0.00397	0.36	0.01820	0.01900	0.01860	1.67	0.00353	0.32	0.08100	7.29	1.28	1.4%
2200	2300	100	Space Operation, Fixed	0.00000	0.00	0.00000	0.00	0.00021	0.02	0.05270	0.06180	0.05725	5.73	0.00000	0.00	0.21400	21.40	3.58	3.6%
2300	2360	60	Amateur, WCS, DARS	0.00000	0.00	0.12693	7.62	0.17754	10.65	0.20220	0.20530	0.20375	12.23	0.10521	6.31	0.17300	10.38	4.83	8.0%
2360	2390	30	Telemetry	0.00000	0.00	0.00000	0.00	0.00000	0.00	0.06200	0.06420	0.06310	1.89	0.00004	0.00	0.00000	0.00	0.01	0.0%
2390	2500	110	U-PCS, ISM (Unlicensed)	0.00022	0.02	0.00074	0.08	0.12461	13.71	0.13470	0.15510	0.14490	15.94	0.00007	0.01	0.25700	28.27	7.04	6.4%
2500	2686	186	ITFS, MMDS	0.00000	0.00	0.00000	0.00	0.07046	13.10	0.10430	0.10420	0.10425	19.39	0.00137	0.26	0.16100	29.95	7.19	3.9%
2686	2900	214	Surveillance Radar	0.00000	0.00	0.00000	0.00	0.02123	4.54	0.02860	0.03090	0.02975	6.37	0.00288	0.62	0.00700	1.50	1.01	0.5%
Total		2850			87.62		178.24		292.57				373.97		26.14		333.06	149.52	
Total Available Spectrum					2570		2570		2570				2850		2570		2850	2850	
Average Spectrum Use (%)					3.4%		6.9%		11.4%				13.1%		1.0%		11.7%	5.2%	

5. Conclusions

This section provides the conclusions of the spectrum measurement project.

5.1 Measurements Show That There is Significant Spectrum "Whitespace"

The goal of this study was to determine the spectrum occupancy in each band at multiple locations. The bar graphs in Figure 7 and Figure 8 provide the average of the occupancy in each band and at each location. The average occupancy over all of the locations is 5.2%. The maximum occupancy is 13.1% (New York City) and the minimum occupancy is at the National Radio Astronomy Observatory (1%). These low occupancy levels show that there is significant spectrum for a Dynamic Spectrum Sharing Radio (DSS) to provide service. In rural areas, there is enough unused spectrum for a DSS Radio to provide ten times the capacity of all existing wireless devices together.

The average occupancy varies band by band significantly. The highest average occupancy is in the TV bands (15%-25%). The PCS 1990-2110 MHz band has 9.9% occupancy. The Cell 806-902 MHz band has 23.7% occupancy. There are many bands with very low occupancy such as the Amateur 1240-1300 MHz band, the "Aero Radar", Military 1300-1400 MHz band, the Fixed/Mobile 1710-1850 MHz band, the AV Aux 1990-2110 MHz band ad others. While there are some hard to detect users in parts of these bands, we believe that they are empty because the usage in measurement areas is low. These bands are excellent candidates for DSS use. The unevenness in usage is also an indication that the current method to allocate spectrum is inefficient.

5.2 A Low Agility, Contiguous DSS Radio Waveform Can Provide High Utility

The detail spectrum occupancy plots show that there is a significant amount of spectrum available in continuous blocks that are 1 MHz and wider. An example is the TV Auxiliary band, where there is large, contiguous spectrum segments not used for hours at a time. This shows that a frequency agile, non-contiguous waveform is not necessary for a DSS Radio. The key DSS technical challenge is to reliably detect signals with high sensitivity to decide what channels can be used with out causing interference to the existing users.

5.3 An Extensive Spectrum Occupancy Dataset Is Available for Future Analysis

We have collected an extensive spectrum occupancy data set that is available for others to investigate. The areas for further investigation include analysis of the spectrum gaps statistics, the transmitter mobility and number (using the amplitude probability distribution), the signal bandwidths, and other existing user parameters. It is critical for DSS radios to be able to know what transceiver types actually use spectrum and how they are operated. An additional area to investigate is to compare the NTIA/FCC allocation tables and assignment databases with the measurements to determine what spectrum is used by GPS and other hard to detect signals to understand why band-by-band the measurement occupancy is so low.

6. Acknowledgments

This project was funded by the National Science Foundation (NSF) Award Number: ANI-0335272 under Subcontract No. FY2004-013 to the University of Kansas Center for Research, Inc. under the National Radio Research Testbed (NRNRT) project.

7. References

[1] "Broadband Spectrum Survey at Denver, Colorado" Frank H. Sanders, Vince S. Lawrence, NTIA Report 95-321, September, 1995.

[2] "Broadband Spectrum Survey at San Francisco, California", Frank H. Sanders, Bradley J. Ramsey, Vincent S. Lawrence, NTIA Report 99-367, May-June 1995.

[3] "Spectrum Occupancy Measurements, Location 1 of 6: Riverbend Park, Great Falls, Virginia", Mark A. McHenry, Karl Steadman, Shared Spectrum Company Report, August, 2005.

[4]. "Spectrum Occupancy Measurements, Location 2 of 6: Tyson's Square Center, Vienna, Virginia, April 9, 2004", Mark A. McHenry Karl Steadman, Shared Spectrum Company Report, August, 2005.

[5] "Spectrum Occupancy Measurements, Location 3 of 6: National Science Foundation Building Roof, April 16, 2004, Revision 2", Mark A. McHenry, Shyam Chunduri, Shared Spectrum Company Report, August, 2005.

[6] "Spectrum Occupancy Measurements, Location 4 of 6: Republican National Convention, New York City, New York, August 30, 2004 - September 3, 2004, Revision 2", Mark A. McHenry, Dan McCloskey, George Lane-Roberts, Shared Spectrum Company Report, August, 2005.

[7] "Spectrum Occupancy Measurements, Location 5 of 6: National Radio Astronomy Observatory (NRAO), Green Bank, West Virginia, October 10 - 11, 2004, Revision 3" Mark A. McHenry Karl Steadman, Shared Spectrum Company Report, August, 2005.

[8] "Spectrum Occupancy Measurements, Location 6 of 6: Shared Spectrum Building Roof, Vienna, Virginia, December 15-16, 2004", Mark A. McHenry, Dan McCloskey, Jim Bates, Shared Spectrum Company Report, August, 2005.