Public Safety Statement of Requirements

for Communications & Interoperability

The SAFECOM Program
Department of Homeland Security

Volume II, Version 1.0
August 18, 2006

SAFECOM

Quantitative

SAFECOM

Defining the Problem

Effective interoperable communications can mean the difference between life and death. Unfortunately, inadequate and unreliable communications have compromised emergency response operations for decades. Emergency responders—police officers, fire personnel, emergency medical services—need to share vital information via voice and data across disciplines and jurisdictions to successfully respond to day-to-day incidents and large-scale emergencies. Responders often have difficulty communicating when adjacent agencies are assigned to different radio bands, use incompatible proprietary systems and infrastructure, and lack adequate standard operating procedures and effective multi-jurisdictional, multi-disciplinary governance structures.

Background

SAFECOM is working with the emergency response community and Federal partners to develop solutions to address these interoperable communications challenges. With its Federal partners, SAFECOM provides research, development, testing and evaluation, guidance, tools, and templates on communications-related issues to local, tribal, state, and Federal emergency response agencies. The scope of community-oriented SAFECOM services is broad, and includes more than 60,000 local and state emergency response agencies and organizations. Federal customers include agencies engaged in emergency response disciplines—law enforcement, firefighting, public health, and disaster recovery—and agencies that provide funding and support to local and state emergency response organizations.

A communications program of the Department of Homeland Security's Office for Interoperability and Compatibility, SAFECOM is managed by the Science and Technology Directorate.

Practitioner-Driven Approach

SAFECOM is committed to working in partnership with local, tribal, state, and Federal officials in order to serve critical emergency response needs. In keeping with its unique "bottom-up" approach, SAFECOM has developed a governance structure that promotes the valuable input of emergency responders and policy makers and leaders.

Strategy

- Promote a system of systems approach through the use of standards-based communications equipment.
- Encourage establishment of governing bodies to foster a culture of cooperation and information sharing across agencies and jurisdictions.
- Support prioritization and funding of interoperability among local, state, tribal, and Federal leadership.
- Advance standardization of training and exercise programs.
- Support daily use of interoperable equipment throughout regions.

Long-Term Goals

- Achieve a system-of-systems environment supported by communications standards, tools, and best practices.
- Assist coordination of funding assistance through tailored grant guidance to maximize limited resources available for emergency response communications and interoperability.
- Pilot tools and methods as national models for emergency response at the rural, urban, state, and regional levels.
- Provide policy recommendations to promote efficiency in emergency response communications.

Publication Notice

Abstract

This document contains the assembled requirements for a system of interoperable public safety communications across all local, tribal, state, and Federal "first responder" communications systems.

Change Log

Version	Date	Changes
1.0 Draft	April 24, 2006	Initial Document (speech only)
1.0 Draft	July 7, 2006	Initial Document—one of three (speech, video, network) Volume II documents broken out for review. Intro, References, and Acronyms are general to speech, video, and network pieces which will become one Volume II document at publish time.
1.0	August 18, 2006	Review input and copyedit changes have been incorporated.

Acknowledgements

The SAFECOM Program extends its sincere appreciation to the many public safety practitioners, individuals, and government organizations that directly contributed to the creation of the Public Safety Statement of Requirements (PS SoR) for Communications and Interoperability.

Contact Information

Please address comments or questions to: The SAFECOM Program
P.O. Box 57243
Washington, D.C. 20073

Telephone: 1-866-969-SAFE
Program E-mail: safecom@dhs.gov

Director: David Boyd, Ph.D.

Introduction

> *"In times of emergencies, the public looks to government, particularly their Public Safety officials, to act swiftly and correctly, and do the things which must be done to save lives, help the injured, and restore order. Most disasters occur without warning, but people still expect a rapid and flawless response on the part of government. There is no room for error. Whether a vehicle accident, crime, plane crash, special event, or any other Public Safety activity, one of the major components of responding to and mitigating a disaster is wireless communications. These wireless communications systems are critical to Public Safety agencies' ability to protect lives and property, and the welfare of Public Safety officials."*

This statement comes from the highly regarded *Public Safety Wireless Advisory Committee (PSWAC) Final Report*, presented to the Chairman of the Federal Communications Commission (FCC) and the Administrator of the National Telecommunications and Information Administration (NTIA) in September 1996.[1] The PSWAC Final Report defined and documented critical public safety wireless communication needs in 1996, and projected anticipated needs through the year 2010. The report focused on the requirements for communications resources and the radio frequency spectrum to support those requirements. While the report mentioned the crucial need to promote interoperability, its emphasis was clearly on the necessity of taking immediate measures to alleviate spectrum shortfalls. Fortunately, for public safety and for the benefit of all Americans, the report spurred the allocation of precious spectrum for use by public safety practitioners.

Unfortunately, the communication challenges for those working on the front lines in public safety have not been eliminated. In fact, at a time when more attention is being paid to interoperability among different disciplines and jurisdictions within the community, there still exists fundamental communication deficiencies within disciplines and jurisdictions as practitioners strive to perform the most routine and basic elements of their job functions. Agencies must be "operable," meaning they must have sufficient wireless communications to meet their everyday internal requirements before they place value on being "interoperable," meaning being able to work with other agencies.

This document, the *Public Safety Statement of Requirements (PS SoR) for Public Safety Communications and Interoperability*, is the natural follow-on to the PSWAC Final Report, but differs in three ways, as follows:

- First, the PS SoR is not keyed to the issue of spectrum allocation, but focused on public safety requirements from a broader perspective. Operational and functional requirements delineated in the PS SoR are not based on a particular approach or technology.

- Second, the PS SoR was developed eight years after the PSWAC Final Report was published. While the Final Report did not explicitly identify specific technological approaches along with the stated requirements, it is important to realize that advances in technology have helped to fashion the way practitioners think about their jobs over the years. Because practitioners expect more from technology today, their needs and desires have been affected, sometimes subtly, by industry advances and solutions that exist in today's commercial and consumer world. Additionally, current technological advances promote technically advanced thinking about what the practitioner may be able to expect 15 years from now. For instance, the possibility that technology refresh cycles could

---

1. In 1994, the FCC and NTIA established PSWAC to evaluate the wireless communications needs of local, tribal, state, and Federal public safety agencies through the year 2010, identify problems, and recommend possible solutions.

be dramatically reduced for public safety based on these advances is extremely attractive. That said, the methodologies used and the general projections made in the PSWAC Final Report remain as valid today as when they were first published. Based on the rapid changes and potential of technology, the PS SoR addresses current requirements and future requirements for the next five to 20 years.

- Third, the PS SoR emphasizes the "information" aspects of communications; that is, the need for the wireless exchange of data, video, and other non-voice mediums. The need for voice communications was clearly made in the PSWAC Final Report, as well as the need for additional bandwidth for other data resources. The PS SoR defines the information requirements of public safety practitioners more explicitly to guide how practitioners will use information resources in the field in mission-critical situations.

Scope

The PS SoR is currently a two-volume set. This volume, Volume II, provides detailed quality of service methods of measurement for the applications and services identified in Volume I of the PS SoR, along with network parameters to specify the minimum acceptable performance of public safety communications systems carrying these services. This document provides recommended performance parameter values for specific public safety applications. In addition, this document provides a common understanding between manufacturers and practitioners for specifying and maintaining these applications and services.

The initial application of this document is for mission-critical speech and video services, in addition to specifying network performance parameters to meet these applications' quality of service needs. Future revisions of this document will provide detailed quality of service metrics for the balance of the applications and services.

Intended Audience

The PS SoR focuses on the functional needs of public safety first responders—Emergency Medical Services (EMS) personnel, firefighters, and law enforcement officers—to communicate and share information as authorized when it is needed, where it is needed, and in a mode or form that allows the practitioners to effectively use it. The communications mode may be voice, data, image, video, or multimedia that includes multiple forms of information.

Because functional requirements are the focus of the PS SoR, it does not specify technologies or business models (i.e., whether requirements should be addressed through owned products and systems or via commercial services). Similarly, the PS SoR does not specify infrastructure, except to note that consistent with first responder operations, it is assumed that terminal links to and from practitioners are wireless unless stated otherwise.

The PS SoR addresses a number of complementary objectives. Most importantly, it is rooted in the goal of improving the ability of public safety personnel to communicate among themselves, with the non-public safety agencies and organizations with whom they work, and with the public that they serve. The PS SoR can also assist the telecommunication interoperability and information-sharing efforts by and among local, tribal, state, and Federal government agencies, and regional entities, by delineating the critical operational functions and interfaces within public safety communications that would benefit from research and development investment and standardization.

The PS SoR can assist Federal programs that work with public safety practitioners to facilitate wireless interoperability at all government levels to develop a comprehensive vision for public safety communications that satisfies the defined needs. This vision can be reinforced by developing Federal grant

programs that promote government research and development, as well as investment in communications equipment and systems, in a manner consistent with the PS SoR.

The PS SoR provides information that can assist the communications industry to prioritize its research and development investment and product and service development strategies so that they are aligned with public safety communications needs.

The PS SoR is intended to be fully consistent with the National Incident Management System[2] as defined by the Federal Emergency Management Agency (FEMA) in the Department of Homeland Security. Any inconsistency between this document and NIMS is a discrepancy, and will be addressed in later version of the document.

Finally, the PS SoR can be used to clearly identify public safety operational issues so that discussions regarding existing and proposed regulations and laws can be dealt with expeditiously by regulatory and legislative bodies.

Statement of Requirements Organization

Section 1   **Measuring Speech Transmission Performance** describes the speech performance parameters.

(*See Section 1, "Measuring Speech Transmission Performance," on page 1*.)

Section 2   **Mission-Critical Speech Transmission Requirements** lists requirement values for speech performance parameters.

(*See Section 2, "Mission-Critical Speech Transmission Requirements," on page 7*.)

Section 3   **Measuring Video Performance** describes video performance parameters.

(*See Section 3, "Measuring Video Performance," on page 13*.)

Section 4   **Tactical Video Requirements** lists requirement values for video performance parameters.

(*See Section 4, "Tactical Video Requirements," on page 25*.)

Section 5   **Reference Model for Network Performance** describes a path-based reference model.

(*See Section 5, "Reference Model for Network Performance," on page 29*.)

Section 6   **Measuring Network Performance** describes the methods used to measure network performance.

(*See Section 6, "Measuring Network Performance," on page 41*.)

---

2.   Developed by the Secretary of Homeland Security at the request of the President, the National Incident Management System (NIMS) integrates effective practices in emergency preparedness and response into a comprehensive national framework for incident management. The NIMS will enable responders at all levels to work together more effectively and efficiently to manage domestic incidents no matter what the cause, size or complexity, including catastrophic acts of terrorism and disasters.

Section 7   **Network Requirements** lists requirement values for network performance parameters.

(See Section 7, "Network Requirements," on page 51.)

Appendix A   **Glossary and Acronyms** lists terminology and acronyms used in this document.

(See Appendix A, "Glossary and Acronyms," on page 75.)

Appendix B   **Audio Measurement Methods and Tools** provides additional details on the laboratory study described in Section 1, and used to set requirements in Section 2.

(See Appendix B, "Audio Measurement Methods and Tools," on page 79.)

Appendix C   **Video Acquisition Measurement Methods** describes how to measure the video acquisition parameters identified in this document.

(See Appendix C, "Video Acquisition Measurement Methods," on page 91.)

Appendix D   **Video Quality Experiment PS1** describes the experimental results from the narrow field of view tactical video application.

(See Appendix D, "Video Quality Experiment PS1," on page 127.)

Appendix E   **Network Measurement Methods** describes how to measure the speech and video acquisition parameters identified in this document.

(See Appendix E, "Network Measurement Methods," on page 153.)

Appendix F   **References** identifies print, standards, and online references of this document.

(See Appendix F, "References," on page 169.)

Contents

Figures

Tables

1　Measuring Speech Transmission Performance

1.1　Mission-Critical Speech Transmission Services

Services that provide speech transmission between two locations to support public safety operations have become firmly established over the decades. A classic example is a land-mobile radio (LMR) system that allows mobile practitioners to speak with each other, and with a dispatcher. These systems are typically half duplex and require each practitioner to "push to talk." Analog systems were once pervasive, and those systems have been migrating to digital operation in more recent years.

The public safety communications community is looking towards a future family of full-duplex public safety speech transmission services. This family of services would be securable, robust, and scalable for operation over personal area networks (PAN), incident area networks (IAN), jurisdictional area networks (JAN), and extended area networks (EAN). The services would transmit speech with a quality that is suitable for mission-critical communications.

This section and Section 3 describe the technical components and an initial set of some of the minimum performance requirements necessary to provide a future family of full-duplex public safety speech transmission services. Packetizing speech as digital data for wired and wireless transmission supports the following public safety communications goals:

- Highly flexible system architectures
- Different operating modes for Personal Area Networks (PAN), Incident Area Networks (IAN), Jurisdiction Area Networks (JAN), and Extended Area Networks (EAN)
- Speech, video, and data communications across networks

1.2　Reference Model for Speech Performance Measurements

Figure 1 provides a high-level depiction for one direction of a generic full-duplex packetized digital speech transmission system.

Figure 1:   Digital Speech Transmission Reference Model
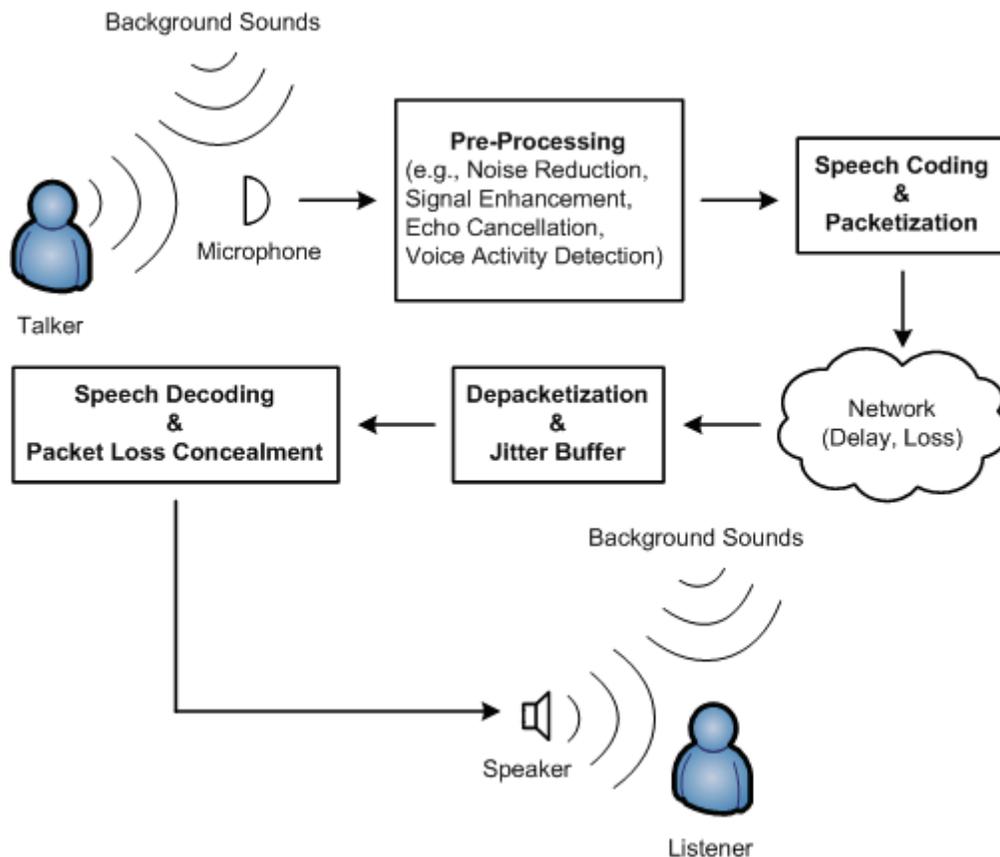
The quality of the speech transmission delivered by such a system will depend on equipment factors including, but not limited to the following elements:

- Microphone

- Pre-processing (noise reduction, signal enhancement, echo cancellation, speech-activity detection) algorithms

- Speech coding algorithm

- Packet size

- Jitter-buffer size and playout algorithm

- Packet loss concealment algorithm

- Earpiece or loudspeaker

In addition, the quality of speech transmission will depend on the following operational factors including, but not limited to:

- Network delay statistics and network loss statistics

- Background sound types and levels at each end

- Characteristics of the talkers, listeners, and messages

This section discusses these factors and provides some related choices. Appendix B provides a detailed description of description of the laboratory study conducted for this PS SoR to determine an initial set of requirements for network loss statistics, such that speech delivered is "suitable for mission-critical speech communications," as judged by public safety practitioners. The results necessarily reflect a spread of opinions among the practitioners. Section 2 provides the resulting set of initial requirements which are valid under the necessary conditions and constraints presented in Section 1 and Appendix B.

1.3 Speech Transmission Factors

In the most general case of speech transmission using packetized transmission of digital data, there are many factors to consider. It was not possible to consider all possible combinations of these in a single laboratory study. Experts did, however, consider each factor independently and made informed choices to arrive at the most relevant and practical set of factors to include in the laboratory study. These choices reflect the need to balance the goal of creating a realistic and relevant environment against the goal of developing a well-controlled and performed laboratory study. The next subsections discuss each of the speech transmission factors and the related choices.

1.3.1 Speech Coding

Public safety operations are often conducted in environments with significant background sound. While noise-canceling microphones can often reduce the level of undesired background sounds entering the speech coder, they cannot always eliminate them. In some cases, the background sounds may contain important information, and accurate transmission of those sounds, along with speech, can be desirable.

Operations can be enhanced by knowledge of exactly who has transmitted speech and what emotional state is represented by his or her speech.

Thus for future services, speech coders that can very accurately transmit the fine nuances (e.g. talker-specific attributes and emotional-state specific attributes) of a human voice or voices along with arbitrary background sounds are necessary. (Note that transmitting both speech and background sounds allows for the optional use of signal processing-based noise reduction techniques at the receiving location, if and when the practitioner at the receiving location desires.) Many speech coders rely on mathematical models of a single speech signal to achieve bit rate reduction. Such coders can attain very low bit rates, but are not intended to accurately transmit fine speech and background sound nuances, and in general, cannot accurately transmit such nuances.

The highest fidelity is attained by directly coding the actual speech-plus-background sound waveform, making few, if any, assumptions about that waveform. The companded pulse-code modulation (PCM) speech coders specified in ITU-T Recommendation G.711[3] give one efficient approach. These speech coders have very low encode and decode complexity and delay and operate at 64 kbps. Appendix I of Rec. G.711 includes a packet loss concealment (PLC) algorithm.[4] This algorithm minimizes distortions caused by lost channel data, and is well-suited to networks where channel data is packetized and some packets may be lost. As detailed in Appendix B, the laboratory study focuses on G.711 speech coding accompanied by this PLC algorithm.

---

3. ITU-T Recommendation G.711, "Pulse Code Modulation (PCM) of Voice Frequencies," Geneva, 1988.
4. ITU-T Recommendation G.711 Appendix I, "A High-Quality Low-Complexity Algorithm for Packet Loss Concealment with G.711," Geneva, 1999.

Other methods of making G.711 speech coders more robust to packet loss include PLC algorithms and multiple description, or diversity, coding algorithms. Development work in this field continues. However, G.711 Appendix I is the only approach known to be fully disclosed and codified in a formal way, and thus it is most suitable for the laboratory study. If other algorithms are eventually demonstrated to provide higher robustness, it may then be possible to relax the packet loss requirements given in Section 2.

1.3.2   Packetized Transmission of Digitized Speech Data

In packetized data transmission, delivery of all packets within a specified time window is not always guaranteed. Packets that fail to arrive within the required time may be considered lost. In addition, packets that are not lost may experience dissimilar transmission delays. This means that a stream of packets sent at uniform time intervals may be received at non-uniform time intervals. Since speech decoders generally require data at uniform time intervals, a jitter buffer is often used to provide such a data stream, at the cost of some additional fixed delay. For a given network configuration, increasing network traffic can increase network congestion which in turn can lead to an increase in lost packets and a wider variation in transmission delays. The laboratory study requires some model of these processes.

Network models with varying levels of detail and complexity are available. More detailed models may better reproduce specific network behaviors, while less detailed models can show wider applicability. Given that no specific network details are presently available, it seems prudent to use a very basic and general model for the combined effects of the network and the jitter buffer. Thus this study treats the network and the jitter buffer together as a single black box that can be parameterized (for at least tens of seconds) by a pair of packet loss parameters. These are fundamental properties and thus they provide a basic yet relevant model.

The two packet loss parameters are packet loss ratio and packet loss correlation. When packet loss correlation is zero the packet loss process is random. As packet loss correlation is increased, the loss of packets becomes more bursty, and it becomes more likely that multiple packets will be lost in succession. In practice, packetized data networks can exhibit random or bursty packet loss patterns. The model used in the laboratory study to represent the network and jitter buffer is detailed in Appendix B. The model is applied to two different packet sizes; those that contain data representing 10 ms of speech signal, and those that contain 40 ms.

1.3.3   Transducers

The determination of suitable acoustical and electrical properties for transducers (e.g., microphones, loudspeakers, and earpieces) used in public safety operating locations is a potential topic for future studies. Such studies will need to carefully account for the range of acoustic environments that may be present at these locations.

In the present laboratory study we assume that these transducers are not limiting performance, rather G.711 speech coding and packet loss are the limiting factors in attaining "suitability for mission-critical voice communications." We use studio-quality microphones and loudspeakers in the laboratory work to ensure that these transducers do not limit performance.

1.3.4   Voice Activity Detection

Voice activity detection (VAD) may be used in full-duplex speech transmission systems to prevent a speech coder from generating a full rate data stream when no one is talking. The challenges in VAD include accurate detection of voice activity, particularly the starts of utterances, and especially in the

presence of significant background sounds. VAD is not appropriate for cases that require transmission of the most accurate representation of the acoustic environment.

The appropriateness of VAD and the consequences of imperfect VAD operation is a potential subject for future investigations. The present laboratory study does not use any VAD and thus the results are not confounded by issues of VAD performance.

1.3.5   Echo Control

In full-duplex communications, there is the possibility a practitioner will be distracted by an echo of his or her own voice. These echoes can be of electrical or acoustical origin. Echoes can exacerbate the effects of delay since increased delay makes echoes more audible and annoying. In systems where delay and echo levels can cause annoyance, echo cancellers are typically deployed to minimize the levels, and thus the annoyance, of the echoes. The laboratory study supporting the requirements in Section 2 does not include any sources of echoes, nor any echo control devices. Thus the results are not confounded by echo or echo control issues.

1.3.6   Encryption

Some applications require encryption to keep speech transmissions secure. The laboratory study described in Appendix B assumes that any encryption system used is transparent to the data stream, even when data packets are lost. If an encryption system requires additional data handling capabilities, or increases the data transmission delay, this must be considered separately.

1.3.7   Mouth-to-Ear Delay

Mouth-to-ear delay identifies the elapsed time between a sound leaving a talker's mouth and arriving at a listener's ear. (This is different from call-setup delay, which is the time associated with the initial establishment of communications between the parties involved.) In face-to-face communications and in many speech transmission systems, the mouth-to-ear delay is either imperceptible or negligible. In other systems, the mouth-to-ear delay can be significant, and can even be large enough to impair the communications attempted by the two parties. For the speech transmission systems considered here, the mouth-to-ear delay is likely to be dominated by the following types of delays:

- Packetization delays
- Network queuing delays
- Network transmission delays
- Jitter buffer delays

Other contributions to the mouth-to-ear delay that are likely to be negligible in most implementations include G.711 encoding and decoding delays and the mouth-to-microphone and loudspeaker-to-ear acoustic propagation times.

As a practical matter, it is necessary to treat mouth-to-ear delay separately from other speech transmission factors in the present study. In potential future studies, interactions between mouth-to-ear delay and other factors might be evaluated through laboratory studies of human subjects engaged in conversation tasks rather than the listening task employed in the Appendix B laboratory study. Such studies require real-time implementations of each system that is to be evaluated, and each real-time implementation must support

speech traffic in two directions. These studies are thus significantly more complex than studies that use listening tasks.

1.3.8   Background Sound

Background sound in public safety operations is often, if not always, present at some level. Background sound types and levels can vary greatly between locations (e.g., dispatch center, parked patrol car, fire truck responding to alarm). In general, the background sound environment at the talking location and the environment at the listening location have the potential to influence the perception of transmitted speech. A rigorous exploration of the variables of talking location background sounds, as well as listening location background sounds could fill numerous potential future studies. In the Appendix B laboratory study, a single background sound level and type is simulated for the talking location. A single background sound type and level type and two different sound levels are simulated at the listening location. See Appendix B for full details.

2   Mission-Critical Speech Transmission Requirements

This section outlines a requirement for mouth-to-ear delay in speech transmission systems used for mission-critical communications. This requirement is based on existing results in the literature. The section also provides a set of packet loss requirements that when met, are expected to provide speech transmission that is suitable for mission-critical communications. Each requirement is related to an estimated percentage of practitioners that will find the results to be suitable for mission-critical communications. These requirements are based on the laboratory study described in Appendix B.

2.1   Mouth-to-Ear Delay

The issue of mouth-to-ear delay is addressed here primarily through information contained in ITU-T Recommendations G.107 (the E-Model)[5] and G.114.[6] These recommendations reflect what is known about the relationship between mouth-to-ear delay, and practitioner satisfaction in the telecommunications context. They are based on extensive research conducted by various telephone operating companies and administrations around the world, over a period of years.

Figure 2 expands on Figure 1 of G.114. The solid line on this graph matches that of G.114 Figure 1, is generated by equations given in G.107, and relates mouth-to-ear delay to the "Transmission Rating" factor, R. This solid line treats the case where mouth-to-ear delay and G.711 encoding are the only significant factors impairing speech transmission. G.107 further indicates that $90 \leq R$ will result in "very satisfied" users, $80 \leq R < 90$ will result in "satisfied" users, and $70 \leq R < 80$ will result in "some users dissatisfied."

---

5. ITU-T Recommendation G.114, "One-Way Transmission Time," Geneva, 2003.
6. ITU-T Recommendation G.107, "The E-Model, a Computational Model for Use in Transmission Planning," Geneva, 2005.

Figure 2:   Rating as Delay Varies for Two G.711-Based Speech Transmission System
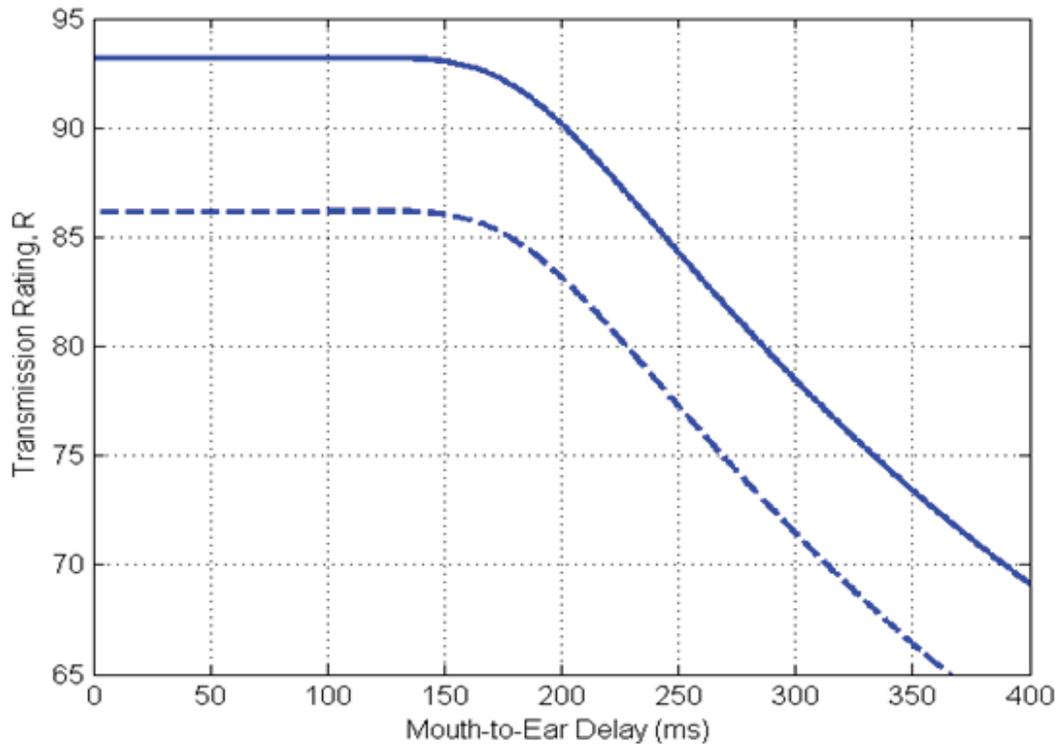
The dashed line in Figure 2 is also generated by the equations given in G.107, but for the case of packetized G.711 speech coding, with 10 millisecond (ms) packets, G.711 Appendix I PLC, and a random packet loss ratio of 2 percent. This result also requires the use of a packet loss robustness factor, found in Appendix I of G.113.[7] This second curve is simply a shifted version of the first curve, and this is a visual manifestation of the key underlying principle used in G.107; "psychological factors on a psychological scale are additive." In the context of mission-critical speech transmission requirements, the principle indicates that perceived speech transmission degradation due to mouth-to-ear delay and due to G.711 coding, packet loss, and PLC combined, are additive.

Based on the extensive research behind G.107, this principle is expected to hold for all speech transmission systems included in this study. Thus, each system-specific curve showing R versus mouth-to-ear delay is simply a downward-shifted version of the original curve, with greater downward shifts associated with more highly impaired systems. The key feature of interest, common to all of these curves is that they show nearly no negative effects of mouth-to-ear delay until that delay is about 150 ms.

Based on the information available at this time, the largest permissible mouth-to-ear delay for mission-critical communications is 150 ms, and this assumes that no audible echoes are present. The use of any greater value would first require laboratory evaluation (using conversation tasks rather than listening tasks) of the combined effect of that delay and the other impairments allowed by the packet loss requirements given in Section 2.2.

---

7.   ITU-T Recommendation G.113 Appendix I, "Provisional Planning Values for the Equipment Impairment Factor Ie and Packet-Loss Robustness Factor Bpl," Geneva, 2002.

We can restate this conclusion and further highlight the logic behind it using the language of additive degradations introduced above. When meeting the packet loss requirements given in Section 2.2 it is possible to "use up the entire degradation budget." No portion of the degradation budget remains for mouth-to-ear delay. This means that mouth-to-ear delay must remain in the range that does not add any degradation (i.e., 0 to 150 ms.). It is natural to ask if the degradation budget can be used differently; one might seek to use a less stringent delay requirement coupled with more stringent packet loss requirements. The definitive answer can only be found through laboratory studies of human subjects engaged in conversation tasks where the relevant mouth-to-ear delay and the packet loss characteristics are simulated in real time for both directions of the conversation.

It is also natural to ask if the mathematical tools provided with G.107 could provide an alternative to the required laboratory studies. At this time the answer is no. G.107 and the associated recommendations do not cover all of the higher packet loss ratio or non-random packet loss cases of interest, nor do they cover the case of 40 ms G.711 packets. An additional complication arises from the fact that at present, there is no well-established relationship between the R values produced by G.107 and the notion of "suitable for mission-critical communications."

Finally, recall from Section 1.3.7 that for the speech transmission systems considered here, the mouth-to-ear delay is likely to be dominated by packetization delay, network queuing delays, network transmission delays, and jitter buffer delay. In other words, the 150 ms delay budget must be allocated among these various sources of delay.

2.2 Packet Loss

The packet loss requirements given in Table 35, Table 36, and Table 37 are based on analysis of 12,320 votes collected in the laboratory study, as described in Appendix B. We view these votes to be samples of the pool of all possible votes that could be cast by the entire body of public safety practitioners in this country, hearing all possible messages. If the samples (votes collected in this study) are representative (with respect to parameters that effect the votes) of the larger pool, then we can use the collected votes to find an estimate of the votes in that larger pool. Specifically, we can find estimates for the fraction of "yes" votes. The fraction of "yes" votes is the fraction of public safety practitioners that find a speech transmission to be "suitable for mission-critical communications."

The combinations of packet loss ratio, packet loss correlation, and packet size detailed in Appendix B, define a total of 79, G.711 based speech transmission systems. In each table, systems with the attributes marked by "*" are expected to meet the requirement stated in the paragraph preceding each table. Table 35, Table 36, and Table 37 address attaining estimated "yes" votes from 70, 80, and 90 percent of the users respectively. As expected, as requirements become more and more stringent (moving from 70, to 80, to 90 percent estimated "yes" votes) only lower and lower levels of packet loss and packet loss correlation will support the requirement. As described in Section 2.1, a single mouth-to-ear delay requirement of 150 ms appears in all three tables.

2.2.1   Requirements for Mission-Critical Speech: 70 Percent Suitability

It is expected that at least 70 percent of public safety practitioners will judge a speech transmission system "suitable for mission-critical communications" when the system:

- Has acceptable packet loss ratio and packet loss correlation combinations as identified by cells marked with "*" in Table 35.

■ Conforms to the constraints and assumptions detailed in Section 1 and Appendix B of this document.

Table 35:   Speech Requirement Set 1

Packet Loss Correlation	Packet Size = 10 ms				Packet Size = 40 ms			
	Packet Loss Ratio				Packet Loss Ratio			
	0%	2%	5%	10%	0%	2%	5%	10%
0.0	*	*	*	*	*	*	*	
0.1	*	*	*	*	*	*	*	
0.2	*	*	*	*	*	*	*	
0.3	*	*	*		*	*	*	
0.4	*	*	*		*	*		
0.5	*	*	*		*	*		
0.6	*	*	*		*	*		
0.7	*	*			*	*		
0.8	*	*			*	*		
0.9	*	*			*	*		

Mouth-to-Ear Delay Requirement: No greater than 150 ms.

2.2.2   Requirements for Mission-Critical Speech: 80 Percent Suitability

It is expected that at least 80 percent of public safety practitioners will judge a speech transmission system "suitable for mission-critical communications" when the system:

■ Has acceptable packet loss ratio and packet loss correlation combinations as identified by cells marked with "*" in Table 36.

■ Conforms to the constraints and assumptions detailed in Section 1 and Appendix B of this document.

Table 36:   Speech Requirement Set 2

Packet Loss Correlation	Packet Size = 10 ms				Packet Size = 40 ms			
	Packet Loss Ratio				Packet Loss Ratio			
	0%	2%	5%	10%	0%	2%	5%	10%
0.0	*	*	*		*	*		
0.1	*	*	*		*	*		
0.2	*	*	*		*	*		
0.3	*	*	*		*	*		
0.4	*	*	*		*	*		
0.5	*	*			*	*		
0.6	*	*			*	*		
0.7	*	*			*	*		
0.8	*	*			*	*		
0.9	*	*			*	*		

Mouth-to-Ear Delay Requirement: No greater than 150 ms.

2.2.3   Requirements for Mission-Critical Speech: 90 Percent Suitability

It is expected that at least 90 percent of public safety practitioners will judge a speech transmission system "suitable for mission-critical communications" when the system:

■   Has acceptable packet loss ratio and packet loss correlation combinations as identified by cells marked with "*" in Table 37.

■   Conforms to the constraints and assumptions detailed in Section 1 and Appendix B of this document.

Table 37:   Speech Requirement Set 3

Packet Loss Correlation	Packet Size = 10 ms				Packet Size = 40 ms			
	Packet Loss Ratio				Packet Loss Ratio			
	0%	2%	5%	10%	0%	2%	5%	10%
0.0	*	*			*	*		
0.1	*	*			*	*		
0.2	*	*			*	*		
0.3	*	*			*	*		
0.4	*	*			*			
0.5	*				*			
0.6	*				*			
0.7	*				*			
0.8	*				*			
0.9	*				*			

Mouth-to-Ear Delay Requirement: No greater than 150 ms.

Note that jitter buffer design can be used to trade loss and delay. That is, when the attained delay is smaller than the delay requirement but the packet loss requirements are not met, increasing the size of the jitter buffer will increase the delay and may (depending on network loss and delay variation statistics) reduce packet loss. On the other hand, delay can be reduced by making the jitter buffer smaller, but this will typically (depending on network loss and delay variation statistics) increase the packet loss. In potential future work, an additional goal might be to separate out jitter buffer operations to arrive at a set of pure network requirements.

Note also that these results address only a single communications link between two points. In potential future work, one might develop models of practitioner-generated traffic and apply them, building on the present results to develop more general network requirements for larger numbers of practitioners communicating between larger numbers of network nodes.

3   Measuring Video Performance

3.1   Mission-Critical Video Services

Video applications are important in the mission of public safety. They will only increase in importance over time. The initial video performance effort in this document focuses on mission-critical video services. Mission-critical video services include applications in tactical public safety situations where there is a potential risk to human life (i.e., either to the lives of the first responders or to the individuals the first responders are aiding).

Qualitative descriptions of example mission-critical video services are:

- Ground-based and aerial video taken at the scene of a fire or other emergency sites to provide immediate tactical firefighting response, to coordinate rescue efforts, and to help distant EMS staff estimate required medical support, treatment, etc.

- Specialized non-visual video (such as infrared [IR]) to warn of spreading fire, heat sources, etc.

- Robotics video at an emergency site to control robotics devices and to assist with tactical decision making by the incident commander.

- Video in support of telemedicine taken by EMS staff at the scene of a fire or other emergency sites to help distant medical personnel evaluate patient condition and treatment, etc. Telemedicine techniques may require high-resolution video or pictures to allow viewing a patient's burns, skin and bone details, etc.

- Video taken at the scene of a stakeout, a traffic stop, or an arrest to send assistance in the case of trouble. This video may also be recorded for later use as evidence, for further investigation purposes, and to document officer conduct.

- Video used for a mutual aid operation; where there is a requirement to rapidly assess damage caused by the disaster, sending on-site views to recovery coordination officials at remote command posts. Real-time video may also be used by robotics operators and search and rescue teams, where the situation is too risky for first responders.

Mission-critical video services may include special situations and constraints that should be considered when developing a complete system specification. In these cases, the performance parameters in this document may not adequately characterize the required system performance and it is thus advisable to consider including additional performance parameters and specifications. Special situations and constraints may include, but are not limited to, the following:

- Where the environment lighting can range from very bright lighting to no lighting, or where there is an extremely wide range of lighting within the video scene.

- Where the video needs to be encrypted to preserve patient's HIPAA rights, to preserve individual's privacy rights, and to protect first responders' tactical operations from those without a need to know.

- Where the video may have to be stored, either at the scene or at a remote location, with very high resolution and clarity so it may be used as evidentiary material.

- A remote control (zoom, pan, focus, etc.) video collection system.

3.2   Reference Model for Video Performance Measurements

Figure 3 provides a reference model for specifying video performance measurements. To fully quantify the user perceived quality of service, the performance of three primary video subsystems must be specified.

1.   The Video Acquisition Subsystem—That normally consists of a camera system and may also include a built in video coder.

2.   The Video Transmission Subsystem—May include a network and its associated interfaces, encryption, etc., and may also include the video coder and decoder and a video storage medium.

3.   The Video Display Subsystem—Includes a monitor, playback computer, etc. May also include a build in video decoder.

The exact demarcation for each of the three subsystems can vary from application to application due to integration of various functions within the end-user's equipment. The approach adopted here is to specify performance parameters for the application as a whole (i.e., System Parameters—see Section 3.3), as well as to specify performance parameters that are unique to each video subsystem (i.e., Acquisition Parameters—see Section 3.4, Transmission Parameters—see Section 3.5, and Display Parameters—see Section 3.6). These generic sets of performance parameters are capable of characterizing the performance of different public safety applications. However, some public safety applications may only use a selected subset of these performance parameters. Figure 3 depicts a reference diagram for the performance measurements. The letters in the figure denote measurement access points that may or may not be available on all video systems.

Figure 3:   Video Performance Measurements Reference Diagram
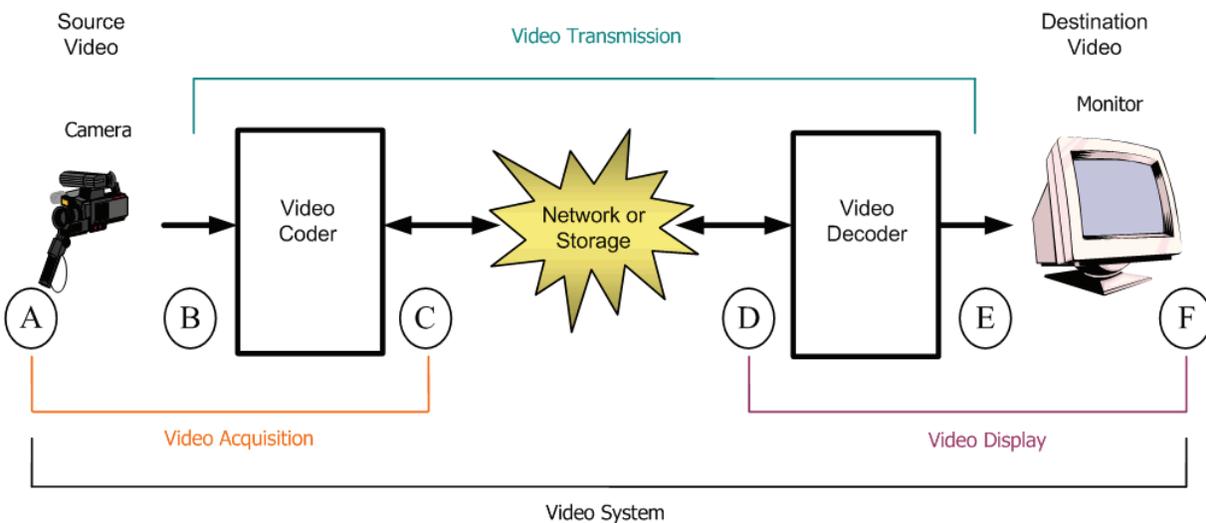
Figure 4 shows two example video systems with reference points identified. In the upper system, access point B is inside the camcorder and probably not available. In the lower system, access point E is inside the computer and may or may not be available. Some video may travel over multiple networks or storage medium.

Figure 4:   Two example video transmission systems, with reference points identified
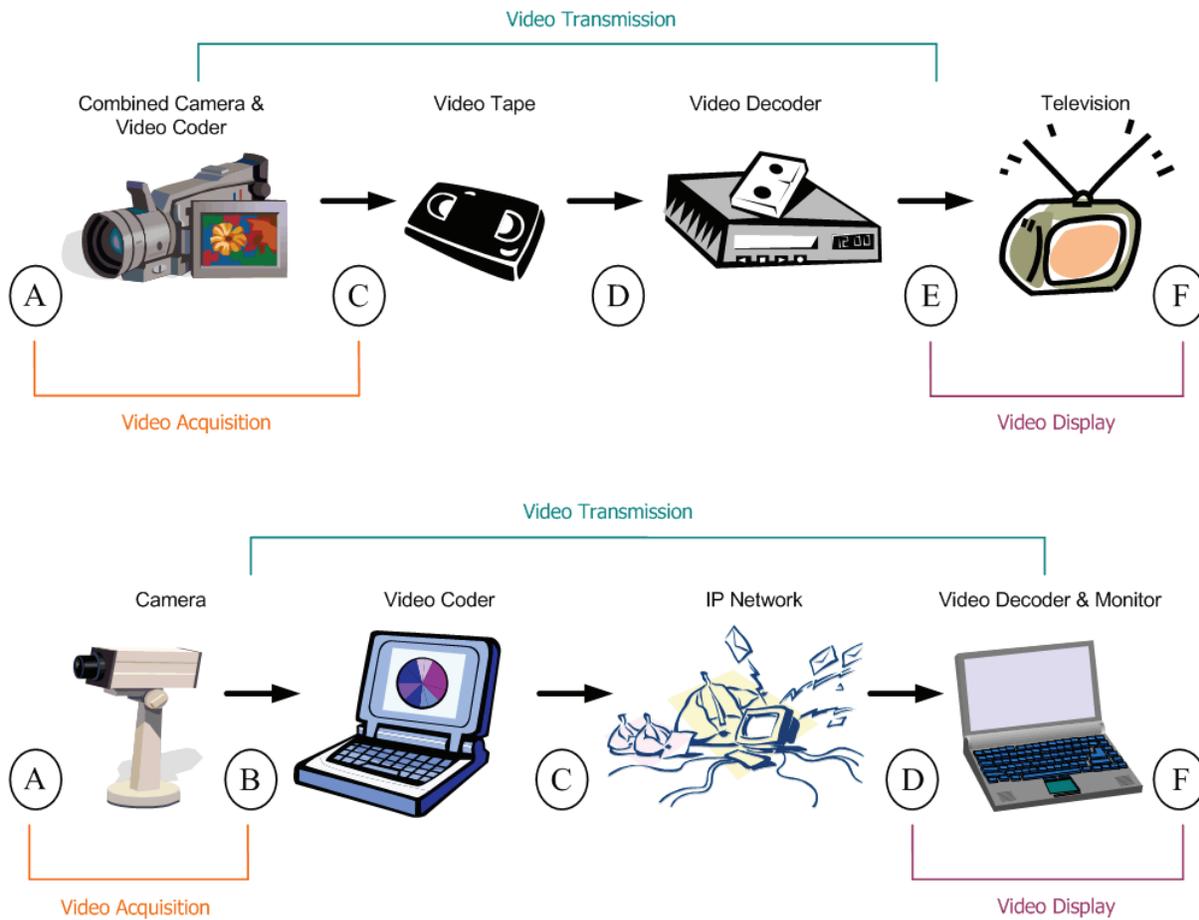
3.3   Video System Parameters

Parameters in this section characterize the performance of the entire video system, from point A to point F in Figure 3. When a specification for a video system parameter is included, all the video subsystems (video acquisition, video transmission, and video display) must support this specification.

3.3.1   One-Way Video Delay

For interactive video services, an important performance parameter is the length of the time delay that is required to send video through the video system. Since coding and decoding (see Figure 3) can add substantial delay, some compression algorithms will not be suitable for public safety applications that have a low video delay requirement. ITU-T Recommendation P.931 contains the recommended standard method of measurement for measuring the one-way video delay of the video transmission subsystem (i.e., from points B to E in Figure 3). To obtain the total one-way video delay (from points A to F in Figure 3), one must add to this the one-way delays of the video acquisition and video display subsystems. The video delays of the video acquisition and video display subsystems are specified by the manufacturers of the equipment.

For some public safety applications, the two-way delay (i.e., round-trip delay) may be the more important specification. To obtain two-way delay, the one-way delay in each direction should be measured separately and then combined. This approach is taken because the delay may not be the same in both directions.

3.3.2   Control Lag

For some interactive applications there may be a control lag. Control lag is defined as the lag time between a user's request and the time that the request is actually implemented. For example, there may be a control lag between a controller requesting a camera zoom, and the implementation of that request at a remote surveillance site. In this case, control lag would measure the time delay required to actually implement the request on the remote end, but the controller would not actually see the change in the video scene until the new video scene was transmitted back to the controller (i.e., after undergoing a one-way video delay). Thus, the two-way delay (i.e., round-trip delay) for this application would be the sum of the control lag and the one-way video delay. Control lag is highly dependent on the type and nature of the control being employed. When required, the manufacturer or service provider should be able to provide a specification for control lag.

3.3.3   Luma Image Size and Interlaced Versus Progressive Scan Type

The luma video signal is the black-and-white portion of the video picture, denoted as Y. The image size used to represent the luma video signal is an important video system parameter since this will limit the perception of resolution. Luma image sizes will be specified as <pixels horizontally> by <frame lines vertically>. Luma image size should be considered as an upper limit of usefulness for what *could* be achieved by an optimal video system. For example, just because a video system *could* produce, say, 352 by 288 useful pixels does not mean that it actually *does*. After coding and decoding, the user might only see the equivalent of 176 by 144 pixels! As another example, High-Definition TV (HDTV) monitors commonly display an image size that is less than 1920 by 1080 (e.g., video might be displayed at only 1366 by 768). In light of the above discussion, additional methods for quantifying resolution and video fidelity need to be specified. (See Section 3.4, Section 3.5, and Section 3.6.)

Scan type specifies whether the image is interlaced (i) scan or progressive (p) scan. With interlaced scan, the video frame consists of two interlaced fields; one field contains the even numbered lines and the other field contains the odd numbered lines. Fields are updated sequentially one field at a time. With progressive scan, the entire video frame is updated at the same time. Interlaced scan can produce smoother looking motion for a given frame rate (see Section 3.3.6), since the individual fields are updated at twice the frame rate. Progressive scan has advantages for reading text and displaying images on computer monitors.

Common luma image sizes and scan types are listed below:

- ■ HDTV, 1080i—1920 by 1080, HDTV, interlaced video.

- ■ HDTV, 720p—1280 by 720, HDTV, progressive video.

- ■ NTSC, 525i—525-line interlaced video (horizontal image size depends upon the sampling rate). The National Television Systems Committee (NTSC), U.S. standard definition television (of the 525 video lines, only 486 contain picture information). Note that there are various flavors of analog NTSC video, including composite (lowest quality), s-video (higher quality), and component (highest quality).

- ■ ITU-R Recommendation BT.601 (i.e., Rec. 601)—720 by 486 (for 525i), interlaced video. The studio quality sampling format for standard definition television signals. The electrical signal used

to transport the digitized video is commonly known as SDI (Serial Digital Interface), and is defined by SMPTE 259M.

- VGA—640 by 480, Video Graphics Array (VGA), progressive video. Used by computer monitors and as a square pixel sampling format for displaying 525i on computer monitors. (Rec. 601 pixels are non-square, so they will not display properly on computer monitors unless they are scaled.)

- CIF—352 by 288, Common Intermediate Format (CIF), progressive video. Used by video conferencing equipment.

- SIF—360 by 240, Source Input Format (SIF), progressive video. Used to encode Rec. 601 video at 1/4 resolution.

- QVGA—320 by 240, Quarter VGA (QVGA), progressive video. Used by personal digital assistants (PDAs).

- QCIF—176 by 144, Quarter CIF (QCIF), progressive video. Used by low-resolution video conferencing equipment.

- QSIF—180 by 120, Quarter SIF (QSIF), progressive video.

The luma image size and scan type of a video system are defined by the manufacturer and service provider specifications.

3.3.4   Chroma Sub-Sampling Factors

For video signals that contain color information, it is common to sample the chroma signals (e.g., the blue and red chroma signals, denoted as C_B and C_R) at a lower rate than the luma signal (i.e., the luma signal, denoted as Y). The reason is the human visual system is not as sensitive to this color information. For instance, the chroma signals are normally sub-sampled by a factor of two in the horizontal direction or sub-sampled by a factor of two in both the horizontal and vertical directions—with only minimal impact on perceived quality. The chroma sub-sampling factors of a video system will be specified as <horizontal factor> by <vertical factor>. Hence, the chroma images (C_B and C_R) will be smaller than the luma image (Y) by the chroma sub-sampling factors. For example, Rec. 601 video has chroma sub-sampling factors of 2 by 1, since the C_B and C_R signals are sub-sampled by a factor of 2 horizontally and 1 vertically (i.e., no vertical sub-sampling) with respect to the Y signal.

The chroma sub-sampling factors of a video system are defined by the manufacturer and service provider specifications.

3.3.5   Aspect Ratio

Correct aspect ratio should be maintained when the video is passed through multiple pieces of equipment (i.e., acquisition, transmission, and display). Aspect ratio for a video picture is defined as the ratio of the displayed image width divided by the displayed image height, expressed as <horizontal width> : <vertical height>. Common aspect ratios are 4:3 (NTSC) and 16:9 (HDTV). Use of a simple video scene that contains a square can verify that aspect ratio is maintained throughout the video system (i.e., by measuring and dividing the horizontal and vertical sides of the square).

3.3.6   Frame Rate

Frame Rate (FR) is the rate at which a video system can produce unique consecutive images called frames. FR is measured in frames per second (fps). For example, an NTSC video system can display 59.94

interlaced fields per second, so the FR of this system is 59.94/2 which is approximately 30 fps. (Since NTSC is an interlaced scan system, half of the picture is updated every 1/59.94 seconds and thus 2/59.94 seconds are required to update the entire frame, see Section 3.3.3).

Unlike NTSC, where FR is a fixed characteristic of the video standard, the FR of many new video systems can be independently specified to achieve the desired motion rendition. Some video codecs (i.e., coder-decoder pairs as shown in Figure 3) have even adopted an adaptive approach to FR, particularly when the transmission bandwidth is fixed or constrained. Thus, when the scene contains still or nearly still video, a high FR can be used (e.g., 30 fps). However, as scene complexity increases (e.g., lots of motion and detail), the video codec drops back to a slower FR (e.g., 10 fps). This results in an improvement to the user's overall perception of quality.

In this document FR will always refer to instantaneous FR. In other words, a specification for a minimum frame rate of 10 fps means that the video system must take no longer than 100 ms to present a new video frame. ITU-T Recommendation P.931 contains the recommended standard method of measurement for measuring the FR of the video transmission subsystem (i.e., from points B to E in Figure 3). To obtain the effective FR for the whole video system (from points A to F in Figure 3), one must compare this FR to the FRs of the video acquisition and video display subsystems, and use the minimum FR over all three subsystems. The FRs of the video acquisition and video display subsystems are normally constant (i.e., not time varying) and specified by the manufacturers of the equipment.

3.3.7   Acceptability Threshold

The acceptability threshold for a video system is defined as the lower bound on the probability, with 95 percent confidence, of obtaining acceptable video clips for a given public safety video application. The acceptability threshold is measured by conducting controlled, subjective evaluations of video clips using viewer panels of public safety practitioners (see Appendix C for an example procedure). These subjective assessments should be conducted in accordance with ITU-R Recommendation BT.500.

3.4   Video Acquisition Parameters

Video acquisition parameters measure the performance of the video acquisition subsystem (see Figure 3 and Figure 4), which reflects the creation of the video imagery itself. For some systems, the camera and coder are distinct and thus video acquisition spans from point A to point B. For other systems, the camera performs coding, and thus video acquisition spans from point A to point C—that is, the camera and coder cannot be separated.

Wherever possible, existing video acquisition performance metrics that are commonly used by industry have been specified. However, some public safety applications present unique video acquisition requirements that may require development of new performance metrics. Thus, this section is likely to evolve over time as additional public safety applications are examined.

The video acquisition parameters are identified as either *Primary* or *Secondary*. Primary parameters (listed first) are more important and should be specified for most public safety applications that require a video acquisition subsystem. Secondary parameters can be specified when they are applicable to meeting a specialized public safety requirement.

3.4.1   Resolution

(Primary Video Acquisition Parameter; method of measurement in Appendix C.4.1)
Resolution, as in sharpness, is the ability to resolve fine spatial detail in the video picture. Resolution will be quantified by MTF50P, the spatial frequency where the modulation transfer function, i.e., contrast, drops to 50 percent of its peak value. MTF50P is measured by analyzing near-vertical and near-horizontal edge responses of the video camera to the International Standards Organization (ISO) 12233 test chart (e.g., see Figure 20 on page 94) under standard lighting conditions. MTF50P is then converted into line widths per picture height (LW per PH) to produce a measure of total image resolution.

3.4.2   Noise

(Primary Video Acquisition Parameter; method of measurement in Appendix C.4.2)
Noise is the unwanted random spatial and temporal variations (e.g., "snow") in the video picture. One method of measuring noise is to capture and analyze images of the Kodak Q-14 test chart (e.g., top strip chart in Figure 21 on page 95). The Q-14 test chart consists of 20 patches with densities from 0.05 to 1.95 in steps of 0.1. Noise and signal-to-noise ratio (SNR) can be measured for each patch. SNR tends to be worst in the darkest patches. Several lighting conditions with various intensities (e.g., standard, reduced, dim) and color temperatures (e.g., tungsten, daylight) may be required to adequately characterize noise. Noise can be measured using a similar approach with the GretagMacbeth ColorChecker (the bottom checkerboard chart in Figure 21). It is a standard color chart consisting of 24 patches: 18 color and 6 grayscale. Noise and SNR are used in the calculation of dynamic range, as described in Section 3.4.3.

3.4.3   Dynamic Range

(Primary Video Acquisition Parameter; method of measurement in Appendix C.4.3)
Dynamic range (DR) is the range of luminance levels (from lowest to highest) that can be captured with reasonable quality, without clipping, by the video acquisition system. Two methods will be presented to measure dynamic range: an indirect method that infers or extrapolates dynamic range using a Kodak Q-14 reflection test chart (e.g., Figure 21, top strip), and a direct method that uses transmission test charts (e.g., Figure 35 on page 110).

3.4.4   Color Accuracy

(Primary Video Acquisition Parameter; method of measurement in Appendix C.4.4)
Color accuracy is the ability to reproduce colors with minimal chromatic distortion so that they are as close to real-life as possible, given the color-space limitations of the video standard being used (e.g., NTSC, Advanced Television Systems Committee (ATSC)). A GretagMacbeth ColorChecker test chart (the bottom checkerboard chart in Figure 21) is used to measure color accuracy. Several lighting conditions with various intensities (e.g., standard, reduced, dim) and color temperatures (e.g., tungsten, daylight) may be required to adequately characterize color accuracy.

3.4.5   Capture Gamma

(Secondary Video Acquisition Parameter; method of measurement in Appendix C.4.5)
Capture Gamma is a measure of camera contrast. It is the average slope of the equation that relates scene luminance to image pixel level, approximately, log(pixel level) = (Capture Gamma) × log(luminance). For image files intended for display on devices with display gamma = 2.2, Capture Gamma should be approximately 0.45.

3.4.6   Exposure Accuracy

(Secondary Video Acquisition Parameter; method of measurement in Appendix C.4.6)
Exposure accuracy is the ability of the video acquisition system to properly match the grayscale tonal levels of the scene being shot. All video cameras can be set for automatic exposure which controls the shutter speed, lens aperture, and gain combination that is used to achieve proper exposure. Some video cameras may have a manual override that lets the user select the aperture. It is anticipated that most public safety applications will use video acquisition systems with automatic exposure. Several lighting conditions with various intensities (e.g., standard, reduced, dim) and color temperatures (e.g., tungsten, daylight) may be required to adequately characterize exposure accuracy for these systems. Exposure accuracy is measured by photographing the GretagMacbeth ColorChecker or Q-14 test charts (Figure 21) against a gray background (which affects the automatic exposure setting), and comparing pixel levels for a range of gray patches from light to dark gray with standard values. Exposure accuracy may be affected by illumination history: it may change following exposure to bright light.

3.4.7   Vignetting

(Secondary Video Acquisition Parameter; method of measurement in Appendix C.4.7)
Vignetting, which identifies light falloff and uniformity, is the reduction in image brightness at the edges of the image versus the center of the image. Illumination in many inexpensive optical systems is not uniform: it decreases with distance from the image center. Some camera modules compensate for this digitally. Vignetting can interfere with other performance parameter measurements, particularly those that use charts consisting of patches of known density or color that span a good portion of the image. For instance, it may be necessary to compensate for vignetting to obtain a valid dynamic range measurement.

3.4.8   Lens Distortion

(Secondary Video Acquisition Parameter; method of measurement in Appendix C.4.8)
Barrel and *pincushion* are terms that describe two types of lens distortion. Barrel distortion is a lens distortion that produces greater magnification at the center of the image versus the edges of the image, resulting in rectilinear grid lines (see Figure 22 on page 95) that are bowed outward. Pincushion distortion is a lens distortion that produces less magnification at the center of the image versus the edges of the image, resulting in rectilinear grid lines (see Figure 22) that are bowed inward. Lens distortion can be measured from an image of a square or rectangular grid or from a simple rectangle near the image margins.

3.4.9   Reduced Light and Dim Light Measurements

(Secondary Video Acquisition Parameter; method of measurement in Appendix C.4.9)
UUnlike still cameras, long exposures are not an option with video cameras. Thus, whereas a still camera can create a high quality photograph (e.g., noise free) by increasing the exposure time, the maximum exposure time for a video camera is the reciprocal of the video frame rate. Hence video performance at low light levels (e.g., noise) cannot be inferred from measurements at high light levels.

3.4.10   Flare Light Distortion (Under Study)

(Secondary Video Acquisition Parameter; method of measurement under study in Appendix C.4.10)
Flare light distortion is caused by light that bounces between lens elements and off the interior barrel of the lens. It reduces the usable dynamic range of the camera system under adverse lighting conditions, such as might occur in a night-time police traffic stop where spotlights are used.

3.5   Video Transmission Parameters

The video transmission subsystem includes everything that occurs after the video has been rendered by the camera until just before the video is displayed on the monitor (from point B to point E in Figure 3). Video transmission can be significantly more complex than what is depicted in Figure 3. Instead of a simple coder/network/decoder, the video transmission subsystem could route the video through an H.264 video coder, over a congested IP network, through an H.264 decoder, through an MPEG-2 coder to a DVD recorder, to be decoded with a PC's DVD player and rendered by a computer that pauses the playback occasionally.

In most cases, the video transmission subsystem is a major contributor to video impairments.

3.5.1   Parameters for Measuring Calibration Errors

In addition to introducing video delay, a video transmission system may introduce other fixed distortions due to improper calibration. These calibration errors include spatial scaling of the picture (both horizontal and vertical), spatial shifts of the picture (both horizontal and vertical), a reduction of the picture area (valid region), and changes in gain (contrast) and level offset (brightness) of the video signal. For some applications, distortions to the video signal that result from calibration errors may not be important. For other applications (e.g., telemedicine), very accurate system calibration may be required.

This section describes a set of parameters that may be used to objectively measure the proper calibration of video transmission systems. One application of these measurements is to "tune" and correct potential calibration problems *before* the video transmission system is deployed. Ideally, network errors (see Section 3.5.3) should not be present when conducting the calibration measurements in this section as they may adversely impact the measurement accuracy.

3.5.1.1   Gain

Gain is a multiplicative scaling factor that has been applied to all pixels of an individual image plane (e.g., Y, C_B, and C_R) by the video transmission subsystem. Gain of the luma signal (Y) is commonly known as contrast. The ideal gain of the video transmission subsystem is 1.0 (i.e., no multiplicative scaling). The recommended method of measurement for gain is given in ANSI T1.801.03-2003.

3.5.1.2   Level Offset

Level offset is an additive factor that has been applied to all pixels of an individual image plane (e.g., Y, C_B, and C_R) by the video transmission subsystem. Level offset of the luma signal (Y) is commonly known as brightness. The ideal level offset of the video transmission subsystem is 0 (i.e., no additive shift). The recommended method of measurement for level offset is given in ANSI T1.801.03-2003.

3.5.1.3   Valid Region

Valid region is the rectangular portion of the image that is not blanked or corrupted by the video transmission subsystem. The valid region only includes those image pixels that contain usable picture information. Ideally, the valid region size is equal to the luma image size for the video standard being used (see Section 3.3.3) The recommended method of measurement for valid region is given in ANSI T1.801.03-2003.

3.5.1.4   Spatial Shift

Spatial shift is the shift of the image in the horizontal and/or vertical directions, measured in pixels. Spatial shift is considered positive when the video images exiting the video transmission subsystem are shifted to the right or down with respect to the video images entering the video transmission subsystem. The ideal spatial shift of the video transmission subsystem is zero (i.e., no spatial shift). The recommended method of measurement for spatial shift is given in ANSI T1.801.03-2003.

3.5.1.5   Spatial Scaling

Spatial scaling is an expansion or shrinkage of the image in the horizontal and/or vertical directions. The ideal spatial scaling of the video transmission subsystem is zero (i.e., no spatial scaling). If spatial scaling is present, the aspect ratio (Section 3.3.5) will be affected. The recommended method of measurement for spatial scaling is under study.

3.5.2   Parameters for Measuring Coding/Decoding Impairments

Coding and decoding are a reality of today's digital video systems. Coding entails compression, which enables a video service to be transmitted using a bandwidth that cannot accommodate the non-compressed video signal. The flip side of this coin is that some of the quality of the video signal may be lost. There is a tendency to demand or require "uncompressed" video or completely lossless coding for particularly important applications. However, in light of recent advances in video coding technology, the requirement for lossless coding should be carefully examined. Consider that most high-definition (HD) video cameras output a digital signal that has been compressed *with some loss*. Likewise, all HD recording medium in common use perform some degree of *lossy* compression. The resulting video is nearly perfect, but it cannot be called "uncompressed". Very high-quality codecs can reduce the transmission and storage bandwidths dramatically while causing only a minute drop in the perceived video quality. On the other end of the spectrum, some types of coding losses might be unacceptable, or some types of coding loss might be acceptable only because it enables a new video service to be available that otherwise would not be.

Motion and spatial detail jointly determine the compressibility of video scenes, so this is a key detail that must be considered in any method of measurement for quantifying coding/decoding impairments. Thus, the performance measurements in this section use actual video scenes content. This scene content should be selected to span the full range of motion and spatial detail that is required for the given public safety application.

3.5.2.1   Lossless Impairment

Lossless video transmission means that the video stream entering the video transmission subsystem (i.e., point B in Figure 3) is bit-identical to the video stream leaving the video transmission subsystem (i.e., point E in Figure 3). Peak-Signal-to-Noise-Ratio (PSNR), as defined in the Alliance for Telecommunications Industry Solutions (ATIS) Technical Report T1.TR.74-2001, is the recommended method of measurement for measuring lossless impairment. For the video transmission system to be truly lossless, the measured noise must be zero. This will produce an infinite PSNR, as you will be dividing by zero. Lossless impairment, as specified by PSNR, can also be effective for quantifying minor impairments to the uncompressed video stream (i.e., these impairments cannot include lossy video compression—see Section 3.5.2.2).

3.5.2.2 Lossy Impairment

Lossy video transmission means that the video stream leaving the video transmission subsystem (i.e., point E in Figure 3) has undergone lossy compression when compared to the video stream entering the video transmission subsystem (i.e., point B in Figure 3). The General Video Quality Model (henceforth abbreviated as VQM_G), as standardized by ANSI T1.801.03-2003,[8] will be used to specify the amount of perceptual lossy impairment. The recommended method of measurement is as follows:

1. Select at least eight video scenes with durations of 8 to 12 seconds each that span the range of scene content for the public safety application being deployed (e.g., tactical video—see Section 4). These scenes should be of the highest possible quality (uncompressed recording formats are recommended) and each scene should contain content that is substantially different from the other scenes being used. Scene characteristics to consider include a range of spatial detail, motion, color, and lighting levels. Development of a standard set of scenes to use for this purpose is under study.

2. Inject the scenes from step 1 into the video transmission subsystem (i.e., point B in Figure 3) and record the scenes from the output of the video transmission subsystem (i.e., point E in Figure 3). The recording of the output scenes should be of the highest possible quality—uncompressed recording formats are recommended.

3. Compute VQM_G for each pair of source and destination video streams from steps 1 and 2, respectively. The nominal range of VQM_G values is from 0 (i.e., no perceptual impairment) to 1 (i.e., maximum perceived impairment).[9]

4. The lossy impairment is computed as the average of the VQM_G values from all scene pairs.

3.5.3 Parameters for Measuring Impact of Network Impairments

Impairments present in the network from point C to point D in Figure 3) can significantly affect the perceived quality of the video service (point F in Figure 3). Network impairments can impact video quality in many ways including brief appearances of false image blocks and/or strips, a frozen image that resumes with a loss of content (i.e., skip), a frozen image followed by fast forwarding through the missing content, a sudden drop in image resolution, the image replaced with a blank screen, etc.

This section provides a set of network performance parameters that are known to have a potential impact on video quality. Specification of acceptable levels for these network parameters can only be made after the video coder and decoder are selected. This is because the efficiency of the video coding algorithm, the use of error concealment by the decoder, and the use of forward error correction (FEC) and/or retransmission methods by the coder/decoder pair, all influence the impact that network impairments have on the final perceived quality. Thus, specification of these network parameters *must always* be associated with a specific coder-decoder (codec) pair configuration.

3.5.3.1 Coder Bit Rate

Coder bit rate is the amount of information (in bits per second) output by the video coder to the network (point C in Figure 3), excluding all transport and protocol overhead and retransmissions. For the purposes

---

8. The General Video Quality Model has also been internationally standardized by ITU-T as Recommendation J.144 and by ITU-R as Recommendation BT.1683.
9. For extremely impaired video sequences, VQM_G can produce values greater than 1.0 but this is not common.

of this document, coder bit rate will always refer to the minimum instantaneous bit rate. Coder bit rate is specified by the manufacturers of the coder equipment.

3.5.3.2   Packet Loss Ratio

Packet loss ratio (PLR) is the fraction expressed in percentage (from 0 to 100 percent) of packets lost by the network (from point C to point D in Figure 3). The recommended method of measurement for PLR is given by the Internet Engineering Task Force (IETF) RFC2680.

3.5.3.3   Packet Size

Packet size (PS) is the size of an IP packet in octets, including all overhead as well as payload information. The influence of PS on video transmission quality is currently under study. One consideration is to force the network errors to occur at the same time slice for two identical video streams that have been encapsulated with different packet sizes. This would allow direct comparisons of the effects of PS on video quality. Another consideration is the decreasing goodput (e.g., useful application information that is transmitted by the network) that might be available to the video coder when packet sizes are reduced. A third consideration is how the video decoder will be affected by losing information in different locations of the coded video stream, and how this might impact error concealment algorithms, if they are present.

3.6   Video Display Parameters

Video display parameters measure the performance of the video display subsystem (see Figure 3 and Figure 4), which reflects the presentation of the video imagery to the user. For some systems, the display and video decoder are distinct and thus the video display spans from point E to point F. For other systems, these functions are inseparable and thus the video display spans from point D to point F. Wherever possible, existing video display performance metrics that are commonly used by industry should be specified. However, some public safety applications present unique video display requirements that may necessitate the development of new performance metrics. Thus, this section is likely to evolve over time as additional public safety applications are examined.

Video display parameters are under study.

4 Tactical Video Requirements

4.1 Description

Tactical video is used in real time during an incident by public safety personnel to make decisions on how to respond to that incident. This section will consider performance specifications for two categories of tactical video: narrow field of view and wide field of view.

In the narrow field of view, the features of interest occupy a relatively large percentage of the video frame. The camera is "zoomed in" relative to the objects pertinent to the application. Examples include:

- Video used to provide the incident commander with situation information, such as 1) a camera carried by a public safety practitioner, looking for victims, into a burning building looking for victims; 2) a body-worn camera during a SWAT raid; and 3) an aerial camera following a suspect on foot.

- Close-up videography from a camera on a robot being used to dismantle a bomb.

In the wide field of view, the features of interest occupy a relatively small percentage of the video frame. The camera is "zoomed out" relative to the objects pertinent to the application. Example scenarios include:

- Aerial videography used during wildfire suppression

- Aerial video to pursue an automobile

- A sweep of the whole incident scene to aid decision-making on how to deploy personnel

4.2 Feature Recommendations for Forensic Video Analysis

Forensic Video Analysis (FVA) is the scientific examination, comparison, or evaluation of video in legal matters. If there is a possibility that tactical video will be used for FVA, give consideration to including the features listed in Table 38.

Table 38:   Tactical Video Feature Recommendations for FVA

Feature	Description
FVA support	Tactical video systems should record video to support FVA when necessary.
Export of proprietary compression to standard format	Video systems with proprietary compression algorithms should export uncompressed video to a standard format, such as Audio Video Interleave (AVI).
Export of standard compression	Video systems with standard compression algorithms, such as MPEG-2 or H.264, should export either compressed video in that standard format or uncompressed video to a standard format, such as AVI.
Export of data to standard format	Video systems that associate data with the video, such as time, date, or Pin number used to open a secure door, should export that data to a standard format.

Table 38:   Tactical Video Feature Recommendations for FVA (Continued)

Feature	Description
Atomic clock synchronization	Video systems that record time information should be automatically synchronized to standard time taken from a U.S. atomic clock.
Automated authentication	Video recording systems should have an automated authentication mechanism. Video authentication should be attached to the video sequence when the video is first recorded. Preferably, video recording equipment should use a digital video signature that has been standardized and approved by the America Bar Association (ABA).
Swappable recording medium	The video recording medium (e.g., hard drive) in the video system should be easily swappable without disabling the system. Thus, in the event of removal, an alternative recording medium should be available.

4.3   Performance Recommendations

This section identifies a set of performance measurements for tactical video systems. The performance parameters and their recommended values are listed in Table 39 according to those that apply to the:

- Entire video system (Section 3.3)
- Video acquisition subsystem (Section 3.4)
- video transmission subsystem (Section 3.5)
- Video display subsystem (Section 3.6).

The recommended values for some of the "narrow field of view" tactical video performance parameters were obtained by conducting a survey and controlled subjective evaluations of video clips, using viewer panels of public safety practitioners (see Appendix C).

Table 39:   Tactical Video Performance Recommendations

Performance Recommendations	Narrow Field of View	Wide Field of View
Video System Parameters (Section 3.3)	Proposed Values	Under Study
One-Way Video Delay (Section 3.3.1)	Maximum of 1 Second	
Control Lag (Section 3.3.2)	Not Specified	
Luma Image Size and Scan Type (Section 3.3.3)	Minimum of 352 by 240, Progressive Scan[a]	
Chroma Sub-Sampling Factors (Section 3.3.4)	Maximum of 2 by 2	
Aspect Ratio (Section 3.3.5)	Not Specified	
Frame Rate (Section 3.3.6)	Minimum of 10 fps	
Acceptability Threshold (Section 3.3.7)	Minimum of 0.7	

Table 39:   Tactical Video Performance Recommendations (Continued)

Performance Recommendations	Narrow Field of View	Wide Field of View
Video Acquisition Parameters (Section 3.4)	**Proposed Values**	**Under Study**
Resolution (Section 3.4.1)	Under Study	
Noise (Section 3.4.2)	Under Study	
Dynamic Range (Section 3.4.3)	Under Study	
Color Accuracy (Section 3.4.4)	Under Study	
Capture Gamma (Section 3.4.5)	Not Specified	
Exposure Accuracy (Section 3.4.6)	Not Specified	
Vignetting (Section 3.4.7)	Not Specified	
Lens Distortion (Section 3.4.8)	Not Specified	
Reduced Light and Dim Light Measurements (Section 3.4.9)	Not Specified	
Flare Light Distortion (Section 3.4.10)	Not Specified	
Video Transmission Parameters (Section 3.5)	**Proposed Values**	**Under Study**
Gain (Section 3.5.1.1)	$0.95 \leq \text{Gain} \leq 1.05$	
Level Offset (Section 3.5.1.2)	$-10 \leq \text{Level Offset} \leq 10$ (for video systems with 255 quantization levels for each image plane)	
Valid Region (Section 3.5.1.3)	Minimum 95 Percent of the Luma Image Size (horizontally and vertically)	
Spatial Shift (Section 3.5.1.4)	Maximum of 2 Pixels (horizontally and vertically)	
Spatial Scaling (Section 3.5.1.5)	Under Study (No spatial scaling is preferred)	
Lossless Impairment (Section 3.5.2.1)	Not Required	
Lossy Impairment (Section 3.5.2.2)	Maximum of 0.41	

Table 39:   Tactical Video Performance Recommendations (Continued)

Performance Recommendations	Narrow Field of View	Wide Field of View
Coder Bit Rate (Section 3.5.3.1)	Minimum of 768 kbps for H.264 Video Codec[b] Minimum of 1.5 Mbps for MPEG-2 Video Codec	
Packet Loss Ratio (PLR) (Section 3.5.3.2)	Maximum of 0.1 Percent for H.264 Video Codec with No Error Concealment [c] Maximum of 0.5 Percent for MPEG-2 Codec with No Error Concealment[d]	
Packet Size (PS) (Section 3.5.3.3)	Under Study	
Video Display Parameters (Section 3.6)	**Under Study**	**Under Study**

a. Before being displayed, this image size must be up-sampled by a factor of 2 in both the horizontal and vertical directions using an up-sampling process that utilizes pixel interpolation.
b. A minimum H.264 coder bit rate of 384 kbps is recommended if the packet loss ratio can be held to 0.0%.
c. A PS of 600 was used for the H.264 codec with no error concealment. Acceptable PLRs for H.264 codecs with error concealment are under study. Some preliminary data is available—see Appendix D, "Video Quality Experiment PS1," on page 127.
d. A PS of 1358 was used for the MPEG-2 codec with no error concealment. Acceptable PLRs for MPEG-2 codecs with error concealment are under study.

5 Reference Model for Network Performance

5.1 Mission-Critical Network Services

A communications network permits the transmission of information from one location to another. The type of information transported depends on: the usage scenario such as, the number and type of devices and their location; and the application considered, such as tactical video or speech. These choices impose many constraints on the communications network. This document provides a common understanding of the constraints imposed by the applications and the usage scenarios to reveal insights into the resulting network performance trends.

You can take at least two approaches to choosing a communications network. The first approach is to design a network customized to the type or types of information to be distributed (i.e., based on application requirements). The second approach is to use an existing network design and examine how to accommodate the requirements of a specific application. However, regardless of the network design approach you choose, the central question remains: how well does a network distribute the information generated by a particular application, or in other words, how well does a network meet the application's quality of service requirements? The recommendations provided in this document offer useful planning information for network engineers building and maintaining communication networks for public safety.

This section describes a reference network model based upon a communication path within the System of Systems concept described in PS SoR Volume I. This model is consistent with public safety communications networks that will be partitioned based on functional and jurisdictional boundaries and constraints (i.e., a system of systems), and whose primary objective is to transport information between public safety communications devices (PSCDs).

A network, or system, consists of a group of nodes corresponding to individual communications devices, and links that connect the nodes to each other. The arrangement, configuration, or topology of nodes and connecting links is based on many variables. Using any specific physical or geographical representation of a network topology would provide only one example topology out of an almost endless set of possible network configurations. To keep our performance analysis tractable, therefore, we use a path-based reference model. This considers the path that information will traverse through the network between a pair of PSCDs.

This network performance reference model considers all possible paths, using the hierarchy network model presented across the various area networks. Every path will not be applicable in all situations. Some paths might seem very unlikely to occur. However, from an engineering perspective, it is important to consider the performance requirements of a range of possible scenarios to support all necessary cases, even if some are unlikely.

This section provides a definition of the path model first, followed by a discussion of the parameters considered in the model. These include the function, capacity, number, and performance characteristics of the nodes as well as the connecting links that directly influence the path characteristics.

5.2 Path Model Definition

A path is the set of links and nodes that the information traverses from the originating PSCD to the destination PSCD. Links in the path represent physical cables that connect pairs of nodes (e.g. fiber optic

lines linking high-speed routers), or they can correspond to point-to-point or one-to-many radio links in the case of wireless networks.

Generally, if we consider a path of interest, it has two types of nodes: end nodes and transit nodes. The end nodes are PSCDs and represent the origination or termination points for the connection. Transit nodes, also known as intermediate nodes, provide for the distribution, or routing, of the information stream carried by the connection.

In addition to the links and nodes along any given path, area network boundaries are defined by the administrative, jurisdictional, and coverage areas of various network segments that constitute the larger network. Figure 7 and Figure 8 illustrate this path-based model in the context of the PS SoR Volume I hierarchical reference network illustration (Figure 5) and link-based description (Figure 6).

Figure 5:   Natural Network Hierarchy
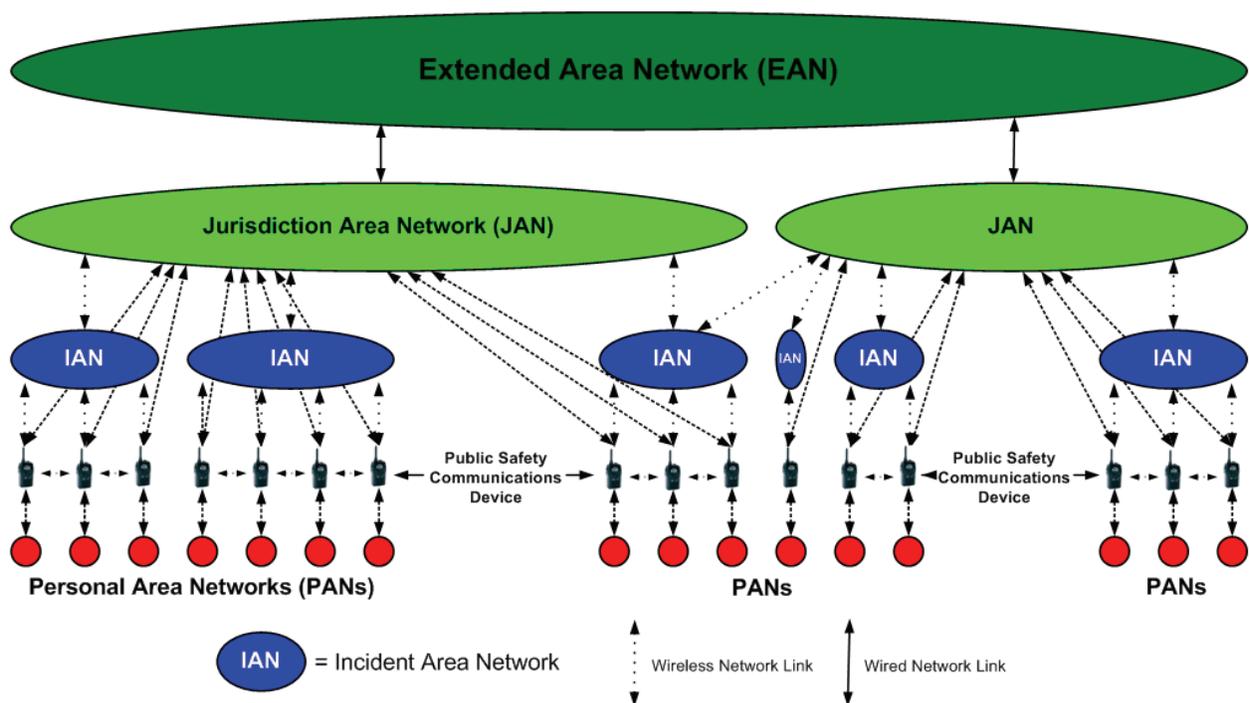
Figure 6:   Link Diagram

For this study, we develop two sets of path models we refer to as Model A and Model B, using the basic network architecture from PS SoR Volume I. Model A assumes the connection follows a strictly hierarchical path from the source PSCD through progressively larger area networks to an inflection point, after which the connection runs back down the hierarchy to the destination PSCD. The longest path allowed by Model A runs from the source PSCD on a personal area network (PAN), to the local Incident Area Network (IAN), then to the local Jurisdiction Area Network (JAN), next to the Extended Area Network (EAN), then to the destination JAN, destination IAN, and finally the destination PSCD and PAN.

Model B assumes that the hierarchical path includes peer-level communication links, i.e., from one IAN to another IAN.

In Model A, the communication path takes the least number of levels and the least number of links between the originating and terminating PSCDs. Two distinct sets of hierarchical paths exist. The first set contains all the symmetrical reference paths where the sequence of area network types traversed from the source PSCD to the inflection point is the reverse of the sequence of area network types traversed from the inflection point to the destination PSCD. An example is PAN-IAN-JAN-IAN-PAN. The second set contains all the asymmetrical reference paths, where the sequence of area network types traversed from the source PSCD to the inflection point is not the reverse of the sequence of area network types traversed from the inflection point to the destination PSCD. An example of this is the path, PAN-IAN-JAN-PAN. Figure 7 illustrates the following symmetrical and asymmetrical path types.

Symmetrical path types:

a.   PSCD via first responder's vehicle (FRV) to PSCD (PAN-IAN-PAN)

   (involves one IAN; two wireless links)

b.   PSCD via jurisdiction communication tower to PSCD (PAN-JAN-PAN)

   (involves one JAN; two wireless links)

c.   PSCD via FRV to jurisdiction communication tower to FRV to PSCD
   (PAN-IAN-JAN-IAN-PAN)

   (involves two IANs and one JAN; four wireless links)

d.   PSCD via FRV to jurisdiction communication tower to EAN to another jurisdiction
   communication tower to first responder's vehicle to PSCD
   (PAN-IAN-JAN-EAN-JAN-IAN-PAN)

   (involves two IANs, two JANs, and one EAN; six wireless links)

e.   PSCD to jurisdiction communication tower to EAN to another jurisdiction communication tower
   to PSCD (PAN-JAN-EAN-JAN-PAN)

   (involves two JANs and one EAN; four wireless links)

Asymmetrical path types:

f.   PSCD via FRV to jurisdiction communication tower to PSCD (PAN-IAN-JAN-PAN)

   (involves one IAN and one JAN; three wireless links)

g.   PSCD via FRV to jurisdiction communication tower to EAN to another jurisdiction
   communication tower to PSCD (PAN-IAN-JAN-EAN-JAN-PAN)

   (involves one IAN, two JANs and one EAN; five wireless links)

Figure 7:   Hierarchical Reference Paths Based on "Natural Network Hierarchy"
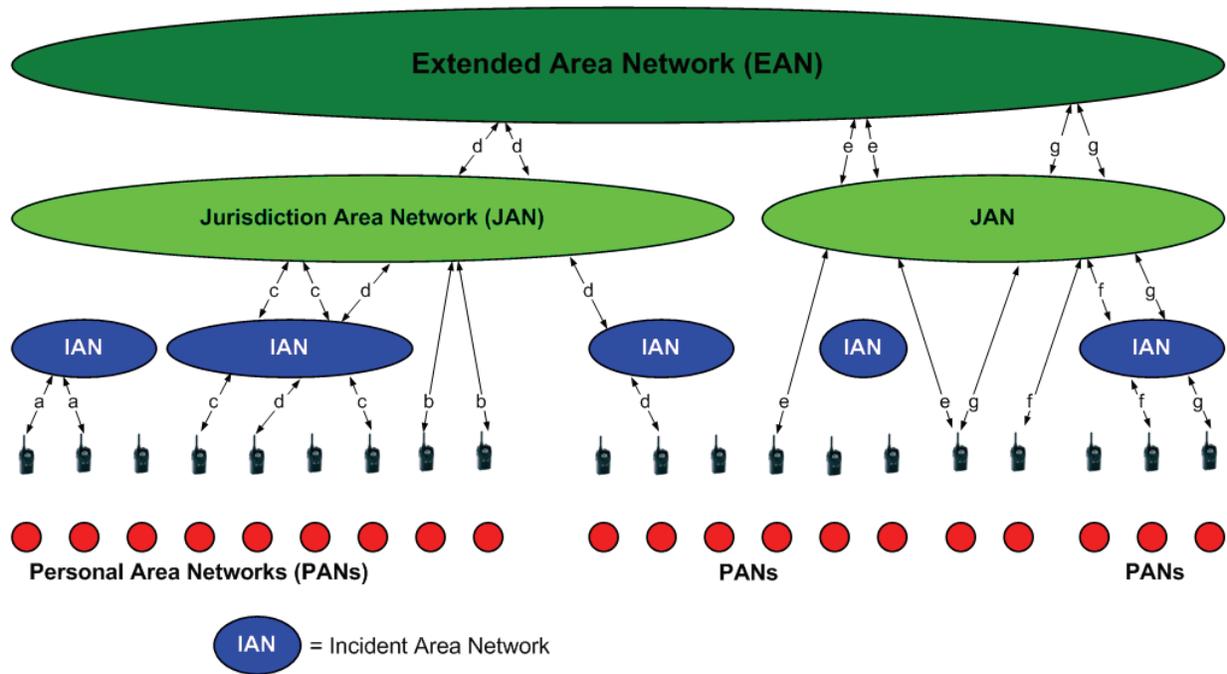
In Model B, the communication paths include peer-to-peer communication links between PSCDs and FRVs (IAN) communications equipment. Figure 8 shows the following peer-to-peer path types:

- PSCD to PSCD

  (where the PSCD to PSCD uses either the IAN or JAN interface with no infrastructure)

- PSCD to FRV to FRV to PSCD

  (where the PSCD to FRV uses the IAN and the FRV to FRV could use either the IAN or JAN interface)

Figure 8:   Peer Reference Paths (by links) Based on "Network Diagram Link Descriptions"
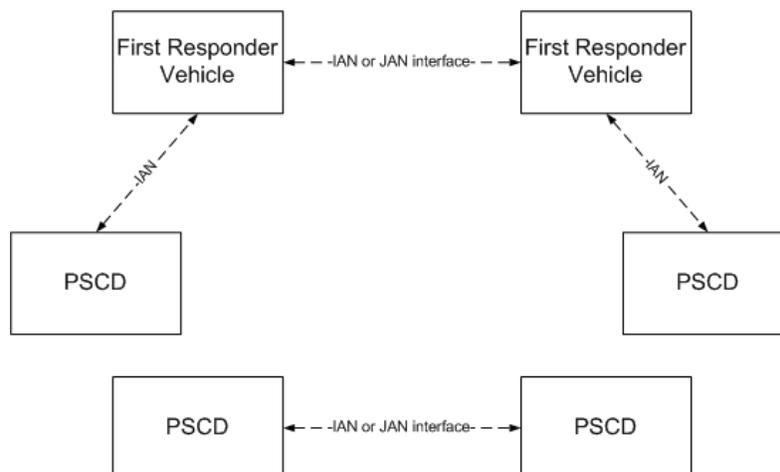
5.3   Path Model Parameters

The parameters characterizing the path are the links, nodes, and area networks that make up the path and directly influence the path characteristics. The links are characterized by the raw data rate available, and the signal propagation time through the different media types. Nodes are characterized by the average time to queue and process a packet and the protocols used from the ingress point to the egress point, including those used to access the transmission on the link. Area networks describe network segments composed of nodes that the path traverses.

This section identifies some of the most important parameters affecting path performance.

5.3.1   Medium Access Control

A Medium Access Control (MAC) protocol is designed to permit access to a shared medium on a fair and equitable basis. The design of a MAC will consider characteristics of the medium or link being shared, the number and arrangement of nodes accessing the medium, and application traffic characteristics. Many MAC protocols exist. For example, some well-known MACs are those defined in IEEE Std. 802.3[10] (commonly known as Ethernet) and IEEE Std. 802.11[11] (WiFi).

Our discussion of network performance is technology-neutral, so the PS SoR considers only the generic functions and features of a MAC. A MAC's two important functions are regulating the transmission of packets on the shared medium, and dealing with loss in case packets do not make it through.

The two general categories of MACs are time division multiple access (TDMA) and carrier sensed multiple access (CSMA). A TDMA MAC divides the shared medium so that once access is granted to a node; it is guaranteed access until the communication session is completed. Thus, the contending method is on a session-by-session basis to gain the initial granting of access for the communication. From the viewpoint of the node, the session is either a success (was granted access to the medium) or a failure (was denied access to the medium). This contending method happens in the CSMA MAC, but on a packet-by-packet basis.

Many flavors of CSMA allow multiple transmitting stations to share a communications medium in an uncoordinated fashion. Slotted Aloha is one the simplest examples. Unlike other more sophisticated versions of CSMA, Slotted Aloha does not require users to monitor the channel to see if it is busy before they begin transmitting. Users must schedule packet transmissions using a common clock that segments time into regular slots such that the length of a slot is equal to the amount of time required to transmit a packet. Any slot can potentially be used by any node. Thus it is possible for more than one node to transmit a packet in a given slot. If this happens, all packets transmitted in that slot collide and are assumed to be lost.

Another function performed by a MAC protocol is dealing with packet loss. Packets can be lost because of corruption by noise or interference, or because they have collided with packets that are being sent by other users on a shared channel. In some cases, a retransmission mechanism is implemented. This requires a procedure to transmit some kind of acknowledgment if the transmitter is incapable of sensing the channel or hidden terminals that are out of the transmitter's detection range. A collided packet is held by its transmitting station until it is successfully transmitted, or the amount of time allotted for it to be sent

---

10. Institute of Electrical and Electronics Engineers, IEEE Std. 802.3, 2005. Revision of IEEE Std. 802.3-2002 including all approved amendments.
11. IEEE Std. 802.11, 1999 (Reaffirmed June 12, 2003), "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications."

expires, or a set maximum limit on the number of attempts is reached. These retransmissions have the effect of increasing the traffic on the channel, thereby increasing the likelihood of collisions. While retransmissions are typically handled at the layer that is closest to the medium, in some cases this retransmission function may be performed in higher layers, such as the transport layer.

5.3.2   Propagation

The physical signal is transmitted over the link between two nodes as an electromagnetic wave. As such, there is a propagation delay that reflects the time the signal takes to travel through the link's medium, whether the medium is an optical fiber, or the air between two radio antennas. The propagation delay is the ratio of the distance between the link endpoints to the speed of light through the link medium. The speed of light in a transmission medium, such as air in the case of a wireless link, is $c_{medium} = c_{vacuum}/n_{medium}$, where $c_{vacuum} = 299,792,458$ m/s is the speed of light in a vacuum, and n_{medium} is the index of refraction of the medium. The index of refraction is a function of the frequency of the electromagnetic wave that is propagating through the medium. In the case of air, the index of refraction is approximately 1.0003 over a wide range of frequencies, while the index of refraction of the core of an optical fiber at the frequencies used for optical communications is often around 1.48. For example, a radio link between a FRV and a JAN tower that is 6 km (3.73 miles) away has a propagation delay of approximately 0.02 ms.

5.3.3   Channel Data Rate

The channel data rate represents the maximum rate at which data can be transmitted onto the channel or link. It is usually dependent on the rate the transmitter in the node can send data. It is expressed in the number of bits that can be sent in a 1-second interval, e.g., kbps.

5.3.4   Public Safety Communications Device

The origination and destination nodes in the reference path are PSCDs. A PSCD is assumed to have at least three separate physical wireless interfaces as shown in Figure 6. Link 1 and Interface 1 between the PSCD and its PAN constitute a local, very short distance wireless communications link, which must support from very low to very high data rates. Interface 2 uses Link 2 and Link 3 to respectively connect the PSCD to a FRV or to another PSCD. This is a pair of medium-range wireless links. When peer PSCDs are using Link 3 to communicate, they are doing so on a separate channel from that used by Link 2. Link 3 does not contend with or share the same channel as the link with the FRV or jurisdiction communication tower. Interface 3 and Link 5 between the PSCD and the JAN communication tower must allow the PSCD to transmit over long distances, since they are used when the PCSD is out of range of the IAN associated with the FRV.

The issue of RF resource limitations is another reason for assuming multiple physical wireless interfaces. It is not efficient to have a single device dividing its time while using the same resource. It will already need an extensive amount of overhead to permit the automatic discovery of devices and access points and handovers. This overhead can exceed resource limits and will affect the performance of the applications' data, especially service class 0.

5.3.5   First Responder's Vehicle

The the first resonder's vehicle (FRV) represents another node type in the reference path model. Examples of FRVs include a patrol car, a truck, an ambulance, a van, and a fire truck. This node's wireless access point is designed to link, or interface, with the PSCD, peer FRVs, and the jurisdiction communication

tower (i.e., using Links 2, 4, and 6 and Interfaces 2 and 3 in Figure 6). Since the purpose of the FRV is to provide a mobile access point for any PSCD associated with the FRV, and to provide a communication path back to the jurisdiction communication tower if the PSCD cannot reach the tower alone, the FRV must support at least two physical wireless technologies.

Even though Link 4 uses Interface 2 along with Link 2 and Link 3, when it is present it will be using a channel that is not the one generated by the FRV for the IAN. This is to reduce the probability of hidden nodes, to reduce contention, and to better coordinate resources. Link 4 can be considered an aggregate of the IAN for the FRV. If all communications in the IAN use Link 4 to reach their end destinations, then Link 4 is the aggregate of the entire IAN. If some of the communication sessions stay within the IAN, then Link 4 is only a partial aggregation.

5.3.6   Jurisdiction Communication Tower

The jurisdiction communication tower represents a node in the reference path and connects to two links using one interface, according to Figure 6. Thus, one wireless technology serves both the individual PSCDs and the FRVs. Link 7 joins jurisdiction communication towers and is assumed to be point-to-point. It is used for redundancy or when Link 8, which is wired, is not available.

5.3.7   Generic Nodes

Generic nodes include access gateway, interworking gateway, distribution, and core nodes. These nodes implement generic network functionality, such as routing packets and relaying signaling information, and do not have any additional functionality specific to public safety communications.

5.3.8   Node Delay

Node delay consists of several components, including processing, packetization, look-ahead, and transmission delays. In the reference model we use a constant delay for each node type that represents the sum of these node delays.

5.3.8.1   Processing Delay

Processing delay represents the time a device takes to do any of its tasks associated with forwarding a packet (e.g., read or write memory location, execute an instruction). Therefore, the faster the device can do its tasks or the smaller number of tasks it has to do, the less delay will be introduced. The source node processing delay consists of the coder delay and the algorithmic delay. The former is the delay associated with encoding raw sampled data, such as 64 kbps pulse code modulated (PCM), to produce a lower bit-rate data stream that retains most of the quality of the original signal. The amount of delay depends on the type of coder that is being used, but the worst-case coder delay is typically in the range of 10-20 milliseconds for speech applications. Video applications have bigger delay. Unencoded speech, described in ITU-T recommendation G.711,[12] has no coding delay.

5.3.8.2   Packetization Delay

The source PSCD, which is the first node on the path, imposes additional application-specific processing delays to convert application data into packets. This is the case with packetization delay, when sampled

---

12. Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.711, 1988, "Pulse Code Modulation (PCM) of Voice Frequencies."

speech signals or images are turned into packets of data before they are sent across the network to the destination PSCD.

5.3.8.3   Look-Ahead Delay

The source PSCD also introduces look-ahead delay, which results from compression algorithms' need to look ahead from the block of data they are processing to the next block of data as part of the algorithm. G.711's lack of encoding means this delay does not occur if raw 64 kbps speech is being transmitted. The amount of algorithmic delay depends on the encoder being used; for example, the G.723.1 coder introduces a look-ahead delay of 7.5 ms.

5.3.8.4   Transmission Delay

Each node also imparts a transmission delay, which is simply the amount of time required to send a packet on a link from one node to the next node. It is inversely proportional to the data rate of the transmitter on the link. For example, if we consider a 1500-byte frame transmitted on an 11 Mbps link, the transmission delay is 1500 x 8/11e6 = 1.091 ms.

5.3.9   Area Networks

This section describes various area network (i.e., PAN, IAN, JAN, and EAN) assumptions based on the information in PS SoR Volume I and other relevant documents from various standards bodies. We use these assumptions to generate default parameter values for sizing the various network segments that a path traverses.

5.3.9.1   PAN

The PAN, as currently defined in PS SoR Volume I, does not use wireless links to communicate between its sensors and PSCD. Instead the PAN is defined as a single link between an intermediate device that aggregates all of the data from the sensors and the PSCD by means of a wired link. A maximum distance of 2 meters (average height of a human) for the length of the wired communication link is assumed (with an average distance being 1 meter, where the aggregating device is located on the torso and the PSCD is handheld). This link represents the separation between the intermediate device and the PSCD.

5.3.9.2   IAN

We assume the coverage area for the IAN (one mobile FRV and one or more PSCDs) is dependent on the technology in use and environmental conditions. However, a minimum coverage radius must be at least 250 meters, which is the greater of the following two reference points:

- Minimum fire hose length is 244 meters (800 feet) in NFPA 1901[13]

- Minimum distance between incident at high-rise and base location is 60 meters (200 feet) in NFPA 1561[14]

---

13. National Fire Protection Association, NFPA 1901, 2003, "Standard for Automotive Fire Apparatus."
14. National Fire Protection Association, NFPA 1561, 2005, "Standard on Emergency Services Incident Management System."

5.3.9.3 JAN

We assume the coverage area for the JAN communications base station with a tower (PSCDs and mobile vehicles) is dependent on the technology in use and environmental conditions. Since this is most likely a fixed network infrastructure, coverage of the entire jurisdiction can be accomplished by systematic placement of base stations (access points), provided sufficient resources are available. Therefore, we assume that the range of the JAN is from 5 kilometers to 110 kilometers. Based on a sample of current paths of communication towers, we use an estimating factor of 2 to multiply the actual distance to represent the networking communications path's distance within the JAN.

5.3.9.4 EAN

Since the EAN represents the embedded network infrastructure, we use the ITU-T Y.1541[15] model for a network section (NS) for calculations with the user-to-network interfaces (UNIs) corresponding to the interface between a JAN and the EAN. We assume a worst-case distance of 12,000 kilometers (within the continental United States).

5.3.10 Number of Nodes

The number of nodes on a path consists of the number of PSCDs, FRVs, and jurisdiction communication towers on the path. We derive the following default values from the assumptions presented in PS SoR Volume I:

- The number of PSCDs per PAN is 1, by definition of a PSCD.

- The number of PSCDs per JAN is 30, based on statistical data of less than 100 sworn officers for 95 percent of the law enforcement agencies and an 8-hour work shift.

- The number of IANs is 15, 4, 10, or 3 per JAN, depending on the FRV's grouping.

  □ The number of IANs per JAN can be as high as one for every PAN PSCD, or one per vehicle. (We assume the following occupancy levels per vehicle: car: 2, van: 8, ambulance: 3, fire truck: 10.)

  □ Given that there are 30 PSCDs per JAN, the number of IANs within a JAN is based on the type of FRV and its assumed PSCD occupancy. For example, 30 PSCDs per JAN and 2 PSCDs per patrol car, yielding 15 IANs per JAN.

- The number of JANs is a maximum of 50,000.

  (There are 18,000 law enforcement agencies; 32,000 fire, EMS, local, state, tribal, and other organization agencies; and 100 Federal agencies.)

- The number of EANs is 1, as currently defined in PS SoR Volume I.

5.3.11 Protocols

To provide information concerning overall network packet and application overhead, we assume use of the Internet Protocol version 6 (IPv6), User Datagram Protocol (UDP), and Real-Time Transport Protocol

---

15. Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation Y.1541, 2002, "Network Performance Objectives for IP-based Services."

(RTP). To provide MAC information, we assume use of the TDMA and Slotted Aloha protocols. Figure 9 shows the protocol stack at the source and destination PSCDs.

Figure 9: Protocol Stack for End User's PSCDs

Application Layer (speech or video over RTP)
Transport Layer (UDP)
Network Layer (IPv6)
MAC Layer (TDMA or Slotted ALOHA)
Physical Layer (coax, fiber, wireless)

5.3.11.1   Internet Protocol Version 6

Internet Protocol Version 6 (IPv6), described in the Internet Engineering Task Force (IETF) Request For Comments 2460,[16] is currently being deployed as a major overhaul to the current Internet network layer technology (IPv4). With its new features, especially support for quality of service differentiation, it is expected to satisfy the requirements of any application for many years to come. An IPv6 packet header contains 20 bytes more than the IPv4 packet header does; most of this extra overhead is devoted to supporting the 128-bit source and destination addresses that were developed to preclude address space exhaustion. We assume the average node processing delays recorded for IPv4 apply to IPv6.

5.3.11.2   User Datagram Protocol

User Datagram Protocol (UDP), defined in RFC 768,[17] supplies a method for distinguishing among multiple applications by providing source and destination part identifiers in an 8-byte header. UDP also includes a 16-bit checksum to check for errors in the UDP header and payload. UDP does not provide guarantees of successful packet delivery. This is in contrast to the Transmission Control Protocol (TCP), which provides a reliable transport mechanism above the network layer. We use UDP in this analysis since the two applications (speech and video) favor timely delivery at a cost of potentially lost packets, instead of reliable delivery at a cost of potentially long delays.

5.3.11.3   Real-Time Transport Protocol

Real-Time Transport Protocol (RTP), defined in RFC 3550,[18] is a protocol framework designed for real-time applications, such as speech and video. RTP exists above the transport layer and applies a header to the application data that is 12 bytes long at minimum, although it can be up to 72 bytes long if the full set of up to 15 Contributing SouRCe (CSRC) identifiers is used. RTP allows additional header extensions, but their use is discouraged. The RTP header length can result in a significant amount of overhead, so the RTP framework allows for header compression on point-to-point links. The compression algorithm removes information that appears in the header of every packet data stream, and can reduce a 40-byte header to 2 to 4 bytes.

---

16. The Internet Engineering Task Force, IETF RFC 2460, 1998, "Internet Protocol, Version 6 (IPv6) Specification."
17. The Internet Engineering Task Force, IETF Standard 6/RFC 768, 1980, "User Datagram Protocol."
18. The Internet Engineering Task Force, IETF RFC 3550, 2003, "RTP: A Transport Protocol for Real-Time Applications."

RTP usually runs over UDP (although it runs over TCP if it carries RealAudio data), which has no acknowledgement mechanism like the one in TCP. For this reason, RFC 3550 also defines the Real-Time Control Protocol (RTCP), for creating and transmitting sender and receiver reports that contain descriptions of the data sent and received and that act as acknowledgements.

5.3.12   User Applications

From the view of the network, a user application is modeled as a traffic generator. In other words at regular time intervals the user application produces a certain-sized packet (in bytes). Using these packet size and generation interval parameters, we can calculate an average application data rate, which gives us the offered load. Packet size and generation interval parameters are sufficient if the user application is a constant bit rate service. Additional parameters may be necessary to model a user application other than a constant bit rate service.

The next two sections describe tactical speech and video as constant bit rate user applications.

5.3.12.1   Example Speech Application

ITU-T G.711[19] defines a speech encoding scheme known as Pulse Code Modulation (PCM), which produces an 8-bit sample every 125 microseconds. This results in an application data rate of 64 kbps. Since we assume the network is a packet network, and not a circuit-switched phone system, a number of samples must be grouped together to form an application packet. The packet has RTP, UDP, and IPV6 headers applied, and is given to the network to deliver. We assume that neither the size of the packet containing the original speech sample(s), nor the number of speech samples in a packet, changes over any link or section in the transmission path. That is, no fragmentation occurs along any link on the path. We use two speech sampling packet sizes: 80 samples per packet and 320 samples per packet.

5.3.12.2   Example Video Application

The H.264[20] and MPEG-2[21] standards define different video encoding and transmission schemes. The output application data rate is modified so that it produces an average constant bit rate. We assume a 600-byte packet for H.264 video, and a 1358-byte packet for MPEG-2. The video application packet is encapsulated using the protocol stack of RTP (IETF RFC 3550) over UDP over IPv6. We also assume that the size of the packet containing the original video application packet does not change over any link or section in the transmission path. That is, no fragmentation takes place along any link on the path.

---

19. Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.711, 1998, "Pulse Code Modulation (PCM) of Voice Frequencies."
20. Recommendations of the International Telecommunication Union, Radiocommunication Sector, ITU-T Recommendation H.264, 2005, "Advanced Video Coding for Generic Audiovisual Services."
21. International Organization for Standardization, ISO/IEC 13818 (commonly known as MPEG-2), 2000, "Information Technology—Generic Coding of Moving Pictures and Associated Audio Information."

6   Measuring Network Performance

PS SoR Volume I lists the following end-to-end PAN, IAN, JAN, and EAN upper-bound performance metrics:

- Packet transfer delay

- Packet delay variation

- Packet loss ratio

- Packet error ratio

This section provides an overview of several major factors affecting network performance, and describes the methodology used to create upper-bound network performance measures for end-to-end packet transfer delay and packet loss ratio. A future PS SoR version will provide measures for packet delay variation and packet error ratio.

6.1   Factors Affecting Network Performance

This section lists the most common factors that can impact the network performance. These include noise and interference as well as delays due to queuing and packet collisions.

6.1.1   Noise

Every packet is ultimately a long string of individual bits (zeros and ones) that are used to generate a modulated waveform that is transmitted through the air or over wires from a sending station, or source, to a receiving station, or destination. Individual bits in a received packet can be corrupted by noise on the communications channel. Two general types of noise cause errors: internal noise and external noise.

Appendix A of Ziemer and Tranter's *Principles of Communication: Systems, Modulation, and Noise*[22] provides a detail discussion of noise sources.

Internal noise originates within communications devices themselves and is caused by a variety of physical phenomena, which we briefly list here.

- *Thermal noise* (also known as Johnson noise or Nyquist noise) occurs because of random electron motion due to heat that occurs in any conducting or semiconducting material. Thermal noise power levels do not change with respect to frequency; the power density near baseband is the same as the power density at very high frequencies. For this reason, this type of noise is also known as white noise.

- *Shot noise* (also known as Schottky noise) occurs in electronic devices like diodes or transistors and arises because each electron that gets swept across the junction where two different types of semiconducting material meet carries a fixed amount of charge. The electrons cross the junction at random times and the result is a randomly fluctuating current exiting the device that can be modeled as a constant average current plus a random component.

---

22. R. E. Ziemer and W. H. Tranter, *Principles of Communications: Systems, Modulation, and Noise*, Houghton Mifflin Co., Boston, 1985.

- *Generation/recombination noise* occurs in semiconductors because of random generation and absorption of free negative charges (electrons) or free positive charges (holes) by the bonds between the silcon and dopant atoms in the material. These events happen randomly over time, so this kind of noise can be thought of as a kind of shot noise.

- *Temperature-fluctuation noise* occurs because elements of electronic systems such as resistors and transistors get hot, but their temperature is not constant. This is because the rate at which these devices give heat to the surrounding environment varies randomly over time. As the temperature of the device changes, so does the amount of thermal noise. Thus, temperature-fluctuation noise can be thought of as an additional noise component within the thermal white noise process itself.

- *Flicker noise* (also known as pink noise or 1/f noise) is noise whose power is greater at lower frequencies than at higher frequencies. This type of noise has been observed in many types of electronic devices, from vacuum tubes to field effect transistors. The exact cause of this type of noise is still a matter of debate in the research community.

External noise originates in the environment surrounding the communications devices. External noise can be generated by natural phenomena in the earth's atmosphere, mechanical devices, or energy sources in deep space. Examples of atmospheric effects include lightning strikes and auroras: the latter are generated by charged particles from the sun striking the earth's magnetic field and then traveling along field lines into the atmosphere. Human-made devices that generate noise include rotating machinery, automobile ignitions, and discharge from overhead power lines. Cosmic noise sources include the sun, other stars, pulsars, and quasars. While the sun is obviously the dominant noise source because of its close proximity to the earth, the more distant objects are collectively a significant source of noise due to their sheer numbers. Cosmic objects are broadband noise sources, and typically generate signals in the MHz to GHz portion of the spectrum.

A number of these external noise sources, particularly lightning and ignitions, generate noise that is highly impulsive, that is, a plot of the noise signal strength vs. time will show many large isolated spikes. You can observe this phenomena if you listen to a radio station transmitting in the AM band when a lightning storm is in the vicinity. The audible cracks and pops superimposed on the station's signal are the impulses generated by individual lightning events. As you might expect, these impulse events can introduce localized but severe errors in received packets.

6.1.2   Interference

In addition to channel noise, communications systems must contend with interference produced by other communications devices that operate in the same frequency band but are not necessarily attempting to communicate directly with either the transmitting or receiving device of interest. This type of interference is known as radio frequency interference (RFI). The following are several different sources of RFI:

- *Co-channel interference* (CCI) occurs when another carrier is present on the channel being used. This kind of interfering signal is not from another user attempting to communicate with a shared access point; it can be due to a user on another network who is trying to communicate with a nearby access point. This tends to occur more often in urban environments, where different providers' wireless systems are clustered close together.

- *Adjacent channel interference* (ACI) is caused by transmitters whose center frequency lies outside the passband of the receiver, but whose spectrum still overlaps the passband. If there is enough overlap, the additional energy entering the receiver can cause bit errors.

- *Inter-symbol interference* (ISI) is produced by the mechanism that turns the ones and zeros composing the bit stream into a modulated signal for transmission to the receiver. In an ideal

situation, the signal energy associated with a given bit would be confined to the time interval corresponding to that bit. However, because the shaping filters used to generate the transmitted waveform are not perfect, the transmitted energy associated with each bit bleeds into the bit intervals that occur after the bit in question. Communications systems designers use channel equalization to reduce the amount of ISI; this is preferable to more brute force methods such as simply increasing the transmitter's power level.

6.1.3   Packet Collisions

Packet collisions are another form of interference that occurs in a shared medium such as a radio channel in which the signals from two or more stations overlap while the stations are attempting to transmit data simultaneously. Because the signal strengths of the contending stations are generally of the same order of magnitude, the probability that the receiver will be able to determine correct bit values and associate them with the correct transmitter is very low. In this analysis, we assume that if any overlap occurs between two or more packets, all the colliding packets are lost. For this reason MAC layer contention resolution protocols, beginning with Aloha in the early 1970s and continuing with various types of CSMA protocols today, were designed to increase throughput and reduce the probability of collisions in situations where multiple users contend for access to a shared transmission medium.

6.1.4   Packetization

Packetization delay occurs because the source node needs time to collect several blocks of application data into a payload and form a packet with the necessary protocol headers. For simple coders like G.711 or G.726,[23] the packetization delay is computed as the ratio of the payload size (not including headers) in bits to the sampling rate in bits per second. For example, a G.711 coder that generates 8,000 8-bit samples per second and produces a payload of 640 bits (80 samples) incurs a packetization delay of 10 ms.

6.1.5   Queuing

Network equipment uses queues, also known as buffers, to hold packets for transmission at a later time. In particular, they are employed when a node receives more data than it can send out; in such a situation the node may chose to drop the data, causing packet loss, or place it in a queue until it can be sent on an outgoing link.

The time that a packet spends in a queue depends on many factors. The primary determiners of queuing delay are the queue utilization and the queue polling discipline. The former is related to the ratio of the arrival rate of packets into the queue to the rate at which they are transmitted on the outgoing link. If packets arrive at the queue according to a Poisson process, so that the amount of time between arrivals is exponentially distributed with mean $1/\lambda$ seconds, and if time to service a packet is distributed according to some arbitrary distribution with mean $1/\mu$ seconds and standard deviation σ_S seconds, the average queuing delay, W, measured in seconds is

$$W = \frac{1}{\mu} + \frac{1}{\lambda}\frac{\rho^2 + \lambda^2\sigma_s^2}{2(1-\rho)}$$

---

23. Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.726, 1990, "40, 32, 24, 16 kbps Adaptive Differential Pulse Code Modulation (ADPCM)."

where:

$\rho = \lambda/\mu$ is the queue utilization.

Note that as ρ gets close to 1, *W* goes to infinity, meaning that a queue will tend to back up when packets arrive nearly as fast as the queue can get rid of them.

Queuing delay can also be affected by queue polling to support differentiated quality of service for different applications. Weighted fair queuing (WFQ) is used to make sure that packets associated with delay-sensitive applications do not linger too long at a given node. WFQ assumes some mechanism exists to classify packets by application type (speech, data, etc.). The node uses this information to sort arriving packets into different queues; each queue has its own priority level. The node is then able to pull packets from high-priority queues more often on average than from low-priority queues, allowing delay-sensitive applications to enjoy better service than low-priority applications like routine file transfers.

6.1.6   Packet Loss and Retransmission

Packet loss and retransmissions have a significant impact on the overall network performance. While packets lost in the lower layers may not necessarily affect the packet loss ratio experienced by the application (due to retransmission mechanisms in lower layers), the time taken to retransmit lost packets is a direct contributor to the end-to-end delay. A packet can be lost due to link-related issues such as collisions with other packets at the MAC layer or data corruption caused by bit errors. Packets can also be lost because of node-related issues, especially buffer overflow. The penalty for retransmitting lost packets is increased delay.

Packet retransmissions can be managed by the MAC layer (layer 2) or by the transport layer (layer 4, specifically if some variety of TCP is used). Many MAC protocols, from Slotted Aloha through CSMA, are designed to allow retransmission after a random delay when a packet is lost due to collision. For example, the IEEE 802.11 layer 2 protocol supports packet retransmission with an exponential backoff algorithm; after each collision, the expected waiting time before the next transmission attempt increases by a factor of 2. High packet loss rates will thus produce a large average delay. Some versions of Slotted Aloha or CSMA allow a maximum number of retransmissions before the packet is declared lost; at this point, higher-layer protocols like TCP may attempt to retransmit the packet from the source node.

In contrast to the MAC, which is managed on a link-by-link basis, TCP uses acknowledgement messages from the receiver node to the transmitting node to allow the latter to identify TCP segments that have been lost in transit so they can be resent. TCP also uses a slow start mechanism that increases the number of packets that can be sent at one time every time the transmitter receives a packet acknowledgement. If there are many losses, the number of packets that TCP will allow to be sent per second will remain small, thus increasing the average delay over the link.

6.2   Packet Loss Ratio Computations

This section describes the computations used to obtain an upper bound for the packet loss ratio as experienced by an application through the reference network model described above. Packet loss occurs when a packet is lost in transit, or when it is so corrupted by bit errors that it arrives at its destination but in an unusable state. The packet loss ratio is defined as the probability that a transmitted packet never reaches its destination.

Figure 10: Nodes and Links Composing a Path, with Numerical Identifiers
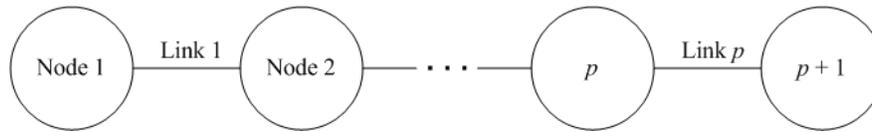
Given a path consisting of a set of p links $\{L_1, L_2, ..., L_p\}$ as shown in Figure 10, where the packet loss probability associated with link L_i is $PL_{link}(i)$, the success or failure of a transmission of a packet on a given link is independent of what happens on the other links on the path. The total packet loss probability for the path is 1 minus the probability that the packet is not lost on any of the p links. The probability that the packet is not lost on any of the links is the product of the complements of the loss probabilities for each of the links. Therefore, the path loss probability for a packet is:

$$PL_{path} = 1 - \prod_{i=1}^{p} (1 - PL_{link}(i))$$

where:

> PL_{path} is the probability that a given packet will be lost somewhere on the path between the source and destination PSCDs.

We compute upper bounds for the packet loss probability, $PL_{link}(i)$, for two types of link layers. In the first case we examine a dedicated channel, in which individual users are able to reserve bandwidth and do not have to contend for access to the receiver. This type of channel sharing is common among long-range access networks such as IEEE 802.16[24] networks. In the second case we consider Slotted Aloha, which is a CSMA MAC protocol and a simpler version of the protocol used by IEEE 802.11 networks.

6.2.1   Dedicated channel

If we have a dedicated channel with time division multiplexing (TDM), and the offered data rate is less than or equal to the channel rate, the loss rate for packets is zero since there is no contention. If the offered load is greater than the channel data rate, the packet loss rate for a new user is 1 because it is blocked from accessing the channel. We assume there are no retransmissions at the transport level, and we define G to be the offered load from the user population, normalized by the channel data rate. The following expression gives the worst case packet loss probability from the perspective of the application:

$$PL_{link}(i) = \begin{cases} 0, G \leq 1 \\ 1, G > 1 \end{cases}$$

This means that if any TDMA link on the path is oversubscribed, in the worst case, all packets will be lost and the attempted connection will be blocked.

---

24. Institute of Electrical and Electronics Engineers, Standard for Local and Metropolitan Area Networks, IEEE Std. 802.16, 2004 (Including IEEE Std. 802.16-2001, IEEE Std. 802.16c-2002, and IEEE Std. 802.16a-2003), "Part 16: Air Interface for Fixed Broadband Wireless Access Systems."

6.2.2   Slotted Aloha

Slotted Aloha has been analyzed extensively in the technical literature, although most analyses compute throughput as a performance metric, rather than the collision probability. Our analysis follows the development in Rom and Sidi, *Multiple Access Protocols: Performance and Analysis*,[25] as a starting point.

First, we describe the assumptions we used in our model. We assume there is a single receiver and there are a finite number of stations, given by the parameter *M,* that are all located at the same distance from the receiving station. The user population does not change and none of the users move. There are no hidden terminals, that is, every station is in within transmitting range of every other station. We assume that packets arrive at each station according to a Poisson arrival process with a mean arrival rate of *g* packets per second. All packets have the same length, and require *T* seconds to transmit, given the channel data rate. For the case where the MAC protocol is Slotted Aloha, described in Section 5.3.1, we assume that a station holding packets for retransmission will send its packet in a given packet transmission interval with probability σ. Because of these assumptions, it follows that σ, which is the probability that a single station transmits during a given interval of time equal to the packet transmission time *T*, is related to the other parameters by the expression $gT = M\sigma$.

The length of a slot in seconds, *T*, is computed from the network parameters as:

$T = (L / C) / 1000$

where:

       T is the length of the slot in seconds.

       C is the channel data rate in kilobits per second.

       L is the packet length in bits.

To obtain an upper bound for the packet loss probability $PL_{link}(i)$, we assume there are no retries for collided packets at layer 2, and that UDP is in use at the transport layer. Thus, PL_{link} on the link is equal to the packet collision probability, P_{coll}. To get P_{coll}, we first must compute σ, the probability that a station transmits in a slot of length *T*. σ is related to the offered load *G*.

G is the total offered bandwidth from the *M* stations using the channel normalized with respect to the channel data rate *C*, by:

$G = M\sigma$

where:

       G is the total offered bandwidth from the *M* stations normalized with respect to the channel data rate *C*.

       M is the number of stations.

       σ is the probability that a station transmits in a slot of length *T*.

We get the collision probability by determining the fraction of slots in which the channel is busy has multiple stations transmitting. To do this, we can look at a single slot and compute the probability that two

---

25. R. Rom and M. Sidi, *Multiple Access Protocols: Performance and Analysis*, Springer-Verlag, New York, 1990.

or more stations are transmitting in the slot given that the slot is not idle (i.e., at least one station is transmitting in the slot). Letting N_T be the number of transmitting stations in the slot, this gives us:

$$PL_{link} = P_{coll} = Pr\{N_T \geq 2 | N_T \geq 1\} = \frac{Pr\{N_T \geq 2\}}{Pr\{N_T \geq 1\}}$$

$$= \frac{1 - Pr\{N_T = 1\} - Pr\{N_T = 0\}}{1 - Pr\{N_T = 0\}}$$

$$= \frac{1 - M\sigma(1-\sigma)^{M-1} - (1-\sigma)^M}{1 - (1-\sigma)^M} = \frac{1 - G\left(1 - \frac{G}{M}\right)^{M-1} - \left(1 - \frac{G}{M}\right)^M}{1 - \left(1 - \frac{G}{M}\right)^M}$$

where:

M, σ, and G are as defined above and we use the fact that $G = M\sigma$.

This expression is the probability of at least two stations transmitting in a slot divided by the probability of at least one station transmitting in the slot. It can also be thought of as the ratio of the number of slots with collisions to the number of slots that are active in a long time interval. In other words, if we look at a long time interval and count the number of slots that were active (i.e., had at least one station transmitting), and see what fraction of them had collisions (i.e., 2 or more stations transmitting at the same time), we would get a good estimate for the collision probability, P_{coll}. Increasing the length of the time interval will improve the estimate.

6.3   End-to-End Packet Transfer Delay Computations

In this section we discuss computations to generate upper bounds on end-to-end delay for a given application, using the network model described in the preceding section. The end-to-end delay is measured from the originating PSCD to the terminating PSCD.

Many phenomena affect the end-to-end packet transfer delay; some of these effects, like propagation through free space, produce fixed delays, while others, such as access and queuing, produce random delays. The causes of path delay include the length and type of each link, the number of links and the number of nodes on the path, and the processing speed of each node. Other causes of delay are the time a packet must spend in various nodes' data buffers, and medium access delays resulting from multiple users having to contend for access to a shared channel.

We derive an expression for the path delay that is a function of the delays associated with the individual links and nodes that lie on the path between the originating and terminating PSCDs. We define the component delays, and develop expressions for average link delays, based on the type of MAC in use on the link. As was the case in the discussion of the packet loss metric, we restrict the MAC types to TDMA and Slotted Aloha.

We have a path consisting of p links $\{L_1, L_2, \ldots, L_p\}$ and n nodes $\{N_1, N_2, \ldots, N_{p+1}\}$, as shown in Figure 10, where Node 1 is the originating PSCD and Node $p + 1$ is the terminating PSCD. The total expected one-way delay for the path, D_{path}, is the sum of the expected delays associated with each of the nodes and links. It is given by:

$$D_{path} = \sum_{i=1}^{p} D_{link}(i) + \sum_{j=1}^{p+1} D_{node}(j) + \sum_{j=1}^{p+1} D_{access}(j)$$

where:

$D_{link}(i)$ is the average delay associated with the *ith* link in seconds.

$D_{node}(j)$ is the average delay associated with the *jth* node in seconds.

$D_{access}(j)$ is the average delays associated with accessing the *jth* link in seconds.

6.3.1   Link Delays

The delay incurred on the link depends on the link distances and the signal propagation through the media type. The signal propagation is the time to send a signal through a particular medium per distance covered. For each link in a path, signal propagation delay is calculated by dividing the link length in meters by the signal propagation speed through the medium. For a terrestrial coaxial cable, the signal propagation delay is calculated by taking the link multiplied by a unit delay of 4 ms per km. Signals in optical fiber cables travel a little slower (5 ms per km). (See reference ITU-T G.114[26]). For example, a 2000 km fiber link has a propagation delay of 10 ms.

An expression for $D_{link}(i)$ is as follows:

$D_{link}(i) = l(i) / s(i)$

where:

$D_{link}(i)$ is the average delay on the *ith* link in seconds.

$l(i)$ is the length of link i in meters.

$s(i)$ is the propagation speed on the *ith* link in meters per second.

6.3.2   Node Delays

The nodes on the path will also add processing delay to the total time required to send a packet. The originating PSCD and terminating PSCD are nodes, but so too are the FRV's, the jurisdiction tower, and any other device that lies on the path. The average total delays for four types of network nodes (access gateway, interworking gateway, distribution, and core) are given in ITU-T Y.1541. The average total delay includes queuing and processing delays. This gives us a figure for the node delay, $D_{node}(i)$. We assume that $D_{node}(i)$ is constant per node type.

Some of the processing delay is determined by the protocol stack. The model includes the amount of various protocol overhead in the delay computations. This overhead is added to the amount of data generated per packet by a given user. This is used to determine the total data rate of the traffic generated for the network links and nodes to service. From this we can compute delays that include the effect of overhead.

---

26. Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.114, 2003, "One-Way Transmission Time."

6.3.3   Medium Access Delays

The medium access delay is mainly affected by the MAC protocol, the channel data rate, and the load on a link. This section considers two generic MAC protocols: TDMA (dedicated) and Slotted Aloha.

6.3.3.1   Dedicated Channel (TDMA)

TDMA divides the given link capacity into channels by assigning users to different slots in a large frame that the MAC layer assembles and transmits. A given number of slots is available to accommodate a number of communication sessions at a given quality of service. In this situation, the medium access delay is the average time a station must wait for an assigned slot for a communication that is permitted access. However, once the capacity of the link is reached, no new communication sessions will be accepted. Thus, the delay for an unsuccessful communication attempt is infinite, or at least as long as necessary for established communication sessions to release enough resources so that the new session can get a channel. Because we are interested in the worst case, we model the delay as infinite if the number of stations in an area network exceeds what the network can support, given its channel capacity.

An expression for $D_{access}(i)$ is as follows:

$$D_{access}(i) = \begin{cases} \dfrac{T}{2}, G \le 1 \\ \infty, G > 1 \end{cases}$$

where:

$D_{access}(i)$ is the average delay to access the *ith* link in seconds.

T is the slot length in seconds.

G is the offered load from the user population normalized to the channel data rate.

6.3.3.2   Slotted Aloha

To get an upper bound for the average delay associated with using Slotted Aloha, we assume the worst-case scenario, in which there is no limit to the number of times that collided packets can attempt retransmission. A simple analysis that results in an expression for the average delay can be found in Rom 1990; we cover the main points here.

If a packet is successfully transmitted on the first attempt, its delay is a single slot interval. If the first attempt fails, the station that is attempting to transmit the packet goes into a state known as *backlog*, in which it holds the packet and attempts to retransmit it after a random delay. As long as the backlogged station's attempts to transmit the packet that it is holding fail, it continues to schedule the packet for another attempt after a random delay. Because there is no limit to the number of times a backlogged station can attempt to transmit a packet, eventually the packet will be transmitted successfully (as long as the probability of a collision is not 1), and the backlogged station will move on to the next packet. Because of this policy, the average access delay experienced by a packet whose first transmission attempt is unsuccessful is one slot plus the average number of slot intervals that the packet waits for while the packet is held by the station (i.e., the packet is backlogged within the station).

By using Little's Theorem, we find the average number of slot intervals that a packet spends in backlog is the ratio of the average number of stations in backlog, $\overline{N}$, to the rate that stations become backlogged, b. So the average delay, in slots, is:

$$E\{D_{access}\} = \Pr\{\text{success on 1st try}\} \cdot 1 + \Pr\{\text{failure on 1st try}\}\left(\frac{\overline{N}}{b} + 1\right)$$

The throughput of a Slotted Aloha system, *S*, is the rate at which packets leave the station and are received by the destination node. Some of this throughput is from backlogged stations and the rest is from stations that succeed on the first attempt. The probability that a packet goes into backlog is *b/S*, the ratio of the backlog rate to the throughput. Therefore, we get:

$$E\{D_{access}\} = \frac{S-b}{S} \cdot 1 + \frac{b}{S}\left(\frac{\overline{N}}{b} + 1\right) = \frac{S + \overline{N}}{S}$$

To get the expected delay, we need to get the average number of backlogged stations. We note that the throughput *S* is equal to the offered load $G = M\sigma$, minus the load corresponding to unsuccessful transmissions from the backlogged stations, which is $\overline{N}\sigma$. So:

$$E\{D_{access}\} = 1 + \frac{G-S}{S\sigma} = 1 + \frac{M}{S} - \frac{1}{\sigma}$$

The throughput *S* as a function of *G* and *M* is given by:

$$S = G\left(1 - \frac{G}{M}\right)^{M-1} = M\sigma(1-\sigma)^{M-1}$$

This gives us the average delay in slot intervals:

$$E\{D_{access}\} = 1 + \frac{G-S}{S\sigma} = 1 + \frac{1}{\sigma(1-\sigma)^{M-1}} - \frac{1}{\sigma}$$

Multiplying this quantity by the length of a slot gives us the average delay in seconds:

$$D_{access}(i) = T * E\{D_{access}\}$$

where:

> $D_{access}(i)$ is the average delay in seconds.
>
> *T* is the slot length in seconds.
>
> $E\{D_{access}\}$ is the average delay in slots.

Note that for the case where there is a single user (*M* = 1), the expected delay is one slot interval, which is what we would expect since no other stations are contending for access to the channel, and every packet transmission attempt is successful.

7 Network Requirements

This section provides network performance recommendations to support the tactical speech and video application requirements, also known as user-to-user-perceived quality of service requirements. In addition, this section provides a recommendations summary for all area network performance requirements.

7.1 User-Perceived Quality of Service

Figure 11 illustrates the general performance requirements for user-to-user-perceived quality of service. These requirements can be broken into two components: the application-specific processing component, and the network service component. While performance requirements imposed by the application-specific processing component are important to consider, they depend on the type of application and thus are generally outside the scope of this study. We are mainly concerned with the network service component, which can be further divided into two major segments, as follows:

- The public network (i.e. EAN)
- The various public safety area networks detailed in Section 5.3.9

Depending on the path, a subset of the available area network types will compose the public safety component of the network segment.

Figure 11:   General Performance Requirements for User-Perceived Quality of Service
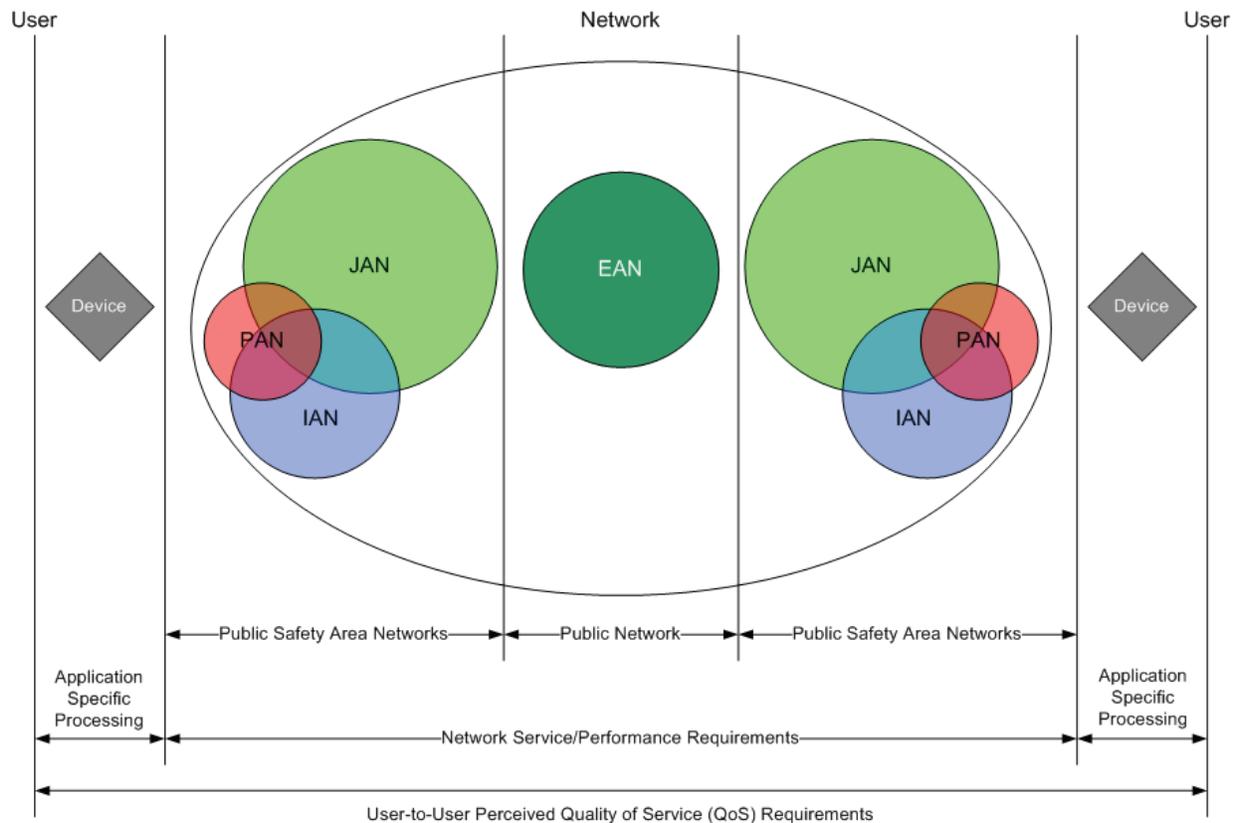
Following is the step-by-step approach we use to provide upper bounds for the network performance requirements:

1.   We use the path-based network reference model and the measurement methodology described in Section 5 and Section 6, respectively, to calculate upper bounds for the packet loss and the end-to-end delay for Paths A through G, inclusive.

2.   We compare these path bounds to the user-to-user-perceived performance requirements established in PS SoR Volume I. In this comparison, we account for the application-specific processing requirement for the given application type.

3.   In case the path upper bounds obtained in step 1 exceed the user-to-user-perceived performance requirements, the path bounds are set to the user-to-user-perceived performance requirements, minus the application specific processing requirements.

4.   We obtain the public safety area networks performance allocations from each path's upper bounds by grouping the links and nodes in the path into area networks.

5.   We obtain upper bounds for area network performance requirements by comparing the set of area network performance requirements for all paths, and selecting the maximum values from the set.

7.2   Speech Applications

The network performance recommendations necessary to support speech applications are based on the user-perceived quality of service study for speech applications described in this volume. In addition, the

maximum acceptable mouth-to-ear packet loss ratio and end-to-end transit delay were obtained from ITU-T G.711 "Pulse Code Modulation of Voice Frequencies." These metrics are summarized in Section 7.2.1 and Section 7.2.2 for packet loss and delay, respectively. Sections Section 7.2.3 through Section 7.2.9 provide the network performance budget for Paths A through G, respectively, in terms of the packet loss and end-to-end delay.

7.2.1   Packet Loss Requirements

The packet loss requirements described in this volume for speech depend upon several factors such as packet size, speech decoding, and post processing algorithms (e.g., packet loss concealment). This section describes packet loss requirements for packet sizes of 80 and 320 bytes corresponding to packet interarrival times of 10 ms and 40 ms, respectively.

Table 40 summarizes the packet loss requirements considered in this study for different percentages of satisfied public safety practitioners, assuming a packet loss correlation of 0.0.

Table 40:   Packet Loss Requirements for Percentages of Satisfied Practitioners

Percentage of Satisfied Practitioners	Packet Voice Sample Size	Packet Loss Percent Requirements
70 percent	80 bytes	10 percent
	320 bytes	5 percent
80 percent	80 bytes	5 percent
	320 bytes	2 percent
90 percent	80 bytes	2 percent
	320 bytes	2 percent

Figure 12 illustrates the budget allocation for speech packet loss ratios. We use a network packet loss budget requirement of 1×10^{-3} specified in ITU-T Y.1541 for the EAN.

Figure 12:   Speech Maximum Packet Loss Ratio Requirements
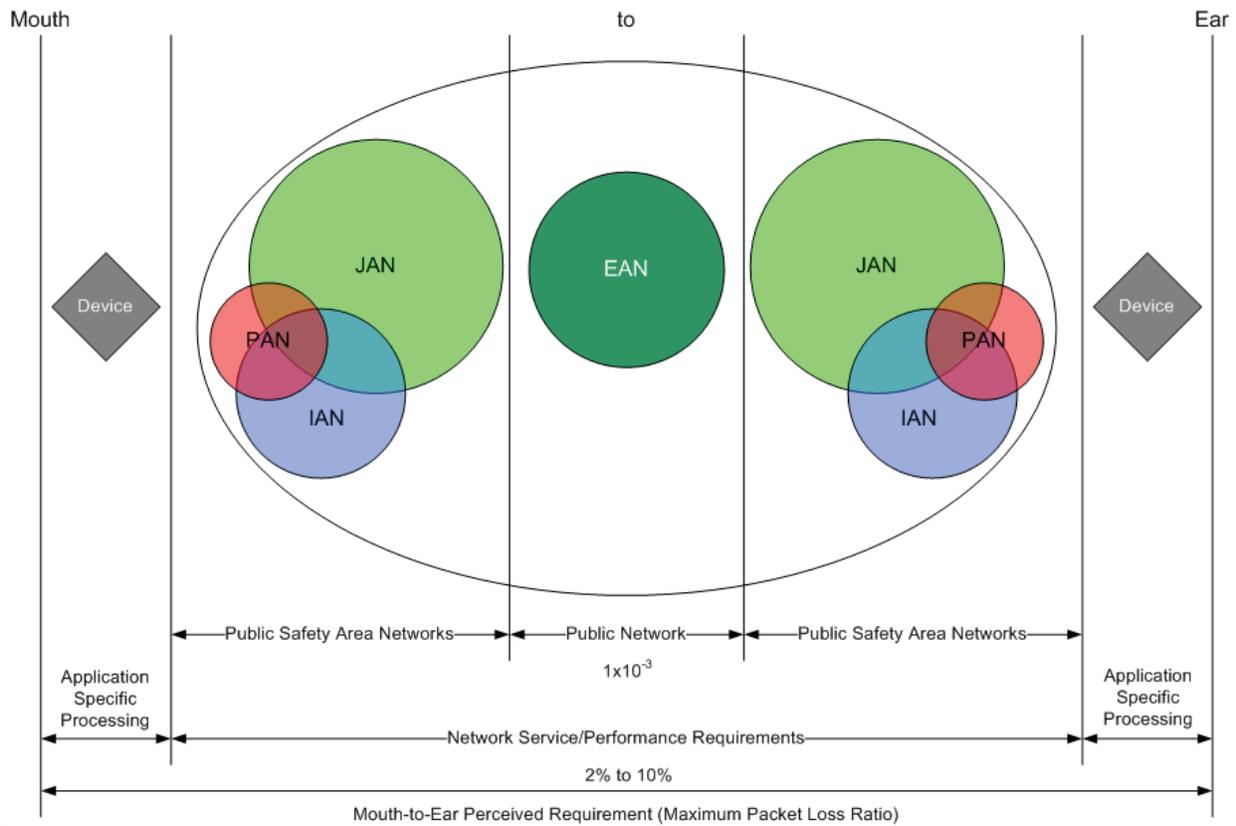
7.2.2   End-to-End Delay Requirements

Given a maximum mouth-to-ear, end-to-end delay of 150 ms and three segments of the mouth-to-ear path (PSCD; PAN, IAN, and JAN; EAN), Figure 13 illustrates the recommended maximum end-to-end delay allocations, presented in terms of path and other input parameters. The value of "x" that represents the delay contributed by the application-specific processing device includes the packetization delay described in Section 6.1.4, and is bounded by 0.375 ms according to ITU-T G.114 for PCM.

We use a network budget requirement for end-to-end transit delay of 100 ms given in ITU-T Y.1541 for the EAN.

Figure 13:   Speech Maximum End-to-End Delay Requirements
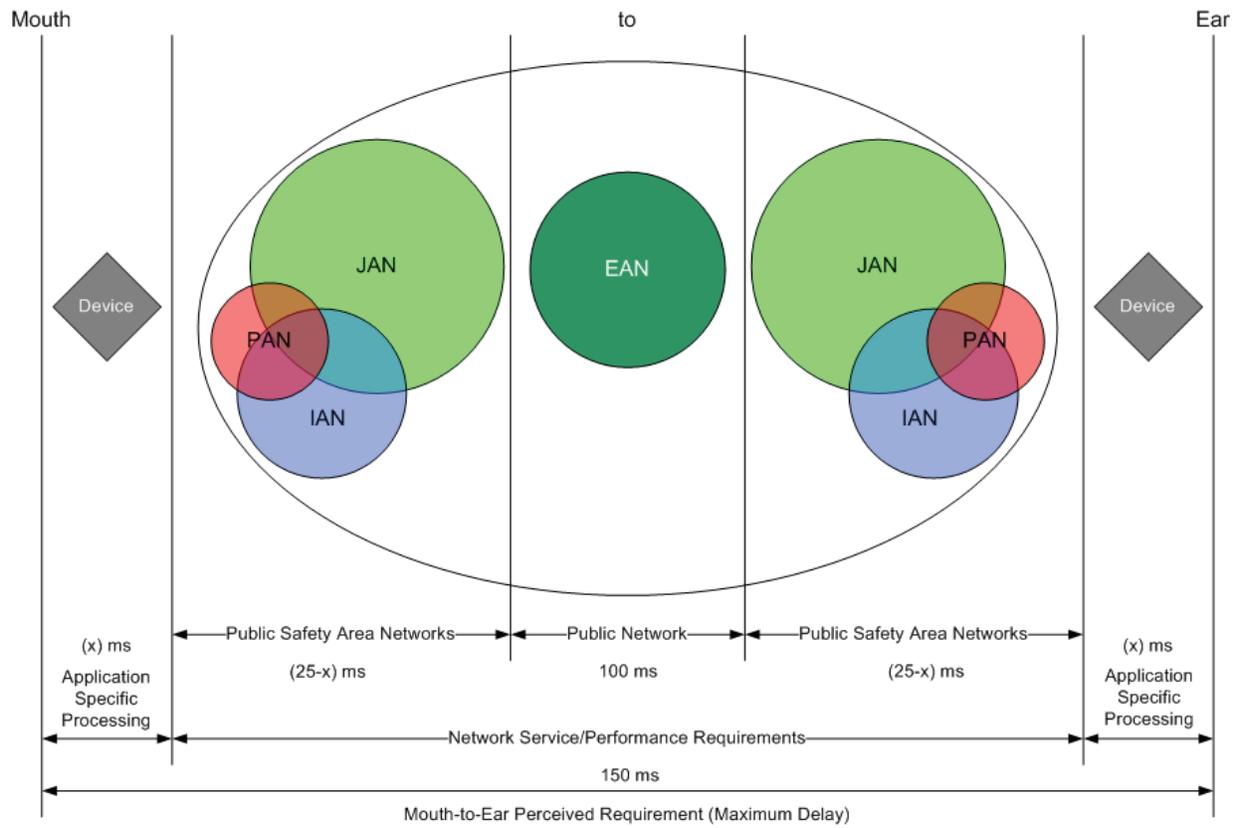
7.2.3   Path A

Path A consists of two PANs and one IAN. This path is one in which it is very easy to allocate the link budget for the various network performance parameters.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel rate of the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

Table 41:   Speech Path A Network Performance Parameter Requirements

| Speech Path A | Calculated | | | | Recommendation | |
| | 80-Byte Packets | | 320-Byte Packets | | 80-Byte Packets | 320-Byte Packets |
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	11 ms	49 ms	11 ms	115 ms	45 ms**	17 ms**
Packet Loss Probability	0.00001	0.3944	0.00001	0.3944	0.1*	0.05*
PAN						
End-to-End Delay	<< 1 ms	<< 1 ms	<< 1 ms	<< 1 ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	49 ms	11 ms	115 ms	45 ms**	17 ms**
Packet Loss Probability	0.00001	0.3944	0.00001	0.3944	0.1*	0.05*

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.2.4   Path B

Path B consists of two PANs and one JAN. This path, unlike path A, may possibly contain multiple links within the JAN (i.e., not simply a communication path from PSCD to a single jurisdictional communication tower to another PSCD).

The PAN component is stable with a very small contribution to network performance.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and the channel data rate on the link.

Table 42:   Speech Path B Network Performance Parameter Requirements

Speech Path B	Calculated				Recommendation	
	80-Byte Packets		320-Byte Packets		80-Byte Packets	320-Byte Packets
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	7 ms	122 ms	11 ms	320 ms	40 ms**	60 ms**
Packet Loss Probability	0.00001	0.3944	0.00001	0.3944	0.1*	0.05*
PAN						
End-to-End Delay	<< 1 ms	<< 1 ms	<< 1 ms	<< 1 ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
JAN						
End-to-End Delay	7 ms	122 ms	11 ms	320 ms	40 ms**	60 ms**
Packet Loss Probability	0.00001	0.3944	0.00001	0.3944	0.1*	0.05*

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.2.5   Path C

Path C consists of two PANs, two IANs and one JAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and the channel data rate on the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, like the IAN component, is significantly affected by the choice of MAC, number of PSCDs and FRVs, packet size, and link channel rate.

Table 43:   Speech Path C Network Performance Parameter Requirements

Speech Path C	Calculated				Recommendation	
	80-Byte Packets		320-Byte Packets		80-Byte Packets	320-Byte Packets
Details	**Minimum**	**Maximum**	**Minimum**	**Maximum**	**Upper Bound**	**Upper Bound**
End-to-End Delay	27 ms	∞	39 ms	∞	32 ms**	39 ms**
Packet Loss Probability	0.0014	1	0.0014	1	0.1*	0.05*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	∞	12 ms	∞	11 ms	12 ms
Packet Loss Probability	0.0001	1	0.0001	1	0.0007	0.007
JAN						
End-to-End Delay	7 ms	∞	16 ms	∞	11 ms	17 ms
Packet Loss Probability	0.0162	1	0.0162	1	0.0163	0.0163

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.2.6   Path D

Path D is similar to Path C, except it contains the EAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel data rate on the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, like the IAN component, is significantly affected by the choice of MAC, number of PSCDs and FRVs, packet size, and link channel rate.

The EAN is assumed to satisfy the network performance objectives given in ITU-T Y.1541. We use a network budget requirement for the packet loss ratio of 10^{-3}, and an end-to-end transit delay of 100 ms given in ITU-T Y.1541.

Table 44:   Speech Path D Network Performance Parameter Requirements

| Speech Path D | Calculated | | | | Recommendation | |
| | 80-Byte Packets | | 320-Byte Packets | | 80-Byte Packets | 320-Byte Packets |
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	54 ms	∞	74 ms	∞	61 ms**	74 ms**
Packet Loss Probability	0.0014	1	0.325	1	0.1*	0.05*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	50 ms	12 ms	115 ms	11 ms	12 ms
Packet Loss Probability	0.0001	0.2218	0.0001	0.2218	0.0001	0.0001
JAN						
End-to-End Delay	10 ms	∞	16 ms	∞	10 ms	16 ms
Packet Loss Probability	0.0163	1	0.0163	1	0.0163	0.0163
EAN						
End-to-End Delay	20 ms	N/C	20 ms	N/C	100 ms***	100 ms***
Packet Loss Probability	0	N/C	0	N/C	0.001***	0.001***

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

Note***: Upper bound from ITU-T Y.1541

N/A: Not included in the model

N/C: Kept constant

7.2.7   Path E

Path E consists of two PANs, two JANs, and the EAN.

The PAN component is stable with a very small contribution to network performance.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and the channel rate of the link.

The EAN is assumed to satisfy the network performance objectives given in ITU-T Y.1541.

We use a network budget requirement for the packet loss ratio of 10^{-3}, and an end-to-end transit delay of 100 ms given in ITU-T Y.1541.

Table 45:   Speech Path E Network Performance Parameter Requirements

Speech Path E	Calculated				Recommendation	
	80-Byte Packets		320-Byte Packets		80-Byte Packets	320-Byte Packets
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	33 ms	264 ms	40 ms	660 ms	100 ms**	140 ms**
Packet Loss Probability	0.00001	0.3944	0.0003	0.3944	0.1*	0.05*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
JAN						
End-to-End Delay	8 ms	122 ms	11 ms	320 ms	40 ms	65 ms
Packet Loss Probability	0.0003	0.3944	0.0003	0.3944	0.0508	0.0264
EAN						
End-to-End Delay	20 ms	N/C	20 ms	N/C	100 ms***	100 ms***
Packet Loss Probability	0	N/C	0	N/C	0.001***	0.001***

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

Note***: Upper bound from ITU-T Y.1541

N/A: Not included in the model

N/C: Kept constant

7.2.8   Path F

Path F consists of two PANs, one IAN, and one JAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel data rate on the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs and FRVs, and the channel rate of the link.

Table 46:   Speech Path F Network Performance Parameter Requirements

Speech Path F	Calculated				Recommendation	
	80-Byte Packets		320-Byte Packets		80-Byte Packets	320-Byte Packets
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	17 ms	∞	27 ms	∞	22 ms**	29 ms**
Packet Loss Probability	0.0007	1	0.0165	1	0.1*	0.05*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	49 ms	12 ms	115 ms	11 ms	12 ms
Packet Loss Probability	0.0001	0.2218	0.0001	0.2218	0.0007	0.0007
JAN						
End-to-End Delay	10 ms	∞	16 ms	∞	11 ms	17 ms
Packet Loss Probability	0.0163	1	0.0163	1	0.0164	0.0163

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

N/C: Kept constant

7.2.9　　Path G

Path G consists of two PANs, one IAN, two JANs, and the EAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel data rate on the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCD and FRVs, and the channel rate of the link.

The EAN is assumed to satisfy the network performance objectives given in ITU-T Y.1541.

We use a network budget requirement for packet loss ratio of 10^{-3}, and an end-to-end transit delay of 100 ms given in ITU-T Y.1541.

Table 47:　Speech Path G Network Performance Parameter Requirements

Speech Path G	Calculated				Recommendation	
	80-Byte Packets		320-Byte Packets		80-Byte Packets	320-Byte Packets
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	44 ms	∞	63 ms	∞	52 ms**	65 ms**
Packet Loss Probability	0.0007	1	0.0165	1	0.1*	0.05*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	49 ms	12 ms	115 ms	11 ms	12 ms
Packet Loss Probability	0.0001	0.2218	0.0001	0.2218	0.0007	0.0001

Table 47:   Speech Path G Network Performance Parameter Requirements (Continued)

Speech Path G	Calculated				Recommendation	
	80-Byte Packets		320-Byte Packets		80-Byte Packets	320-Byte Packets
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
JAN						
End-to-End Delay	10 ms	∞	16 ms	∞	10 ms	16 ms
Packet Loss Probability	0.0163	1	0.0163	1	0.0164	0.0163
EAN						
End-to-End Delay	20 ms	N/C	20 ms	N/C	100 ms***	100 ms***
Packet Loss Probability	0	N/C	0	N/C	0.001***	0.001***

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

Note***: Upper bound from ITU-T Y.1541

N/A: Not included in the model

N/C: Kept constant

7.3   Video Applications

The network performance recommendations necessary to support video applications are based on the user-perceived quality of service study for video applications described in the video portion of this volume. This section addresses MPEG-2 and H.264 encodings and the maximum acceptable eye-to-eye packet loss ratio and end-to-end transit delay given, also described in the video portion of this volume. These metrics are summarized in Section 7.3.1 and Section 7.3.2 for packet loss and delay, respectively. Section 7.2.3 through Section 7.2.9 provide the network performance budget for Paths A through G, respectively, in terms of the packet loss and end-to-end delay.

7.3.1   Packet loss Requirements

The video portion of this volume identifies the end-to-end packet loss requirements. Figure 14 illustrates these requirements from a user's eye (camera) to a user's eye (display). The video reference model described in the video portion of this volume includes a decomposition of the main contributors of loss. The reference lines in Figure 14 where the device has a common interface with the PAN, are equivalent to the C and D reference points described in the video portion of this volume.

Figure 14 shows the packet loss probability across its many components for the video application. From the video portion of this volume, we consider a maximum eye-to-eye packet loss ratio of 0.1 and 0.5 percent for H.264 and MPEG-2 encodings, respectively.

We use a network budget requirement of 10^{-3} given in ITU-T Y.1541 for the EAN.

Figure 14:   Video Maximum Packet Loss Ratio Requirements
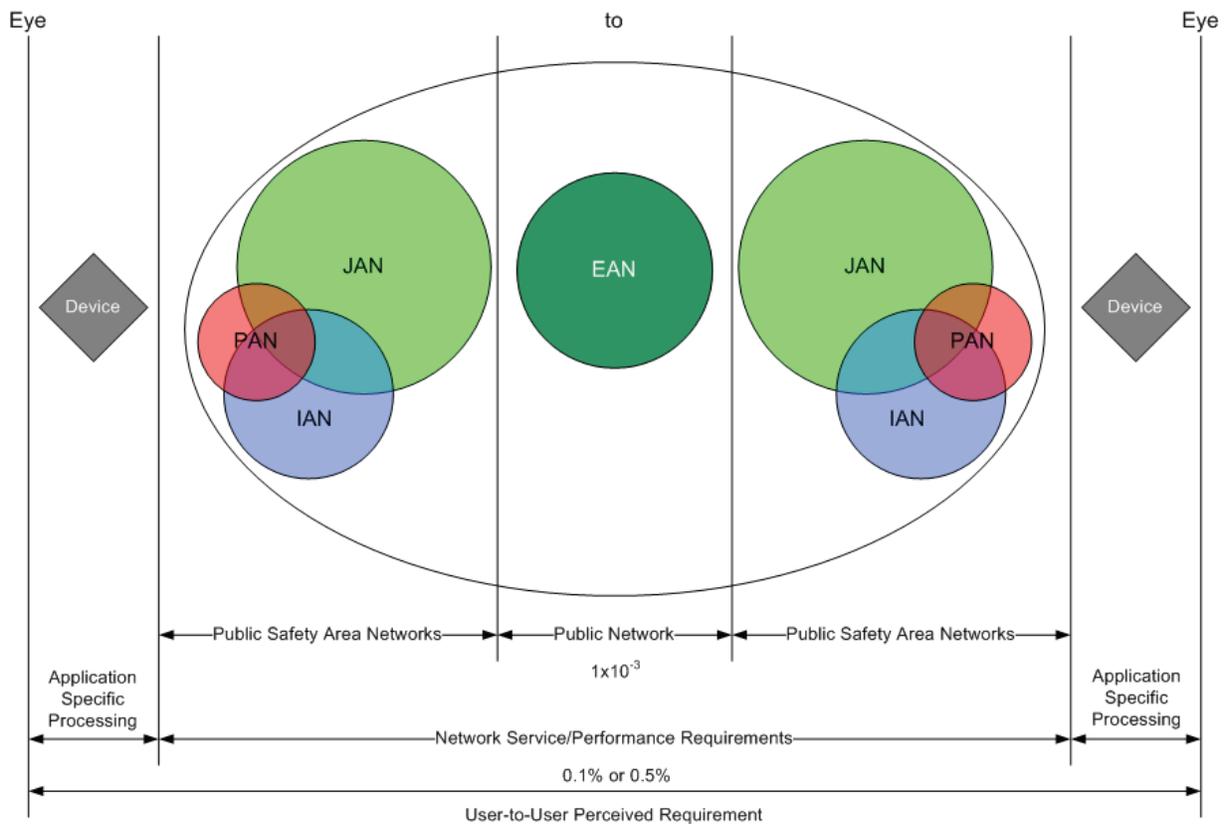
7.3.2   End-to-End Delay Requirements

Figure 15 illustrates the end-to-end delay for video from a user's eye (camera) to a user's eye (display). Given a maximum eye-to-eye end-to-end delay of 1 second described in the video portion of this volume, Figure 15 illustrates the recommended maximum end-to-end delay allocations presented in terms of path and other input parameters. The value of "x," which represents the delay contributed by the application-specific processing device, includes the packetization delay for video.

We use a network budget requirement for end-to-end transit delay of 150 ms given in ITU-T Y.1541 for the EAN.

Figure 15:   Video Maximum End-to-End Delay Requirements
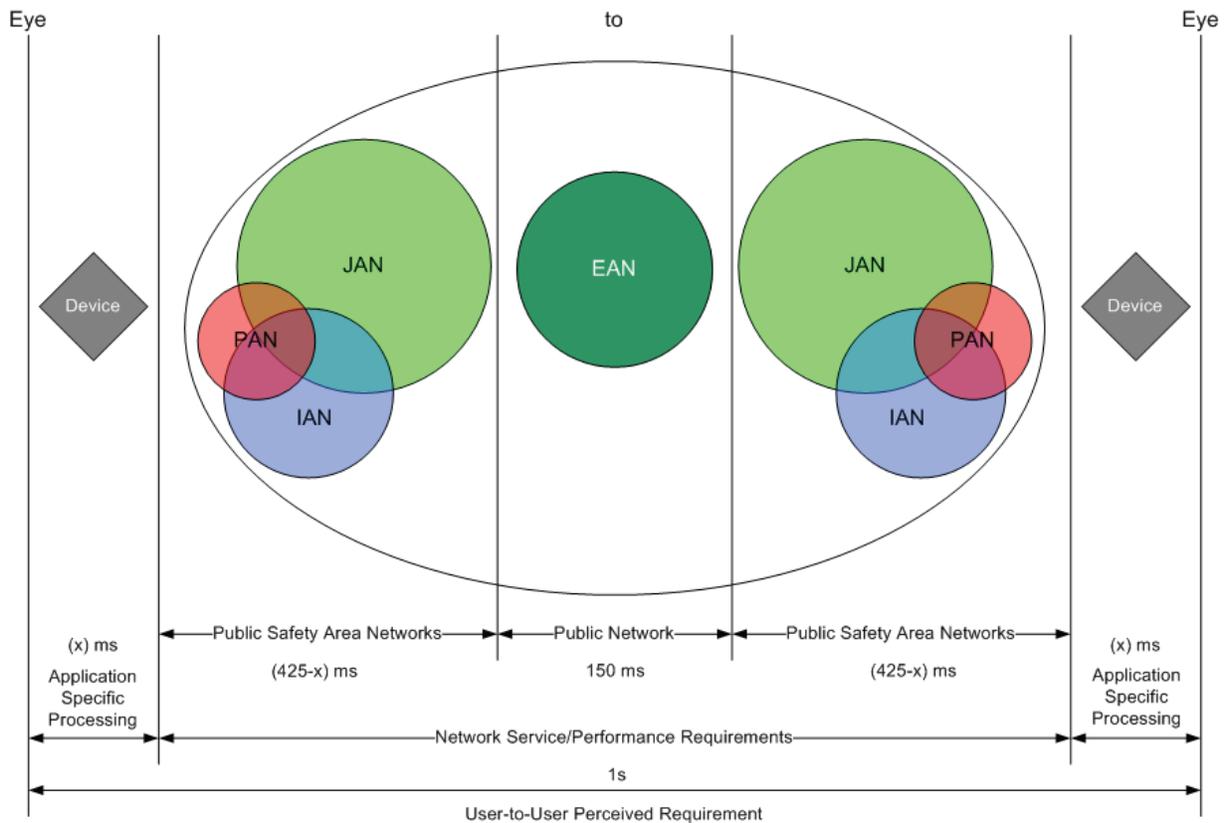
7.3.3   Path A

Path A consists of two PANs and one IAN. This path is one in which it is very easy to allocate the link budget for the various network performance parameters.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel rate of the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

Table 48:   Video Path A Network Performance Parameter Requirements

Video Path A	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	11 ms	42 ms	11 ms	227 ms	11 ms**	11 ms**
Packet Loss Probability	0	0.6239	0	0.9518	0.001*	0.005*

Table 48:   Video Path A Network Performance Parameter Requirements (Continued)

Video Path A	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
PAN						
End-to-End Delay	<< 1 ms	<< 1 ms	<< 1 ms	<< 1 ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	42 ms	11 ms	227 ms	11 ms**	11 ms**
Packet Loss Probability	0	0.6239	0	0.9518	0.001*	0.05*

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.3.4   Path B

Path B consists of two PANs and one JAN. This path, unlike path A, can contain multiple links within the JAN (i.e., not simply a communication path from PSCD to a single jurisdictional communication tower to another PSCD).

The PAN component is stable with a very small contribution to network performance.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and the channel rate of the link.

Table 49:   Video Path B Network Performance Parameter Requirements

Video Path B	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	7 ms	100 ms	7 ms	655 ms	8 ms**	10 ms**
Packet Loss Probability	0	0.6239	0	0.9518	0.001*	0.005*

Table 49:   Video Path B Network Performance Parameter Requirements (Continued)

Video Path B	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
PAN						
End-to-End Delay	<< 1 ms	<< 1 ms	<< 1 ms	<< 1 ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
JAN						
End-to-End Delay	7 ms	100 ms	7 ms	655 ms	8 ms**	10 ms**
Packet Loss Probability	0	0.6239	0	0.9518	0.001*	0.005*

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.3.5   Path C

Path C consists of two PANs and two IANs and one JAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel rate of the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, like the IAN component, is significantly affected by the choice of MAC, number of PSCDs and FRVs, packet size, and link channel rate.

Table 50:   Video Path C Network Performance Parameter Requirements

Video Path C	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	30 ms	∞	34 ms	∞	36 ms**	37 ms**

Table 50:   Video Path C Network Performance Parameter Requirements (Continued)

Video Path C	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
Packet Loss Probability	0.008	1	0.0303	1	0.001*	0.005*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	42 ms	11 ms	227 ms	13 ms	14 ms
Packet Loss Probability	0	0.3867	0	0.7804	0	0
JAN						
End-to-End Delay	10 ms	∞	14 ms	∞	10 ms	10 ms
Packet Loss Probability	0.004	1	0.0153	1	0	0

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.3.6   Path D

Path D is similar to Path C, except it contains the EAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel rate of the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, like the IAN component, is significantly affected by the choice of MAC, number of PSCDs and FRVs, packet size, and link channel rate.

We assume the EAN satisfies the network performance objectives given in ITU-T Y.1541. We use a network budget requirement for packet loss ratio of 10^{-3}, and an end-to-end transit delay of 100 ms given in ITU-T Y.1541.

Table 51:   Video Path D Network Performance Parameter Requirements

Video Path D	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	**Minimum**	**Maximum**	**Minimum**	**Maximum**	**Upper Bound**	**Upper Bound**
End-to-End Delay	59 ms	∞	68 ms	∞	65 ms**	67 ms**
Packet Loss Probability	0.008	1	0.0303	1	0.001*	0.005*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	42 ms	11 ms	227 ms	13 ms	14 ms
Packet Loss Probability	0	0.3867	0	0.7804	0	0
JAN						
End-to-End Delay	10 ms	∞	14 ms	∞	10 ms	10 ms
Packet Loss Probability	0.004	1	0.0153	1	0	0
EAN						
End-to-End Delay	20 ms	N/C	20 ms	N/C	100 ms***	100 ms***
Packet Loss Probability	0	N/C	0	N/C	0.001***	0.001***

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

Note***: Upper bound from ITU-T Y.1541

N/A: Not included in the model

N/C: Kept constant

7.3.7   Path E

Path E consists of two PANs, two JANs, and the EAN.

The PAN component is stable with a very small contribution to network performance.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and the channel rate of the link.

We assume the EAN satisfies the network performance objectives given in ITU-T Y.1541. We use a network budget requirement for the packet loss ratio of 10^{-3}, and an end-to-end transit delay of 100 ms given in ITU-T Y.1541.

Table 52:   Video Path E Network Performance Parameter Requirements

Video Path E	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	33 ms	219 ms	33 ms	1330 ms	35 ms**	38 ms**
Packet Loss Probability	0	0.6239	0	0.9518	0.001*	0.005*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
JAN						
End-to-End Delay	7 ms	100 ms	7 ms	655 ms	8 ms	9 ms
Packet Loss Probability	0	0.3867	0	0.7804	0.0005	0.0018
EAN						
End-to-End Delay	20 ms	N/C	20 ms	N/C	100 ms***	100 ms***
Packet Loss Probability	0	N/C	0	N/C	0.001***	0.001***

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

Note***: Upper bound from ITU-T Y.1541

N/A: Not included in the model

N/C: Kept constant

7.3.8 Path F

Path F consists of two PANs, one IAN, and one JAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel rate of the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs and FRVs, and the channel rate of the link.

Table 53: Video Path F Network Performance Parameter Requirements

Video Path F	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	20 ms	∞	24 ms	∞	23 ms**	24 ms**
Packet Loss Probability	0.0048	1	0.0184	1	0.001*	0.005*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	42 ms	11 ms	227 ms	13 ms	14 ms
Packet Loss Probability	0	0.3867	0.0002	0.7804	0	0
JAN						
End-to-End Delay	10 ms	∞	16 ms	∞	11 ms	17 ms
Packet Loss Probability	0.0008	1	0.0032	1	0	0

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

N/A: Not included in the model

7.3.9   Path G

Path G consists of two PANs, one IAN, two JANs, and the EAN.

The PAN component is stable with a very small contribution to network performance.

The IAN component, assuming a fixed node (FRV) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs, and channel rate of the link. The packet loss probability, however, is affected significantly by the choice of MAC and the number of PSCDs.

The JAN component, assuming a fixed node (jurisdictional communication tower) delay, has as its major performance-affecting factors: the choice of MAC protocols, choice of packet size, number of PSCDs and FRVs, and the channel rate of the link.

We assume the EAN satisfies the network performance objectives given in ITU-T Y.1541. We use a network budget requirement for the packet loss ratio of 10^{-3}, and for an end-to-end transit delay of 100 ms given in ITU-T Y.1541.

Table 54:   Video Path G Network Performance Parameter Requirements

Video Path G	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
End-to-End Delay	48 ms	∞	53 ms	∞	52 ms**	53 ms**
Packet Loss Probability	0.0048	1	0.0184	1	0.001*	0.005*
PAN						
End-to-End Delay	<< 1ms	<< 1ms	<< 1ms	<< 1ms	1 ms	1 ms
Packet Loss Probability	N/A	N/A	N/A	N/A	0.001	0.001
IAN						
End-to-End Delay	11 ms	42 ms	11ms	227 ms	13 ms	14 ms
Packet Loss Probability	0	0.3867	0	0.7804	0	0
JAN						
End-to-End Delay	10 ms	∞	14 ms	∞	10 ms	10 ms
Packet Loss Probability	0.004	1	0.0153	1	0	0

Table 54: Video Path G Network Performance Parameter Requirements (Continued)

Video Path G	Calculated				Recommendation	
	H.264		MPEG-2		H.264	MPEG-2
Details	Minimum	Maximum	Minimum	Maximum	Upper Bound	Upper Bound
EAN						
End-to-End Delay	20 ms	N/C	20 ms	N/C	100 ms***	100 ms***
Packet Loss Probability	0	N/C	0	N/C	0.001***	0.001***

Note*: Upper bound based on user-perceived mouth-to-ear delay study

Note**: Upper bound influenced by packet loss probability

Note***: Upper bound from ITU-T Y.1541

N/A: No included in the model

N/C: Kept constant

7.4 Summary for All Area Networks

Based on the results for speech and video applications in Section 7.2 and Section 7.3, a set of recommendations are created. These involve the maximum allowable packet loss and delay for each type of area network to meet the end-to-end quality of service requirements for both applications. Table 55 lists the upper bounds for them. In the event that area networks are combined together, use the path upper bounds instead.

Table 55: Maximum Allowable Packet Loss and Delay for Each Type of Area Network

Speech and Video Paths	Recommendation
Details	Upper Bound
PAN	
End-to-End Delay	1 ms
Packet Loss Probability	0.0001
IAN	
End-to-End Delay	14 ms
Packet Loss Probability	0
JAN	
End-to-End Delay	65 ms

Table 55:   Maximum Allowable Packet Loss and Delay for Each Type of Area Network (Continued)

Speech and Video Paths	Recommendation
Details	**Upper Bound**
Packet Loss Probability	0
EAN	
End-to-End Delay	100 ms
Packet Loss Probability	0.001

Appendix A   Glossary and Acronyms

A

ABA

America Bar Association

ACI

adjacent channel interference

A/D

analog-to-digital

ANSI

American National Standards Institute

ATIS

Alliance for Telecommunications Industry Solutions

ATSC

Advanced Television Systems Committee

AVI

Audio Video Interleave (file format)

B

B&W

black and white

C

C&I

Communications and Interoperability

CBR

constant bit rate

CC

color correction (a type of lens filter)

CCD

charge-coupled device

CCI

Co-channel interference

CIF

Common Intermediate Format

codec

coder-decoder

CRI

color rendering index

CSMA

carrier sensed multiple access

CSRC

Contributing SouRCe

D

DAT

digital audio tape

dB

decibel

dBA

dB A-weighted sound pressure level

DHS

Department of Homeland Security

DOC

Department of Commerce

DR

dynamic range

DSC

digital still camera

E

EAN

extended area network

EC

error concealment

EMS

Emergency Medical Services

F

FCC

Federal Communications Commission

FEC

forward error correction

FFT

fast Fourier transform

fps

frames per second

FR

frame rate

FRV

first responder vehicle

FVA

forensic video analysis

G

GOP

group of pictures

H

H.264

Also known as, MPEG-4 Part 10, or AVC (advanced video compression). A digital video codec standard that achieves very high data compression.

HD

high-definition

HDTV

high-definition television

HID

high intensity discharge (a type of photography lamp)

HMI

halogen metal iodide (a type of photography lamp)

HRC

hypothetical reference circuit

Hz

Hertz

I

i

interlaced video display scan (e.g., 1080i)

i3a

International Imaging Industry Association

IAN

incident area networks

IEC

International Electrotechnical Commission

IEEE

Institute of Electrical and Electronics Engineers

IETF

Internet Engineering Task Force

I-frame

Intra-coded frame

IP

Internet Protocol

IPv6

Internet Protocol version 6

IR

Infrared

ISI

inter-symbol interference

ISO

International Standards Organization

ITU

International Telecommunication Union

ITU-R

ITU Telecommunication Standardization Radiocommunications. See ITU.

ITU-T

ITU Telecommunication Standardization Sector. See ITU.

J

JAN

jurisdictional area network

K

K

Kelvin

kHz

kilo-Hertz

L

LMR

land-mobile radio

LUT

look up table

LW per PH

line widths per picture height

M

MAC

Medium Access Control (a network protocol)

MMCN

Multimedia Computing and Networking Conference

MOS

Mean Opinion Score

MPEG

Moving Pictures Expert's Group

ms

millisecond

MTF

modulation transfer function

MTF50P

MTF, where contrast drops to 50 percent of its peak value. See MTF.

N

NAL

network abstraction layer

ND

neutral density (a type of lens filter)

NFPA

National Fire Protection Association

NS

network section

NTIA

National Telecommunications and Information Administration

NTSC

National Television Systems Committee

O

OECF

opto-electronic conversion function

OIC

Office for Interoperability and Compatibility

P

p

progressive video display scan (e.g., 720p)

PAN

personal area network

PCM

pulse-code modulation

PDA

personal digital assistant

PLC

packet loss concealment

PLR

Packet Loss Ratio

PS

packet size

PSCD

public safety communications devices

PSNR

peak signal-to-noise ratio

PS SoR

Public Safety Statement of Requirements

PSWAC

Public Safety Wireless Advisory Committee

PSWC&I

Public Safety Wireless Communications and Interoperability

Q

QCIF

Quarter CIF. See CCI.

QSIF

Quarter SIF. See SIF.

QVGA

Quarter VGA. See VGA.

R

RFC

Request for Comments

RFI

radio frequency interference

RTCP

Real-Time Control Protocol

RTP

Realtime Transport Protocol

S

SD

standard definition

SED

Systems Engineering and Development

SFR

spatial frequency response

SIF

Source Input Format

SMIA

Standard Mobile Imaging Architecture

SMPTE

Society of Motion Picture and Television Engineers

SNR

signal-to-noise ratio

SPL

sound-pressure level

S&T

Science and Technology

T

TCP

Transmission Control Protocol

TDM

time division multiplexing

TDMA

time division multiple access

TIA

Telecommunications Industry Association

TV

television

U

UCL

University College London

UDP

User Datagram Protocol

UNI

user-to-network interface

V

VAD

voice activity detection

VCL

video coding layer

VCEG

Video Coding Experts Group

VGA

Video Graphics Array

VHS

Video Home System (recording media)

W

WB

white balance

WFQ

weighted fair queuing

Appendix B   Audio Measurement Methods and Tools

B.1   Laboratory Study

This section describes a laboratory study on the suitability of speech transmission systems. Specifically, public safety users listened to and evaluated a large number of recordings of speech transmission systems. The packet loss requirements given in Section 2.2 are based on the results of this laboratory study.

The study is a "human factors" study in the sense that human subjects experienced controlled stimuli and responded to them. The stimuli are combinations of sound recordings, so the study could be more specifically described as a "human listening" study. This is a well-developed field in telecommunications research and such studies are sometimes described as "listening tests," "subjective speech quality tests," or just "subjective tests." Where appropriate, the study follows the conventions and standards used in this field (e.g., P.800[27] and P.830[28]).To meet the unique demands of the question at hand, and the environment posed by public safety operations, some aspects of the study depart from convention.

B.2   Goal

In broad terms, the goal of the study is to determine how suitable or unsuitable different speech transmission systems would be for "mission-critical communications" in public safety operations. A key ingredient in meeting this goal is creating a well-controlled, yet realistic, simulation of the environment in which mission-critical speech communications occur. The twin goals of realism and control are generally at odds with each other, so the design of this study, like all studies in this field, required informed, calculated trade-offs or compromises.

Once the environment was simulated, practitioners were recruited to enter the environment, listen to recordings, and provide their opinions on what was suitable and what was not suitable. The following describes the various steps of the study in some detail, and provides rationales and results.

B.3   Methods

B.3.1   Message Transcription

To help create a realistic environment, it was desired that recordings carried realistic messages related to public safety operations. Toward that end, Internet-based "scanners" were used to monitor actual public safety radio traffic in several different U.S. locations. The traffic was transcribed, filtered for appropriateness and to remove duplication, and edited as necessary to protect anonymity. The transcribed messages were then classified by length, resulting in four categories of messages, listed as follows with the number of messages for each type:

- Tiny (one word) — 10 messages

- Short (several words) — 64 messages

27. ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality," Geneva, 1996.
28. ITU-T Recommendation P.830, "Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs," Geneva, 1996.

- Medium (typically a full sentence) — 200 messages

- Long (typically more than one sentence) — 200 messages

All messages were in the English language. Two example messages from each category follow:

- Tiny:

    "Negative."

    "Copy."

- Short:

    "They will not extradite."

    "Sure go ahead."

- Medium:

    "Do you need another unit there for traffic control?"

    "White males in their mid 50's, one in a green baseball cap."

- Long:

    "Meet the complainant at 145 Riverside. The citizen thinks he found a body in the woods near by. I also have fire responding."

    "The suspect took a bag out of her vehicle, ran down the street, and put it in another vehicle."

B.3.2   Message Recording

Two female and three male adults were selected to read the messages during recording sessions. Each person, also called a talker, read 40 messages from each category, or the maximum number of messages available. In the tiny category, each talker read all ten messages. In the short category, some of the talkers read messages 1 through 40, and others read messages 25 through 64. In the medium and long categories, each message was read by a single talker (5 talkers × 40 messages/talker = 200 messages).

The recording sessions were conducted in a sound-isolated chamber, with ambient noise level below 30 decibels (dB) A-weighted sound pressure level (dBA). A single, studio-quality microphone with a cardioid pickup pattern was used. This microphone includes an integral analog-to-digital (A/D) converter, and the digital output (48,000 samples per second, 24 bits per sample) was connected to a digital audio interface hosted by a personal computer. A sound editing software tool controlled the recording process. To preserve the most fundamental and pristine speech signal, no equalization was used in the recording process. The original digital recordings were then segmented into a sequence of shorter recordings that included just the desired messages, and excluded any mistakes in reading, or extraneous noises.

B.3.3   Addition of Transmit Location Background Sound

The recorded messages are of studio quality, and have virtually no background sound. Background sound in actual public safety operations is often, if not always, present. Futher, background sound types and levels can very greatly between locations. The use of messages with virtually no background sound would be unrealistic, yet including the effect of background sound at transmit locations as a factor would have been well beyond the scope of the study. As a compromise, background sound was added to the messages at a single low level, resulting in a 25 dB signal-to-noise ratio (SNR). This background sound was recorded

using studio quality equipment in three different field locations: on a street corner, in a passenger car, and in an office.

The 25 dB SNR was selected empirically. It is high enough to be clearly perceptible, yet it is low enough that it does not interfere with judgments of the speech transmission systems of interest. Future studies might be conducted to characterize the effects of higher background sound levels, and specific noise types at transmit locations.

B.3.4   Message Concatenation

Recorded messages were concatenated to form 182 different meta-messages. A meta-message may contain 11, 12, 13, 14, or 15 messages. Each meta-message was at least 30 seconds long, and the average length was 32 seconds.

The messages in each meta-message were selected using a constrained randomization procedure. This procedure produced a good mixture of talkers and message lengths without resorting to a repetitive pattern.

All five talkers were heard once before any talker is heard a second time. No talker was heard twice in immediate succession. The messages within a meta-message were not connected logically, rather they produced the effect of scanning across multiple public safety communications channels. Following are two example meta-messages:

- Example Meta-Message 1:

    "552, criminal trespasses. Ladder 1, move up to Station 32. Standby Tom 33, they already have someone on the way. Sure, go ahead. 426 copy, a missing person. William 2 en route. Alright. Do you want me to call and have county get out of the area? Thanks. Transporting one adult male to West Precinct. Goodnight 1428. This is the Ladder 4, false alarm, going green and back in service. He should be getting a call right now. 073 Henry, I'm back. Engine 32 aid response."

- Example Meta-Message 2:

    "He also said yesterday it looked like there were two pit bulls inside with the suspect. Affirmative. Ladder 10, code red. Disregard. Engine 21 has arrived. A 15-year-old female missing from Riverside Lane. Disregard. If I take him down will they take him? Disregard. Can you see that person, and what house he went into? This is Engine 6, one truck for manpower please. Affirmed. Negative. They have an urgent call on the hill, units are asking for assistance. Please switch to East."

B.3.5   Speech Transmission Systems

The study focuses on G.711 speech coding accompanied by the PLC algorithm specified in G.711 Appendix I. The study uses μ-law G.711, since that is the option used throughout North America. Note that G.711 is typically used to deliver a nominal speech passband that extends to 3400 Hertz (Hz). Potential future studies may investigate any possible benefit in transmitting a passband that extends beyond 3400 Hz (e.g., 7 kilo-Hertz (kHz) wideband speech coding or 15 kHz audio coding).

The speech coding and data transmission processes simulated in the laboratory study are shown in Figure 16. As shown in this figure, incoming speech signals are encoded and the resulting data is placed into packets which are passed to the network. Packets that emerge from the network are processed to extract the G.711 data stream, which is placed into a jitter buffer to accommodate delay variation in the

network. Following G.711 decoding, the PLC algorithm attempts to hide the effects of lost channel data, and generates the final output speech signal.

Figure 16:   Simulated Speech Coding and Data Transmission Processes
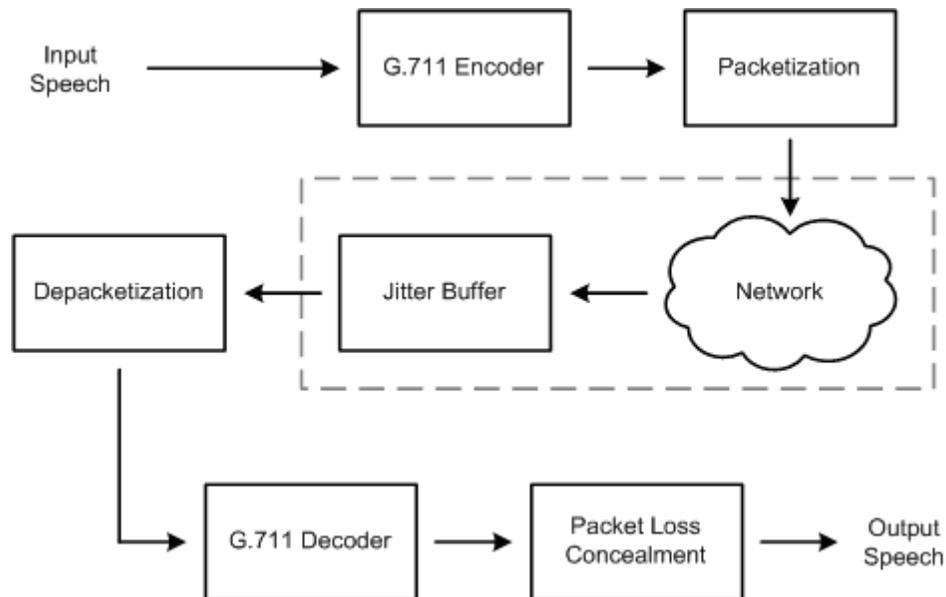
The simulation of these processes treats the network and the jitter buffer together as a single black box that can be parameterized (for at least tens of seconds) by a pair of packet loss parameters. These are fundamental properties and thus they provide a basic yet relevant model.

The two packet loss parameters are packet loss ratio and packet loss correlation. When packet loss correlation is zero, the packet loss process is random. As packet loss correlation is increased, the loss of packets becomes more bursty, and it becomes more likely that multiple packets will be lost in succession. In practice, packetized data networks can exhibit random or bursty packet loss patterns.

When the data network is supporting speech transmission, random packet losses generally are less offensive than bursty packet losses. When packet losses are random, it is more likely that isolated packets (not consecutive packets) will be lost. PLC algorithms can hide much of the effect of an isolated lost packet. As packet losses become more bursty, it becomes more likely that multiple consecutive packets will be lost. PLC algorithms loose effectiveness when asked to conceal extended periods of missing data, and they generally stop working altogether after 60 to 90 ms.

In this study, packet loss is modeled by a two-state Markov channel model[29] as shown in Figure 17. Packet loss ratio $(0 < \mu < 1)$ and packet loss correlation $(0 \leq \gamma < 1)$ determine the transition probabilities in the Markov model according to $p = (1 - \gamma) \times \mu$, and $q = (1 - \gamma) \times \mu + \gamma$. Conversely, we have $\mu = p / (1 + p - q)$, and $\gamma = q - p$.

When $\gamma = 0$, we have $q = p = \mu$. That is, the unconditional probability of loss (μ) and both conditional probabilities of loss (p and q) are identical, as they must be for random (independent) losses. Note also that the mean length of a loss is $1 / (1 - q)$ packets, and the standard deviation of a loss is $\sqrt{q} / (1 - q)$ packets.

---

29.   H. Sanneck and N. Le, "Speech Property-Based FEC for Internet Telephony Applications," *Proc. SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000*, San Jose, CA, January 2000.

Figure 17:   Two-State Markov Channel Model Used in Laboratory
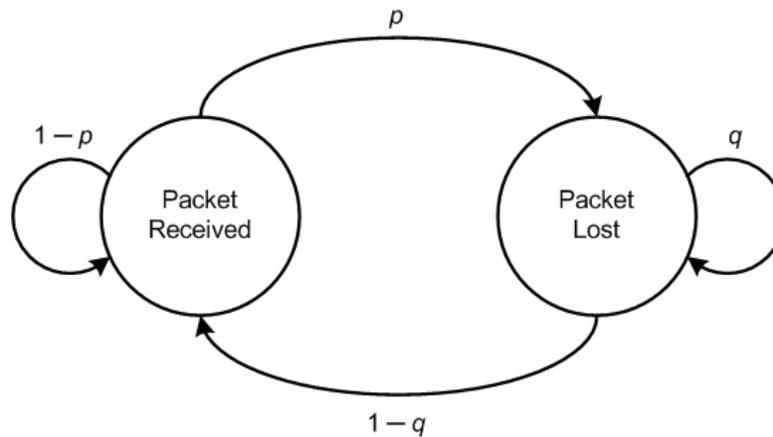
Future studies may incorporate more detailed network models, perhaps specifically tuned to the hybrid network environment (containing both wired and wireless links) that is likely to be used in future systems.

This study uses packet sizes 10 and 40 ms, meaning that G.711 data corresponding to either 10 or 40 ms of speech signal is assigned to a single packet. In practical systems, packet headers add overhead data to each packet, so data efficiency (ratio of payload data per packet to total data per packet) can be increased by using larger packets. On the other hand, larger packets induce larger packetization delay and cause worse speech impairments when a single packet is lost. The choice of a packet size involves balancing these two effects. Based on choices used in current packetized speech transmission systems, 10 and 40 ms packet sizes may approximately represent the endpoints of a viable range.

The meta-messages were passed through software implementations of various speech transmission systems of interest. The majority of these systems were G.711 speech coding, followed by packetization, packet loss, decoding, and PLC, reflected conceptually in Figure 16. The actual implementation followed these steps:

1.   Used software to perform bandpass filtering (nominally 200-3600 Hz) of the speech consistent with G.712.[30] This filtering is the standard preparation for G.711 speech coding.

2.   Used software to perform G.711-compliant speech encoding.

3.   Used software to implement the Markov packet loss model shown in Figure 17, and accordingly deleted blocks of channel data consistent with one of two packet sizes: 10 ms or 40 ms.

4.   Used software to perform G.711-compliant speech decoding.

5.   Used software to perform G.711 Appendix I-compliant packet loss concealment.

---

30.   ITU-T Recommendation G.712, "Transmission Performance Characteristics of Pulse Code Modulation Channels," 2001.

In Table 56, combinations of packet loss ratios and packet loss correlation values used in the study are indicated by "X."

Table 56:   Laboratory Packet Loss Ratios and Packet Loss Correlation Values

Packet Loss Correlation	Packet Loss Ratio						
	2 Percent	5 Percent	10 Percent	20 Percent	30 Percent	40 Percent	50 Percent
0.0	X	X	X	X	X	X	X
0.1	X	X	X	X	X		
0.2	X	X	X	X	X		
0.3	X	X	X	X	X		
0.4	X	X	X	X			
0.5	X	X	X	X			
0.6	X	X	X				
0.7	X	X	X				
0.8	X	X					
0.9	X						

The Markov model software uses a pseudo-random number generator to drive the state transitions. Thus the finite-length loss patterns produced have approximately, but not exactly, the desired packet loss ratio. To tightly control this factor, the actual packet loss ratio of each loss pattern was checked. A loss pattern was retained and used only when the measured packet loss ratio was within 5 percent of the target packet loss ratio. For example, when the target packet loss ratio is 0.1000 (commonly denoted as 10 percent) the actual packet ratio is guaranteed to be between 0.095 and 0.105 (commonly denoted as 9.5 percent and 10.5 percent).

Table 56 shows 39 different combinations of packet loss ratios and packet loss correlation values. When these 39 cases are crossed with 10 ms and 40 ms packet sizes, the result is 78 speech transmission systems. Another system arises from the case of no packet loss, which is unaffected by to packet loss correlation and packet size.

Nine additional "speech transmission systems" were simulated in software for a total of 88 systems in this study. These nine systems were included for their potential use in future calibration processes.

In all cases, proper care was taken to present the proper sample rate (8000 samples per second) and active speech levels (26 dB below the clipping point) to the software implementations. The result of these steps is a second set of meta-messages — the processed meta-messages. Each of these processed meta-messages represents the output of a specific speech transmission system that a public safety practitioner would hear at a receiving location. For each system under consideration, eight different processed meta-messages, containing a total of about four minutes of messages, were produced.

B.3.6 Laboratory Conditions

Public safety practitioners evaluated the processed meta-messages in acoustically controlled laboratory conditions. We attained the necessary acoustic control through the use of sound-isolated chambers, located inside of an already quiet laboratory space. The inside dimensions of each chamber, effectively a room within a room, were about 9-feet wide, 10-feet long and 7-feet high. The background noise level inside of these chambers was below 30 dBA.

Figure 18 illustrates the layout of the sound-isolated chambers (top view). Processed meta-messages are played through speaker marked A. Background sound is played through speakers marked B.

Figure 18:   Laboratory Layout
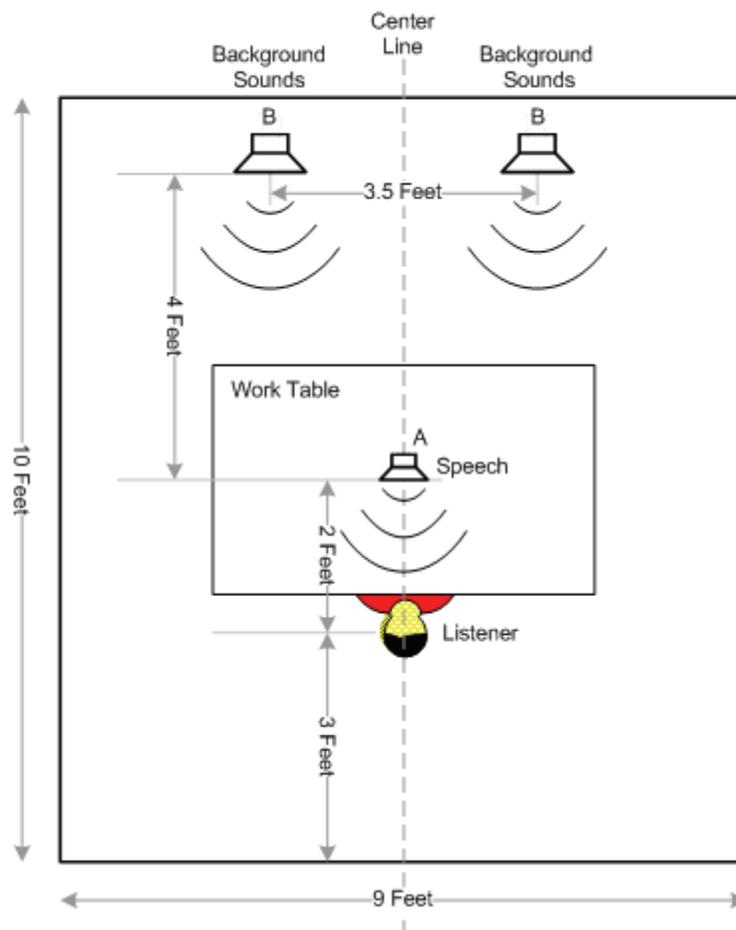
The processed meta-messages were played through a small, studio-quality loudspeaker, equipped with a volume control, located on a table. In each evaluation, a single practitioner was seated at this table, and was encouraged to adjust the volume level at any time. Through this arrangement, the processed meta-messages could be reproduced at nominal levels up to about 80 dBA at the seating location. The default starting position of the volume control resulted in a nominal level of about 65 dBA at the seating location. The majority of the practitioners did not change this volume control (or at least it was in the default position at the end of the session.) Some practitioners did leave the volume control above the default position by 4 to 7 dB; no practitioner left the volume control in the extreme upper position.

While the chambers isolated practitioners from virtually all uncontrolled background noises, such a quiet environment is certainly not typical for public safety operations. In actual public safety operations, background sound types and levels can vary greatly between locations, and may even be high enough to make communications difficult. The effect of high levels of background sound at listening locations is an operational factor that is outside the scope of the present study of equipment factors. On the other hand, any evaluation of equipment factors should certainly be made in an environment that is at least somewhat realistic. Thus to better represent the acoustic environments in which practitioners would actually listen to messages, we introduced additional prerecorded background sound into each chamber in a controlled way.

The two background sound levels chosen reflect this compromise: they are high enough to be clearly audible, yet low enough to have only a minor effect on the votes submitted by the practitioners. This prevents the background sound from obscuring the desired information regarding equipment factors. The higher level was nominally 60 dBA measured at the practitioner head location, and the majority of the level readings for this setting fell in the 55 to 65 dBA range. Preliminary trials with informal self-reporting revealed that this level does not create significant annoyance, fatigue, or distraction. The lower of the two sound levels was 15 dB lower. A significant step in level was desired, yet it was also desirable for the background sound to still be clearly audible. The lower level was nominally 45 dBA and the majority of the sound-pressure level (SPL) readings for this setting fell into the 40 to 50 dBA range.

The background sounds were prerecorded on location using studio quality equipment. They included street corner sounds, a fire truck at idle, and a distant siren. These recordings were mixed and processed to attain a monophonic signal with the desired level characteristics, and this signal was then transferred to digital audio tape (DAT) for storage. While practitioners were in the chamber, the monophonic DAT was played through a pair of studio monitor speakers (identical signal fed to each speaker) marked B as shown in Figure 18.

B.3.7   Evaluation by Users

Thirty-five public safety practitioners were recruited to participate in the laboratory listening evaluations. The practitioners came from across the country. Various local jurisdictions were represented 29 times, state jurisdictions 6 times, and Federal jurisdictions 2 times. Public safety first responders represented include firefighters, law enforcement officers, and Emergency Medical Services (EMS), with 18, 13, and 9 practitioners, respectively. (Some practitioners represented more than one jurisdiction or discipline.) Three of the practitioners had less than 10 years experience, 11 had between 10 and 20 years experience, 14 had between 20 and 30 years experience, and 7 had more than 30 years of experience. Thirty-four of the practitioners were males and one was female. Roughly 25 percent were in their thirties, about 50 percent were in their forties, and 25 percent were in their fifties. Thirty of the practitioners appeared to be of European heritage, four of Hispanic heritage, and one of African-American heritage.

Practitioners participated one at a time. This allows each to proceed at his or her desired pace, and prevents practitioners from exerting any intentional or unintentional influence on each other. A laboratory study administrator read identical instructions to each practitioner, the key portion of which follows:

> Thank you for taking time to participate in this listening experiment. The experiment involves no risk or discomfort, and you are free to end your participation at any time with no penalty. This experiment is one small part of a major process that is being used to design future public safety communications systems.
>
> In this experiment you will hear a recorded sequence of messages and you are asked to decide whether or not the speech quality in the sequence of messages is suitable for mission-critical communications, given the job that you do. Once you have decided, you will press the appropriate

button on the screen in front of you. Each sequence of messages is about 30 seconds in length, but you can make your decision at any time.

If you press the "replay" button, the exact same recording will start again, and it will have the exact same distortions in the exact same places. (Note that this is different than asking a coworker to repeat him or herself.) You can press the "replay" button at any time.

After you have indicated your decision on a recording, the next recording will start immediately.

You are welcome to adjust the volume of the recordings to any level at any time using the knob on the speaker in front of you. Other speakers in this room will play background noise. Several different levels of background noise will be used, and you will not be able to control the volume of the background noise. However, if you start to feel any discomfort from the background noise, please let me know, and I will adjust the level.

Please note that there are no right or wrong answers, we are seeking your opinion.

We want to know what speech quality you personally would consider suitable for mission-critical communications in the job that you do. We ask that you base your judgments on any distortions, noises, or other impairments caused by the communications system. We ask that you not include the pleasantness and diction of the voices, the content of the messages, nor the quality or content of the background noise.

No user asked to have the background sound level adjusted, so the same nominal background sound levels defined in Section B.3.6 were used for all 35 users.

Specialized software aided the actual listening process. This software initiates the playback of a processed meta-message, and waits for the practitioner to respond. The practitioner used a Personal Digital Assistant (PDA) with a wireless network connection. The following question was prominent on the screen: "Is the speech quality suitable for mission-critical communications?" Also prominently on the display were two response buttons marked "yes" and "no." Users were allowed to vote on each processed meta-message at any time after it began playing, and the playback could be restarted at any time with a "replay" button. Once a vote was collected, the playback of the next processed meta-message would begin.

Users did not know what speech transmission systems they were hearing, and a different random order was used in each listening session. Users did receive text indicating their progress through each session (e.g., "Finished with 10 of 88 trials.") Each practitioner first completed a training session that exposed the practitioner to a wide range of speech transmission systems, and allowed the practitioner to become familiar with the test procedure and equipment. Votes from these training sessions were discarded.

After the training session, each practitioner participated in four sessions, each containing 88 processed meta-messages, and thus provided a total of 352 votes. Breaks were offered between each of the sessions so that subjects could refresh themselves if they wished to. The 35 practitioners moved at different paces through the 352 recorded processed meta-messages. The total time spent listening ranged from 18 to 107 minutes, with an average total listening time of 58 minutes, which is about 14.5 minutes per session, or 10 seconds per processed meta-message.

Sessions 1 and 2 used the 45 dBA nominal background sound level, while Sessions 3 and 4 (and the training session) used the 60 dBA nominal background sound level. Sessions 3 and 4 used the same processed meta-messages as Sessions 1 and 2 respectively, but they were played in different random orders. All practitioners heard each system once in each session. To maximize the number of processed

meta-messages heard with each system, four different versions of each session were used, each containing different sets of processed meta-messages. Since Sessions 3 and 4 contained the same processed meta-messages as Sessions 1 and 2 respectively, a total of eight different processed meta-messages (four versions per session × two distinct sessions) were used to evaluate each system. On average, eight meta-messages contain 104 messages.

B.4   Analysis of Votes

In this study, a total of 12,320 votes were collected (88 processed meta-messages per session × 4 sessions per practitioner × 35 practitioners). Of these votes, 7,452 (60.5 percent) were "yes" votes indicating a good overall balance (suitable versus not suitable) of the systems included in the study. The requirements given in Section 2 are based on a statistical analysis of the 12,320 votes. More specifically, the 12,320 votes break into 140 votes (35 practitioners × 4 votes per practitioner) for each system considered. These 140 votes per system can include variation due to the use of eight different processed meta-messages, the use of 35 different practitioners, and the use of two different levels of controlled background sound.

The analysis exercised here views these 140 votes to be samples of the pool of all possible votes that could be cast for that system by the entire body of public safety practitioners in this country, hearing all possible messages. Since each vote can be "yes" or "no," Bernoulli Trials and the underlying binomial distribution provide a model for this voting process. If the samples (the votes collected in this study) are representative with respect to parameters that effect the votes of the larger pool, then we can use the 140 votes to find an estimate of the votes in that larger pool. Specifically, we can find a maximum likelihood estimate of the underlying parameter p (interpreted as the probability of success, or a "yes" vote), in the binomial distribution. Further, we can calculate an interval $[p_{low}, p_{high}]$ that is 95 percent certain to contain the true value of p in the larger pool.[31]

First we consider the effect of the two different background sound levels used during the practitioner evaluations. When we compare statistical variations between sessions, we find that the variations between sessions with different background sound levels are no greater than the variations between sessions with the same background sound level. This analysis concludes that changing receive location background sound level does not have a statistically significant effect in this particular experiment. This is a helpful result. It suggests that the background sound levels used are not influencing the results significantly, yet they certainly increase the realism of the study.

Based on the background sound level result, the remaining analyses aggregate votes from both background sound levels, thus corresponding to a nominal background sound level between 45 and 60 dBA. Compared to many operating environments, this would be a conservative background sound level, and in light of the results above, we would expect the resulting requirements to be conservative as well. The remaining analyses use the 140 votes collected for each system to produce estimates on the true value of p in the larger pool of all users and all messages.

For each system, we compute a range $[p_{low}, p_{high}]$ that is 95 percent certain to contain the true value of p in the larger pool, if the sample is representative in all relevant aspects. From this calculation, subject to the constraints and assumptions detailed in Section 2 and Appendix B of this document, we then conclude that "it is expected that at least p_{low}×100% of the public safety practitioners will find the resulting speech transmission to be suitable for mission-critical communications." Section 2 provides results of this type, organized according to three different thresholds.

---

31.   N. Johnson, S. Kotz, and A. Kemp, *Univariate Discrete Distributions, Second Edition*. New York: Wiley, 1992.

Figure 19 provides an example of the results attained. It shows the estimated values of p (fraction of "yes" votes), and 95 percent confidence intervals $[p_{low}, p_{high}]$, for those estimates for the systems associated with random packet loss $(\gamma = 0)$. The upper (blue) line corresponds to 10 ms packet size, and the lower (red) line corresponds to 40 ms packet size. As expected, estimates of p decrease as packet loss ratio increases. In addition, the 40 ms packet size generally results in lower estimates of p since the loss of 40 ms of speech is harder to conceal than the loss of 10 ms.

Figure 19:   Example Results for Random Packet Loss
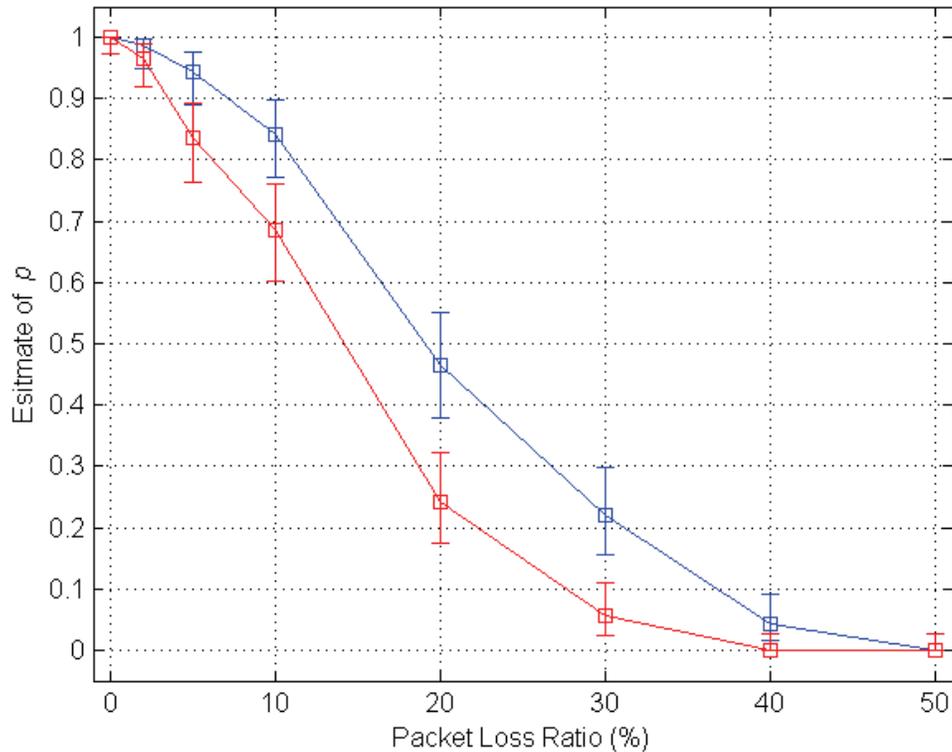
Appendix C   Video Acquisition Measurement Methods

DISCLAIMER

Certain commercial equipment, materials, and software are identified to specify the technical aspects of the reported procedures and results. In no case does such identification imply recommendations or endorsement by the National Telecommunications and Information Administration (NTIA), the Department of Commerce (DOC), or the U.S. Government, nor does it imply that the equipment, materials, and software identified are the best available for this purpose.

No warranty, expressed or implied, is made by NTIA, DOC, or the U.S. Government as to the accuracy, suitability, and functioning of any equipment, materials, or software given herein, nor shall this constitute any endorsement by the U.S. Government.

C.1   Existing Camera Performance Standards

Several sets of standards exist for measuring digital camera performance. Two sources of particular interest are the International Standards Organization (ISO),[32] and the Standard Mobile Imaging Architecture (SMIA), which publishes a camera characterization specification.[33]

The camera performance measurements described here have been designed to be performed at moderate cost with moderately skilled operators. They generally involve photographing simple or standard targets under controlled lighting conditions and then analyzing the resulting images on a computer. The tests do not require expensive or highly specialized equipment.

Within the video transmission system, the tests measure the quality of the video acquisition subsystem (i.e., the video camera). In general, video acquisition quality may be divided into two aspects: still image and motion properties. Motion quality factors are difficult to measure. The most serious arise from image compression artifacts due to video coders. The tests described here are not intended to specify performance parameters for video coders, which may be an integral part of some video acquisition subsystems. Instead, performance parameters for video coders (e.g., frame rate) are considered as part of the video transmission subsystem.

The focus of this appendix is on video acquisition (i.e., camera) performance parameters that are important for public safety applications. Most of the tests that will be described were originally designed for still cameras and adapted for use with video cameras. All the tests require that one or more still frames be captured from the video camera. One major difference between still and video frames is low light performance. With video, there is little choice of shutter speeds and long exposure times cannot be used to compensate for dim lighting conditions. Dim lighting performance must therefore be characterized by exposure accuracy and noise.

Video acquisition quality is primarily affected by two factors that arise at different stages of the imaging process:

---

32. Standards source for measuring digital camera performance. International Organization for Standardization. http://www.iso.org/. Cited June 14, 2006.
33. *SMIA 1.0 Part 5: Camera Characterisation Specification*, Rev A March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

- Capture—Image quality factors affected by the sensor and lens. These include sharpness, noise (total, fixed pattern, and dynamic), dynamic range, exposure uniformity (vignetting), and color quality. Exposure, which is set at capture time, is also important. There is a tradeoff between pixel size and quality: small pixels provide greater image resolution but suffer more from diffraction and photon shot noise, which are fundamental effects of the wave and particle nature of light.

- Post-capture image processing—Factors include white balance, sharpness (as affected by sharpening), color saturation, and tonal response. These factors are not intrinsic to the camera sensor and lens, but they can be important in real-time video systems, where there may be little or no opportunity to enhance the image after capture.

C.2 Lighting Conditions Terminology

The definitions in Table 57 are required to specify lighting conditions to be used for the parameter measurements.

Table 57:   Lighting Terminology

Term	Description
Standard Lighting Intensity	Approximately 200 to 500 lux (a lux is equal to one lumen per square meter) with ± 10% uniformity over the test chart.
Reduced Lighting Intensity	Approximately 30 to 60 lux with ± 10% uniformity over the test chart.
Dim Lighting Intensity	Approximately 5 to 10 lux with ± 10% uniformity over the test chart.
Color Temperature	The color of the illuminating lamp, defined as the temperature (in degrees Kelvin (K)) at which a heated, black-body radiator matches the hue of the lamp. One key issue involving color temperature is the ability of the camera's white balance algorithm to adapt to light with different color temperatures.
Tungsten Light	Light that has a color temperature between 2,800 and 3,200K.
Daylight Light	Light that has a color temperature between 5,500 and 7,500K.
Neutral Density (ND) Filters	Uncolored filters specified by their density ($-\log_{10}$(light absorption)). These are placed in front of the light sources or camera lens to achieve reduced or dim lighting. Typical values are D = 0.3 (2x; 1 f-stop), 0.6 (4x; 2 f-stops), and 0.9 (8x; 3 f-stops). When filters are stacked, the density is summed. For example, if two SoLux Task Lamps located 1 meter from the target provide approximately 250 lux at the target, 0.6 + 0.9 ND filters (for a density of 1.5, which produces a decrease in lighting intensity by a factor of $2^{(1.5/0.3)} = 2^5 = 32$) can reduce the illumination to 250/32 = 7.8 lux, which is in the range of dim lighting. You can make fine adjustments by moving the lamps.

Table 57:   Lighting Terminology (Continued)

Term	Description
Color Correction (CC) Filters	Filters that alter the color temperature of light reaching the camera. These are placed in front of the lens or the light source. Degradation from heat can be a problem near strong light sources. The best-known CC filters are the Wratten series 80 (strong cooling), 81 (subtle warming), 82 (subtle cooling), or 85 (strong warming). "Warming" means *decreasing* color temperature and "cooling" means *increasing* color temperature. Several filters correspond to each number in the series, e.g., 80A, 80B, 80C, etc., each of which alters color temperature by a different amount. CC filters alter color temperature by a fixed number of mireds (micro-reciprocal degrees), where 1 Mired = 10^6/(degrees K). Example: The Wratten 80A filter (the strongest standard cooling filter) changes color by -131 mireds, equivalent to increasing color temperature from 3,200K to 5,500K. It also reduces light by 2 f-stops.

Blue Filters				Amber Filters			
Filter	**Exposure Increase**	**Conversion**	**Mired**	**Filter**	**Exposure Increase**	**Conversion**	**Mired**
80A	2	3,200–5,500K	-131	81	1/3	3,300–3,200K	+9
80B	1 1/3	3,400–5,500K	-112	81A	1/3	3,400–3,200K	+18
80C	1	3,800–5,500K	-81	81B	1/3	3,500–3,200K	+27
80D	2/3	4,200–5,500K	-56	81C	1/3	3,600–3,200K	+35
82C	2/3	2,800–3,200K	-45	81D	1/3	3,700–3,200K	+42
82B	2/3	2,900–3,200K	-32	81EF	1/3	3,850–3,200K	+53
82A	1/3	3,000–3,200K	-21	85C	2/3	5,500–3,800K	+81
82	1/3	3,100–3,200K	-10	85	2/3	5,500–3,400K	+112
				85B	2/3	5,500–3,200K	+131

Term	Description
Middle Gray Surface	A neutral gray-colored surface with approximately 18 percent reflectance. A middle gray surface provides a useful background for test charts since it influences the auto-exposure algorithm and helps to obtain a good exposure. For the tests presented here, it is sufficient to have a good visual match with a surface of approximately middle gray, which includes patch M (7) on the Kodak Q-13or Q-14 Gray Scale or patch 22 (4th from left on the bottom row) in the GretagMacbeth ColorChecker (see Figure 21, fourth from left on the bottom row). Examples of middle gray surfaces that can be used include Crescent Matte board 1074 (Gibraltar Gray), 935 (Copley Gray), and 976 (Bar Harbor Gray).

C.3   Standard Test Chart Setup

Mount all of the test charts described in Section C.3.1on a flat background, preferably a half-inch foam board, because this is lightweight and stays flatter than standard-thickness foam boards (both foam baord types are widely available at art supply stores). Follow the procedures in Section C.3.2 to ensure charts are uniformly illuminated.

C.3.1   Standard Test Charts

Use the standard test charts in this section to measure resolution, noise, dynamic range (indirect method), color accuracy (and white balance), and lens distortion. Later sections describe specialized test patterns and methods for directly measuring dynamic range (see Section C.4.3) and for measuring flare light distortion (see Section C.4.10).

C.3.1.1   The ISO 12233 Test Chart

Figure 20 is a sample video frame of the ISO 12233 test chart that was captured using a high-definition (HD) video camcorder. This chart can be used for measuring resolution.

Figure 20:   ISO 12233 Resolution Test Chart Captured Using an HD Video Camcorder
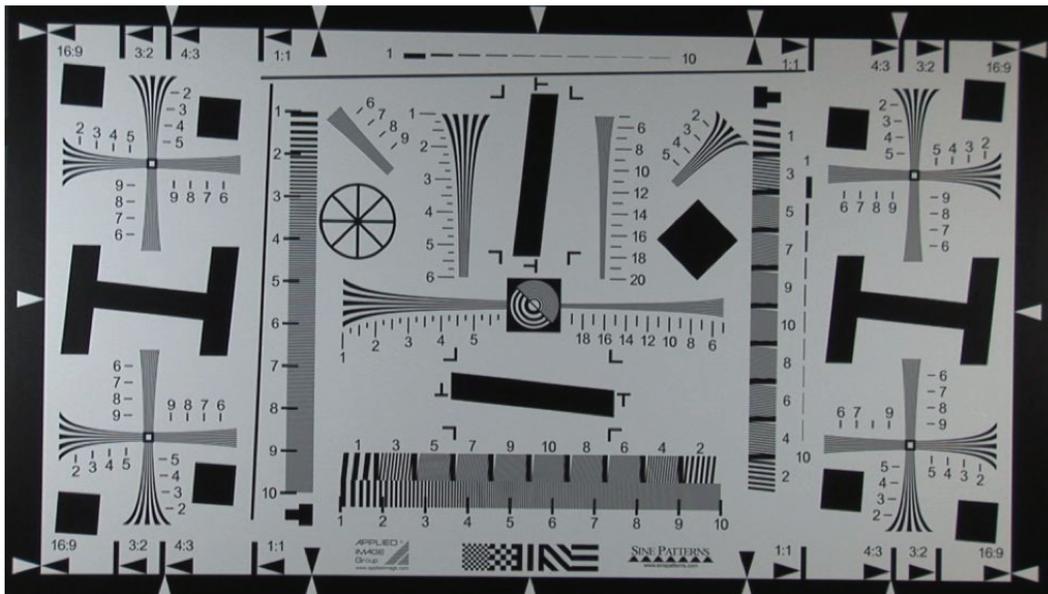
C.3.1.2   Combination Kodak Q-14 and GretagMacbeth ColorChecker Test Chart

Figure 21 is a sample video frame of the combination Kodak Q-14 (top strip) and GretagMacbeth ColorChecker (bottom checkerboard) test chart that was captured using an HD video camcorder. You can use this combination test chart for measuring noise, color accuracy, and dynamic range (indirect method). Mount the two charts on a middle gray surface matte board between 11 by 14 inches and 12 by 16 inches in size. Mount the flimsy Q-14 test chart with adhesive spray to keep it flat. You can mount the more rigid ColorChecker chart by any means. Since you might need to photograph the gray matte board-mounted targets against dark and white backgrounds (e.g., a white background will be required for testing lens flare), back-affix the matte board with "hook and loop" material that allows easy attachment and removal.

Figure 21:   Q-14 and ColorChecker Test Charts Captured Using an HD Video Camcorder
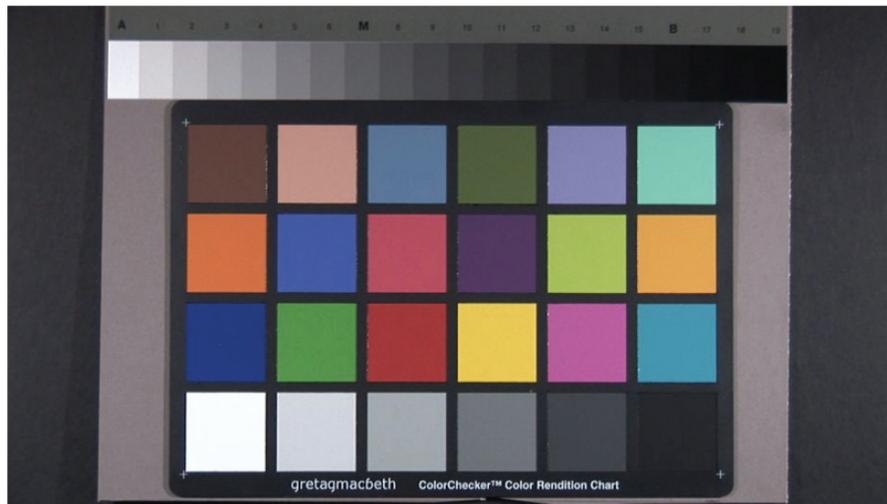
C.3.1.3   Rectilinear Grid Test Chart

Figure 22 presents a simple rectilinear test chart for testing barrel and pincushion distortion of video cameras.

Figure 22:   Rectilinear Grid Test-Chart for Testing Lens Distortion
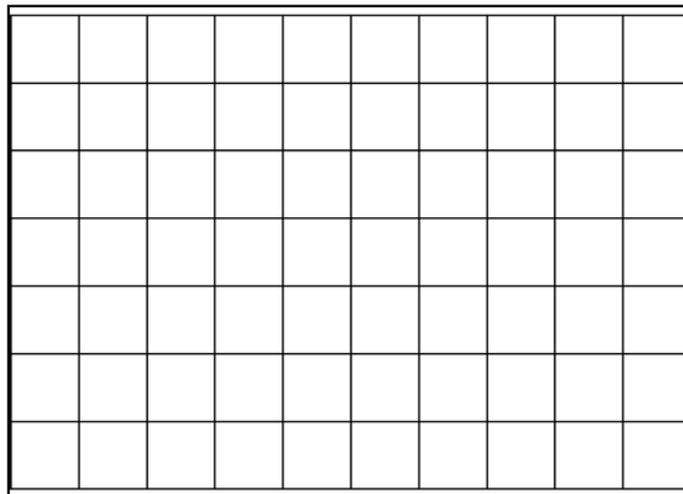
C.3.1.4   Plain White or Gray Background

Use a very evenly lit white or gray background for performing vignetting measurements. A special device, called an integrating sphere, is advantageous for producing uniform lighting. This is especially true for testing wide-angle lenses, where even illumination over a large area may be difficult to achieve.

C.3.2   Lighting Setup for Test Charts

Ensure that the lighting on test charts is uniform and glare-free. To achieve this goal, illuminate reflective test charts by at least two lamps, one on each side of the target, oriented at angles between 30 and 45 degrees, as illustrated in Figure 23. To minimize glare on the test chart, ensure no significant lighting

comes from behind the camera. Check that the optical axis of the camera is perpendicular to the test chart and intersects the center of the test chart (to minimize perspective distortion). Lamp position and angle strongly affect the evenness of illumination across the test chart. To maximize uniformity of the light on the test chart, ensure that the lamps and camera all lie in the same horizontal plane, which also intersects the center of the test chart.

Figure 23:   Lighting Setup for Test Charts
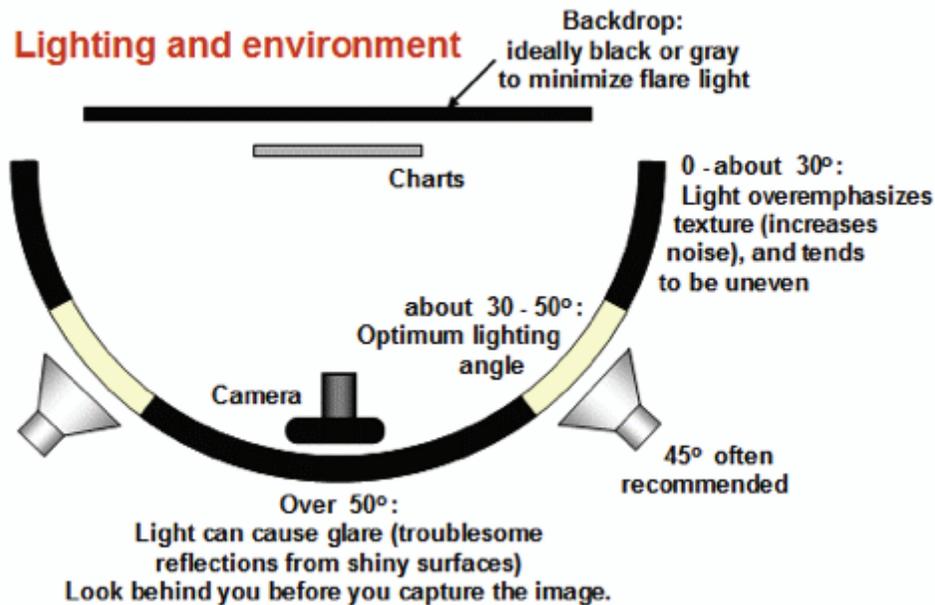
Figure 23 is similar to the default dark room illustration in the SMIA Camera Characterization Specification,[34] which you can use for guidance in setting up the lighting. The SMIA-recommended 45º is not optimum for wide-angle lenses. The angle may need to be reduced to 30º or less to reduce glare near the sides of the test chart, which can be particularly serious in the dark zones of the Kodak Q-14 gray scale step chart, which has a semi-gloss surface.

Uneven lighting on the test chart tends to be less noticeable in the original scene but more obvious in the captured image, so examine the post-exposure images carefully for signs of uneven lighting. If, for example, the gray areas on either side of the ColorChecker (i.e., the background gray matte upon which the ColorChecker is mounted) appear to have the same intensity values, then the lighting is sufficiently uniform from left to right. Use similar examinations to determine top to bottom uniformity of the lighting.

Unless otherwise specified, conduct all performance measurements under standard lighting intensities (see "Standard Lighting Intensity" in Table 57) of approximately daylight color temperature (see "Daylight Light" in Table 57).

C.3.3   Lamps

Many illuminating lamp options are available to fulfill the lighting needs Figure 23 illustrates. Select lamps that have native color temperatures between 4,000K and 7,000K with a color rendering index (CRI) of at least 90 percent. Placing two lamps roughly 1 meter from an 18-inch wide target should, with careful

---

34. Figure 3 "Default Darkroom Set Up." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

adjustment, provide at least 200 lux of even light (no more than ±10 percent variation) on the target. Smaller lamps producing less heat are well-suited for adjusting color temperature using Wratten color correction filters (series 80, 81, 82, or 85) placed in front of the camera lens or the lamp.

Following are a few lamps covering a range of intensity, color temperature, and uniform lighting accuracy.

- **SoLux Task Lamp**. A halogen lamp with a built-in dichroic filter for 4,700K color temperature. Two SoLux lamps at 0.8 to 0.9 meters from the test chart produce approximately 250 lux of incident light.[35]

- **GretagMacbeth Sol-Source Daylight Desk Lamp with Weighted Base**. This is a halogen lamp with a Wratten color-correction filter. You can choose the filter for color temperatures of 5,000K, 6,500K, or 7,500K.[36]

- **North Light Ceramic High Intensity Discharge (HID) Copy Light**. A 4,200K color temperature lamp that is available in different wattage ratings (300, 600, and 900 watts) and useful for achieving even illumination.[37]

- **Dedolight DLH200D Sundance Halogen Metal Iodide (HMI)**. A very high intensity 5,600K color temperature halogen light.[38]

C.3.4 Modifications for Changing Color Temperature and Lighting Intensity

You can modify lamp heads to accept filters for use in reduced and dim light testing (see "Reduced Lighting Intensity" and "Dim Lighting Intensity" in Table 57), color temperature correction, and polarization for glare removal. For illustration purposes, Figure 24 shows the head of the SoLux Task Lamp.

---

35. SoLux Task Lamp. Available from http://www.solux.net/, https://www.nleinternet.net/solux/cgi-bin/tlistore/index.html, and other sources. Cited June 14, 2006.
36. GretagMacbeth Sol-Source daylight desk lamp with weighted base. Available from http://usa.gretagmacbethstore.com/, http://www.colorhq.com, and other sources. Cited June 14, 2006.
37. North Light Ceramic High Intensity Discharge (HID) Copy Light. Available from http://www.northlightproducts.com, http://www.betterlight.com/hid_copylights.html, and other sources. Cited June 14, 2006.
38. Dedolight DLH200D Sundance Halogen Metal Iodide (HMI). Available from http://www.dedolight.com, http://www.bhphotovideo.com, and other sources. Cited June 14, 2006.

Figure 24:   SoLux Task Lamp Head

The modification involves inserting a lens shade that can accept filters over the lamp head. Figure 25, for example, shows a double-threaded rubber lens hood you could use to accept filters.

Figure 25:   Example Lens Shade to Mount Filters

You can use epoxy or cyanoacrylate (Super Glue) to attach the lens shade to the metal rim of the lamp (just outside the bulb). Before attaching the lens shade, ensure there is sufficient clearance so the filters do not contact the lamp head diffuser, and that bulb replacements can occur freely without interference. Due to heat from the lamp, it might be preferable to mount the filters in front of the camera lens.

Use the following filters to adjust the lighting from the SoLux Task Lamp for different color temperatures and lighting intensities. Remember that a mired is 10^6 divided by the color temperature in degrees K.

- 85B warming (yellow) filter. +131 mireds. Changes 4,700K to 2,900K, typical of ordinary incandescent bulbs.

- 80C cooling (blue) filter. –81 mireds. Changes 4,700K to 7,500K, characteristic of cool daylight.

- Neutral Density (ND) filters with D = 0.3 (2x; 1 f-stop), 0.6 (4x; 2 f-stops), and 0.9 (8x; 3 f-stops). Filters can be stacked to obtain densities up to 1.8 (64x; 6 f-stops). For example, if two SoLux Task Lamps located 1 meter from the target provide approximately 250 lux at the target, 0.6 + 0.9 ND filters (for a density of 1.5, which produces a decrease in lighting intensity by a factor of $2^{(1.5/0.3)} = 2^5 = 32$) can reduce the illumination to 250/32 = 7.8 lux, which is in the range of dim light testing. You can make fine adjustments by moving the lamps.

C.4 Methods of Measurement for Performance Parameters

C.4.1 Resolution

Resolution is one of the most important image quality factors; it is closely related to the amount of visible detail in an image. The camera's lens quality, sensor design, signal processing, and especially the application of sharpening or unsharp masking—which can result in "halos" near edges when overdone—all affect resolution.

The traditional method of measuring sharpness uses a resolution test chart. First, you capture an image of a resolution test chart such as the USAF 1951 chart (see Figure 26), which consists of a set of bar patterns. Next, you examine the captured image to determine the finest bar pattern that is discernable as black-and-white (B&W) lines. Finally, you make measurements of the horizontal and vertical resolution by using bars orientated in the vertical and horizontal directions, respectively. Unfortunately, this procedure presents problems because it is manual and its results have a strong dependence on the observer's perception, which can deliver resolution results that correlate poorly with perceived sharpness.

Figure 26:   USAF 1951 Chart
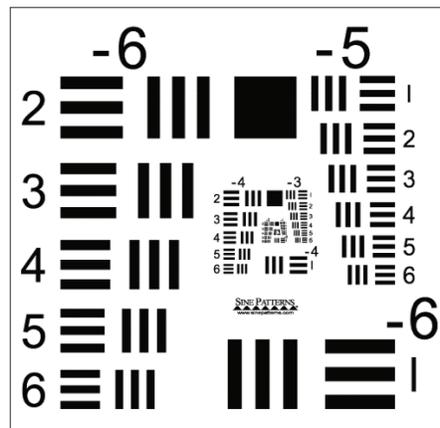
A more contemporary approach is to measure the Modulation Transfer Function (MTF) of the camera system. MTF is the name given by optical engineers to Spatial Frequency Response (SFR). The more extended the MTF response, the sharper the image. The ISO 12233 standard contains a powerful technique for measuring MTF from a simple, slanted-edge target image that is present in the ISO 12233 resolution test chart (see Figure 27).

The International Imaging Industry Association (i3a) offers two free application downloads[39] that implement the ISO standard:

- Slant Edge Analysis Tool sfrwin 1.0 (Windows® executable for most users)
- Slant Edge Analysis Tool sfrmat 2.0 (MATLAB® must be installed)

Both downloads include printable user guides and both provide SFR plots, but little numerical output.

---

39. Slant Edge Analysis tools—sfrmat2 and sfrwin. Available from http://www.i3a.org/downloads_iso_tools.html. Cited June 14, 2006.

Figure 27:   ISO 12233 Resolution Chart
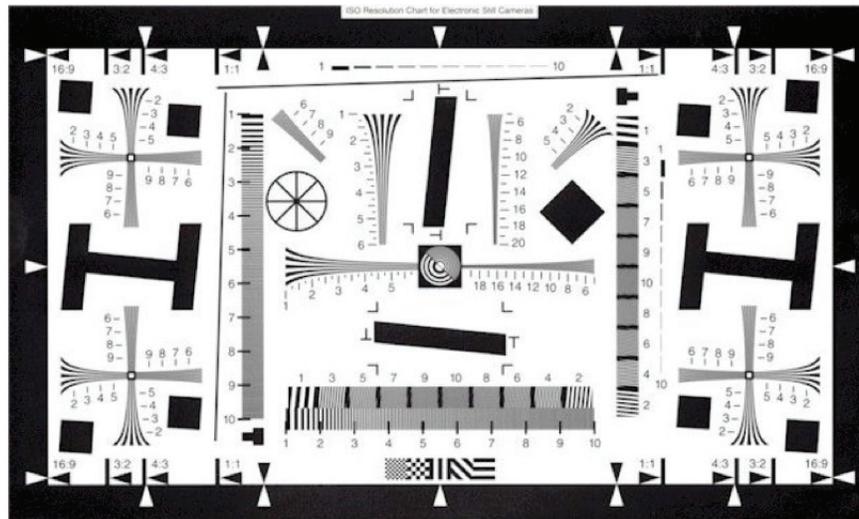
To give accurate results, the sfrmat and sfrwin applications require you to load a tonal response curve, or Opto-Electronic Conversion Function (OECF) file. If the file is omitted, the applications assume gamma = 1, which is atypical of still and video cameras that actually tend to have a capture gamma of around 0.5. Without the proper OECF file, a measurement error of about 10 to 15 percent will result. Since the sfrmat and sfrwin applications do not come with an OECF file for a gamma of 0.5, Section C.5 contains a MATLAB script (makeoecf.m) for creating OECF files.

C.4.1.1   Example Procedure for Measuring Sharpness

The following steps use the sfrwin application as an example to describe the procedure for measuring sharpness.

1.   Download the sfrwin application mentioned in Section C.4.1 for analyzing the slanted-edge pattern in the ISO 12233 resolution chart.

    Extract the sfrwin.zip file into a folder of your choice. (The steps that follow assume the sfrwin application is installed in C:\programs\sfrwin.) Use the makeoecf.m MATLAB program to create an appropriate OECF Look Up Table (LUT) file for the camera system being tested (e.g., a gamma of 0.5 for B&W would produce the OECF file "lut_0.5_1.dat"; a gamma of 0.5 for color would produce the OECF file "lut_0.5_3.dat"). Copy this file into C:\programs\sfrwin\data.

2.   Mount the ISO 12233 test chart on a sheet of foam board (1/2-inch thick preferred) using a spray adhesive to keep it flat. Alternatively, use a test chart consisting of high-quality laser prints of slanted-edges, tilted roughly 5 degrees from horizontal and vertical.

3.   Set up the test chart according to the instructions in Section C.3.2. Frame the test chart within the video picture according to the appropriate aspect ratio markings on the chart (e.g., Figure 20 shows proper test chart framing for an HDTV camera with a 16:9 aspect ratio).

4.   Save a sample video clip from the camera and convert one video frame from this file into a standard still image format. Use TIFF or BMP image formats. You can convert a file to TIFF by opening it with an editor such as Irfanview[40] and saving it as a TIFF file.

---

40. Infran image format conversion tool. Available from http://www.irfanview.com. Cited June 14, 2006.

5.   Run the sfrwin application for slanted vertical and horizontal edges near the center of the image and in the far corner of the image (e.g., one of the edges on the lower-right or upper-left of the ISO 12233 chart in Figure 27). For some cameras, the resolution may vary significantly, depending upon the location in the image (i.e., center vs. edge) and the direction (i.e., horizontal vs. vertical).

Figure 28:   Best Minimum Cropped Region Pixel Dimensions
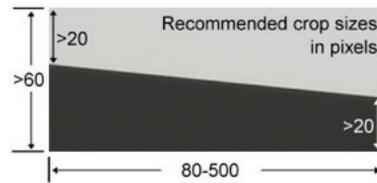
Although the cropped region can be as small as 20 by 20 pixels, ensure the cropped region is at least 60 pixels wide and 80 pixels long to attain the most accurate and consistent results. (Note that the edge is approximately centered in the cropped image.) The horizontal slant edge in Figure 28 is used for measuring the resolution in the vertical direction, while a vertical slant edge (from another part of the ISO 12233 chart) is used for measuring the resolution in the horizontal direction.

In the sfrwin application, leave both **LUT** boxes unchecked for the first run. Leave **Pitch in mm** at **1.0000** to get the output X-axis scaled in cycles per mm. Click **Acquire Image**. Select the input file. Select the region of interest to analyze by clicking and dragging the mouse. In the **Please select the ROI…** window, which might be behind the image window, click **Continue**. Now enter the OECF file name (e.g., lut_0.5_3.dat). Figure 29 shows example MTF results from the sfrwin application for one slant edge (red, green, and blue channels plotted separately).

Figure 29:   Example MTF Results from sfrwin Application
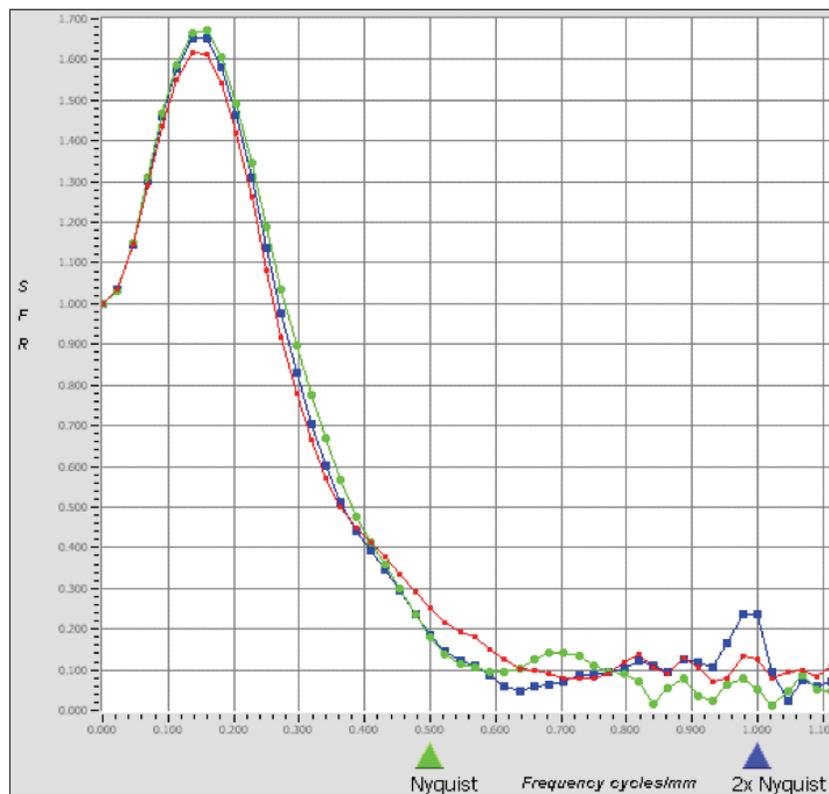
The frequency where MTF drops to 50 percent (MTF50) of its low frequency value is a widely used sharpness metric. But this metric has a serious weakness because it is strongly affected by sharpening applied by software inside the camera. All digital images benefit from some degree of sharpening, but some cameras over-sharpen, resulting in unrealistically high MTF50 values and annoying halo effects near edges.

A better metric for video systems that works in the presence of over-sharpening is MTF50P, the frequency where MTF is half (50 percent) of its peak value. In Figure 29 the peak MTF is 1.65. MTF50P is the spatial frequency where MTF is half that value, 0.82. For this edge, MTF50P = 0.301 cycles per pixel. This example is for horizontal resolution measured using a vertical edge. MTF50P is identical to MTF50 for images that have little or no sharpening, where MTF(0) = MTF(fpeak).

There are several units for measuring MTF50P. While cycles per pixel are produced directly by the sfrwin application, this measures performance on the pixel level. To obtain a measure of the total image resolution, MTF50P is converted into line widths per picture height (LW per PH, where one cycle equals two line widths), using the following equation:

> LW per PH = 2 × cycles per pixel × total pixels

For the example in Figure 29, this would produce a horizontal image resolution value of 2 × 0.301 × 640 (i.e., VGA image), or 385 LW per PH.

C.4.1.2   Algorithm for Calculating MTF

A description follows of the MTF calculation, as derived from ISO 12233 standard slant edges and as implemented by the sfrmat and sfrwin applications. The essential algorithm described here determines the Fourier transform of the impulse response, which is in turn estimated from the derivative of the unit step response:

1. The pixel values in the cropped image are linearized, i.e., the pixel levels are adjusted to remove the transfer curve (also known as the OECF or gamma encoding) applied by the camera.

2. The edge location centers for the Red, Green, Blue, and luminance channels (Y = 0.299 × Red + 0.587 × Green + 0.114 × Blue) are determined for each line (e.g., for measuring resolution in the vertical direction, the vertical lines in the cropped image with a horizontal slant are used). The edge location centers in each line are determined by differencing successive pixel values in the line, and then finding the location of the maximum absolute value.

3. A first- or second-order least-squares fit is calculated for each channel using polynomial regression, where y denotes the edge location centers (from step 2), and x represents the associated pixel locations of each line. For the cropped image, the second-order equation would have the form, $y = a_0 + a_1 x + a_2 x^2$. The a_i coefficients can be found using the MATLAB polyfit function; the fitted y can be determined using the MATLAB polyval function. The fitted y provides an improved estimate for the true edge location centers. A second-order least-squares fit may be required when lens distortion creates a curved rather than straight slant edge.

4. Depending on the value of the fractional part $fp = y_i - int(y_i)$ of the second-order least-squares fit for each line, four average lines are produced, one line for each of the following: $0 \le fp < 0.2\dot{5}$, $0 \le fp < 0.\dot{5}$, $0.5 \le fp < 0.7\dot{5}$, and $0.75 \le fp < 1$. The averaging process centers the edge locations of each line within the averaging buffers. Each of the four average lines form an estimate of the unit step response, each shifted by ¼ pixels.

5.   The four average lines from step 4 are interleaved to produce a 4x oversampled line. This allows analysis of spatial frequencies beyond the normal Nyquist frequency.

6.   The derivative (*d/dx*) of the averaged 4x oversampled edge is calculated by differencing adjacent pixels. A Hamming windowing function is applied to force the derivative to zero at the endpoints.

7.   MTF is the absolute value of the fast Fourier transform (FFT) of the windowed derivative from step 6.

C.4.2   Noise

Noise is the unwanted random spatial and temporal variations (e.g., snow) in the video picture. It has a strong effect on a camera's dynamic range.

One method of measuring noise is to capture and analyze images of a step chart consisting of patches of uniform density, such as the Kodak Q-14 Gray Scale (Figure 21, top). The Q-14 Gray Scale consists of 20 patches with densities from 0.05 to 1.95 in steps of 0.1. Noise and signal-to-noise ratio (SNR) can be measured for each patch. SNR tends to be worst in the darkest patches and for dim lighting. Several lighting conditions with various intensities (e.g., standard, reduced, dim) and color temperatures (e.g., tungsten, daylight) may be required to adequately characterize noise and SNR.

Follow these steps to measure noise and SNR within a patch:

1.   Select a rectangular region that contains most of the patch. The edges of the selected region should be far enough from the patch boundaries to eliminate edge effects. The selected region typically comprises 50 to 70 percent of the total patch area. The pixel values will be represented by $P(x,y)$, where $1 \le x \le m$ and $1 \le y \le n$. The mean pixel level of the region is:

$$\text{mean}(P) = \frac{1}{mn} \sum_{x=1}^{m} \sum_{y=1}^{n} P(x, y)$$

2.   A useful approximation of the noise in the region is the standard deviation σ of:

$$P{:}N_P = \sigma(P) = \left(\Sigma (P(x, y) - \text{mean}(P))^2 / (mn - 1) \right)^{1/2}$$

However, lighting nonuniformity reduces the accuracy of the simple standard deviation in many practical situations. To obtain a good noise measurement, the signal variation due to lighting nonuniformity must be removed, as the following procedure describes:

a. Find the horizontal and vertical mean values of the signal.

$$P_{Ymean}(x) = \frac{1}{n} \sum_{y=1}^{n} P(x, y) \qquad P_{Xmean}(y) = \frac{1}{m} \sum_{x=1}^{m} P(x, y)$$

b. Find the second-order polynomial fits to these means.

$$F_Y(x) = f_{y1}x^2 + f_{y2}x + f_{y3} \qquad F_X(y) = f_{x1}y^2 + f_{x2}y + f_{x3}$$

The f_{xi} and f_{yi} coefficients can be found using the MATLAB polyfit function; the fitted $F_Y(x)$ and $F_X(y)$ can be determined using the MATLAB polyval function. These values represent the slowly varying illumination within the patch.

   c. Subtract the nonuniformity terms of F_Y and F_X from $P(x,y)$ to obtain the uniformly illuminated signal:

$$P_U(x, y) = P(x, y) - f_{y1}x^2 - f_{y2}x - f_{x1}y^2 - f_{x2}y$$

   d. Pixel noise is the standard deviation of P_U for the region $(1 \leq x \leq m, 1 \leq y \leq n)$:

$$N_P = \sigma(P_U) = (\Sigma(P_U(x, y) - \mathrm{mean}(P_U))^2 / (mn - 1))^{1/2}$$

   e. Note that the constant terms, f_{y3} and f_{x3}, have no effect on N_P; using the equation $P_U(x,y) = P(x,y) - F_Y(x) - F_X(y)$ instead of the equation in step c, results in the same value of N_P.

3.   The *pixel* SNR (P/N_P) for the region is equal to $\mathrm{mean}(P_U)/N_P$.

4.   In imaging literature, S/N often refers to the scene-referenced or sensor SNR, S/N_S, prior to the conversion to an image file. The conversion is characterized by a transfer function called the OECF (Opto-Electronic Conversion Function), which is represented as a table with pixel level P as the independent variable and Luminance (linearized response) L as the dependent variable. Figure 30 shows an OECF curve for camera gamma = 0.5.

<p align="center">Figure 30:   OECF Table Plot for Camera Gamma</p>
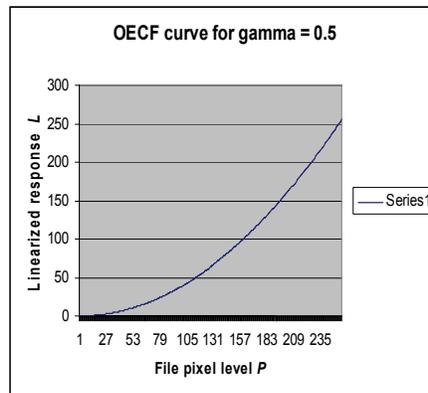
5.   The OECF can be calculated from the image of the Q-14 chart using the knowledge that the chart has density steps of 0.1, where *density* = -$\log_{10}$(exposure).

6.   The OECF is often approximated as an exponential function, though in practice an "S" curve is frequently superimposed on top of the exponential. The exponential transformation from the sensor to the image file is called gamma encoding; it is the inverse function of the OECF, since luminance is transformed to pixel level (see Figure 31). The equation for gamma encoding is Pixel level = $P = L^{\Delta}$, where L is luminance. Camera gamma γ is typically around 0.5 for standard image files[41] designed for display gamma = 2.2.

NOTE: Camera/Capture Gamma Nomenclature

Display (i.e., monitor) gamma is always described by the equation, $L = P^{\gamma}$. But camera (or capture) gamma can be defined in either of two ways: 1) It can be defined under the assumption that output = input$^{\gamma}$, in which case, $P = L^{\gamma}$; or 2) it can be defined under the assumption that

41. "Gamma FAQ—Frequently Asked Questions about Gamma." Charles Poynton. http://www.poynton.com/notes/colour_and_gamma/GammaFAQ.html. Cited June 14, 2006.

L = P$^\gamma$ for both the input and the output, in which case, P = L$^{\frac{1}{\gamma}}$. The former assumption is used in standard film response curves. The latter assumption appears in some imaging literature, for example, in Charles Poynton's well-known Gamma FAQ.[42]

In this document we use the first formula, P = L$^\gamma$. With this nomenclature, camera and display gamma have the same units, so that total system gamma is the product of the camera and display gamma.

7.   Gamma (γ) is a measure of perceived image contrast. It can be determined by plotting $\log_{10}(P)$ as a function of *density*, ($-\log_{10}$(exposure)). γ is the average slope of the relatively linear region of the plot, i.e., where the slope is at least 20 percent of its maximum value. This requirement ensures that a relatively linear portion of the response curve is used. Portions of the image where the slope is lower, typically located in the toe and knee (deep shadow and extreme highlight regions) of the response curve, contribute little to the pictorial content of the image. Strobel, Compton, Current, and Zakia provide justification for this criterion.[43] Gamma can be measured at the same time as noise using the method described in Section C.4.5.

8.   The scene or luminance noise, scaled according to Figure 31 (the inverse of the OECF chart), is

$$N_S = N_P \frac{dL}{dP}$$

where *dL/dP* is the derivative of the OECF.

Figure 31:   Scaled Luminance Noise
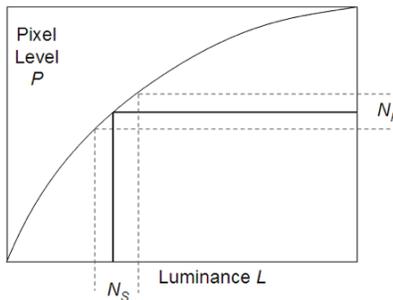
9.   The scene-referenced SNR is:

$$S/N_S = \frac{L}{N_S} = \frac{L}{N_P \frac{dL}{dP}}$$

10.   For an OECF that is approximated by the inverse function of the gamma correction curve, P = L$^\gamma$, and

$$\frac{dP}{dL} = \gamma L^{\gamma-1} \qquad \frac{dL}{dP} = \frac{P^{1/\gamma-1}}{\gamma}$$

The scene-referenced SNR is approximated by:

---

42. For NTSC video systems, camera gamma is equal to 0.45.
43. Strobel, Compton, Current, and Zakia. 2000, *Basic Photographic Materials and Processes*, second edition., Focal Press, Massachusetts, p. 101 (the section on Log Exposure Range in the chapter on Photographic Sensitometry).

$$S/N_S = \frac{L}{N_S} = \frac{\gamma P^{1/\gamma}}{N_P P^{1/\gamma-1}} = \frac{\gamma P}{N_P}$$

where γ is the factor that converts pixel SNR, which is easy to measure, into scene SNR. This approximation holds true only when the OECF resembles an exponential curve.

These equations provide the basis for measuring noise and signal-to-noise ratio (SNR) in individual patches of any of several test charts. It is possible to specify maximum values of noise, or minimum values of SNR, for one or more patches in a chart. An example is patches 2 (light gray) and 10 (dark gray) of the Kodak Q-14 Gray Scale could be used. Noise is generally invisible in white areas, and difficult to see in dark areas (although SNR can be poor in dark areas). Noise tends to be worse in dim light, where amplifier gain in video cameras has to be boosted to recover the signal.

C.4.3 Dynamic Range

Dynamic range (DR) is an important video acquisition performance specification in many public safety applications, especially where lighting is poorly controlled or where video images contain multiple objects under vastly different lighting conditions. An example is nighttime objects illuminated by a spotlight together with objects not illuminated by the spotlight. The measurement of DR in this section is for instantaneous DR, in that the camera's aperture and shutter speed are assumed to be fixed for the duration of the measurement. This is different than tunable dynamic range, where the camera aperture can be opened and closed over time. Instantaneous DR is a measure of the total range of unique luminance levels that can be output by the camera in any given video frame. A camera's effective dynamic range depends primarily on two factors:

- Intrinsic dynamic range of the camera's image sensor—or the range of unique luminance levels that can be captured by the sensor. In video cameras, where the frame rate does not allow long exposures and where low light performance is achieved by increasing the amplifier gain, versus opening up the lens aperture, effective dynamic range will be limited by reduced SNR.

- Flare light—also called "veiling glare." Light that bounces between lens elements and off the interior barrel of the lens can limit the effective dynamic range by fogging shadows and causing ghost images in the proximity to bright light sources.

DR is usually measured in f-stops (factors of two in luminance), but it can also be measured in exposure density units, where one density unit = 3.32 f-stops. You can measure DR by photographing a transmission or reflection step chart consisting of patches with a wide range of densities. Most step charts have uniform density steps of 0.1 or 0.15 (1/3 or 1/2 f-stop). The logarithm to the base 10 of the pixel level ($\log_{10}(P)$) and the scene-referred SNR is calculated for each patch. The camera's dynamic range is then defined as the range of step chart densities (or equivalently, f-stops) where the following criteria are met:

1. The difference in $\log_{10}(P)$ between patches for charts with uniform density steps (or $\Delta(\log_{10}(P))/\Delta(\text{density})$ for charts with non-uniform density steps) is greater than a specified fraction (typically 0.2 to 0.3) of the maximum difference. The difference refers to the maximum difference observed over all the steps. This difference is called the contrast step.

2. The scene-referenced SNR (see Section C.4.2) is greater than a specified level, typically 1, which corresponds to the intent of the ISO 15739 specification, which defines the ISO digital still camera (DSC) dynamic range measurement. The higher the specified level of the scene-referenced SNR, the smaller the resulting dynamic range. This dynamic range will still have a higher effective SNR.

Significant differences exist between DR measurements of still and video cameras. Still cameras, especially digital SLRs with large pixel sizes, often have extremely large dynamic ranges, 10 or more f-stops, which can be realized via post-processing of raw sensor files. This is more than can be easily displayed in prints, so a certain amount of post-processing image manipulation is required to make the full dynamic range useful (e.g., to bring out information hidden in the shadows). On the other hand, users can only access processed sensor information from video cameras that have much less dynamic range.

Because still cameras can have such large dynamic ranges, their DR is best tested using transmission step charts (e.g., Figure 32) such as the Stouffer T4110, which has an exposure density range of 4.0. Measuring DR with a transmission chart takes considerably more care and effort: the chart must be evenly illuminated from behind and photographed in total darkness. Stray room light must be avoided.

Figure 32:   Example Transmission Step Chart Image
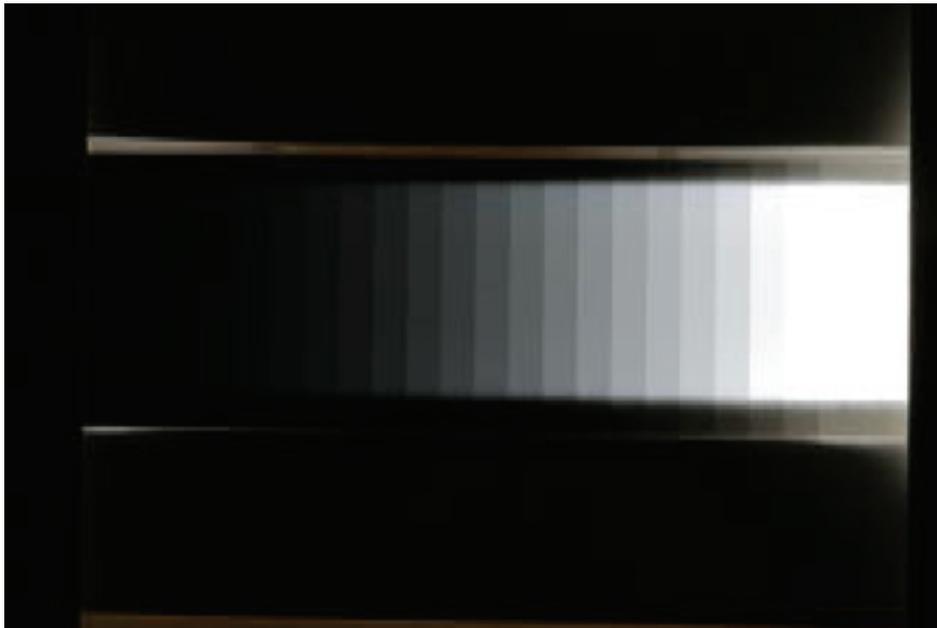
On the other hand, you can photograph the Kodak Q-13 or Q-14 reflection step chart (top strip chart in Figure 21) using the standard lighting setup described in Section C.3.2. But its exposure density range is only 1.9, which is equivalent to $1.9 \times 3.32 = 6.3$ f-stops. This is well below the DR of many digital still and video cameras, but it may be sufficient for specifying whether a video camera has sufficient DR for public safety requirements. You can measure a camera's DR using a chart with a DR less than that of the camera under test by specifying both criteria 1 and 2 described above (i.e., the minimum value of $\Delta(\log_{10}(P))$ and the minimum SNR) in such a manner so as to ensure that the camera has excellent performance within the 6.3 f-stop range of the reflective chart (with high SNR) as well as acceptable performance beyond the 6.3 f-stops (with reduced SNR).

In summary, a camera's dynamic range can be measured by one of two methods:

- **Direct Method.** Uses a transmission step chart with a density range that equals or exceeds the camera's DR. Direct measurements are more difficult to perform than indirect measurements, but they are more accurate and can be used as checks on indirect measurements.

- **Indirect Method.** Uses a reflection test chart, such as the Kodak Q-13 or Q-14, whose DR may be less than that of the camera under test. Rather than estimating the camera's total DR, minimum

acceptable values are set for both the contrast step $(\Delta(\log_{10}(P))/\Delta(density))$ and the minimum SNR. This ensures that the camera's effective DR exceeds the density range of the reflective chart by an acceptable margin. The indirect method is much more convenient than direct method.

C.4.3.1   DR Direct Method

Table 58 lists several transmission step charts, all of which have a density range of at least 3 (10 f-stops). Kodak and Stouffer Photographic Step Tablets can be purchased calibrated or uncalibrated. Calibrated charts, which have individual density measurements for each patch, offer an assurance of quality but little practical improvement in accuracy.

Table 58:   Transmission Step Charts for Measuring Dynamic Range with the Direct Method

Product	Steps	Density Increment	Dmax	Size
Kodak Photographic Step Tablet No. 2 or 3	21	0.15 (1/2 f-stop)	3.05	1 by 5.5 inches (#2) larger (#3)
Stouffer Transmission Step Wedge T2115	21	0.15 (1/2 f-stop)	3.05	0. 5 by 5 inches
Stouffer Transmission Step Wedge T3110	31	0.10 (1/3 f-stop)	3.05	3/4 by 8 inches
Stouffer Transmission Step Wedge T4110	41	0.10 (1/3 f-stop)	4.05	1 by 9 inches
Danes-Picta TS28D (on their Digital Imaging page)	28	0.15 (1/2 f-stop)	4.2	10 by 230 mm (0.49 inches)

Follow these steps to manually measure DR using the direct method:

1.   Prepare a fixture for mounting the transmission step chart. Ensure it is large enough to keep stray light out of the camera.

     Stray light can reduce the measured dynamic range; avoid it at all costs. You can make fixtures from simple materials such as scrap matte board.

2.   Place the fixture with the step chart on top of a light box or any other source of uniform diffuse light. Standard light boxes are fine. If some non-uniformity is visible in the light box, orient the chart to minimize its effects; that is, if there is a linear fluorescent lamp behind the diffuser, place the chart above the lamp, along its length.

3.   Photograph the step chart in a darkened room. Ensure no stray light reaches the front of the target, as this will distort the results. Keep the surroundings of the chart relatively dark to minimize flare light, as Figure 32 shows. The density difference between the darker zones is not very visible in the figure, but it shows up clearly in the measurements.

     ☐ If possible, set camera exposure manually. The indirect method, which Section C.4.3.2 describes, is more suitable for cameras that cannot be set manually because a reflection chart can easily be surrounded with a neutral (approximately 18 percent reflectance) gray background to influence the auto-exposure setting. If your camera displays a histogram, use it to determine the exposure that just saturates the lightest region of the chart. Overexposure (or underexposure) will reduce the measured dynamic range. The lightest region should have a relative pixel level of at least 0.98 (pixel level 250 of 255). Otherwise, the full dynamic range of the camera will not be measured.

     ☐ You can photograph the chart slightly out of focus to minimize noise measurement errors due to

texture in the test chart patches. We emphasize the word "slightly" because the boundaries between the patches must remain distinct.

☐ The distance to the test chart is not overly critical. For an accurate noise analysis, ensure the chart fills most of the image width for cameras with VGA (640 pixels wide) or lower resolution. Increasing the size improves the accuracy of the noise measurement, although in some cases it might increase light falloff (vignetting), which can affect the accuracy of the measurement.

☐ Capture the image from the camera in the highest quality format. If the camera employs data compression, use the highest quality (lowest compression) setting.

4. Determine the mean pixel level and scene-referenced SNR of each patch in the chart image. (These are defined in Section C.4.2.)

5. Visualize the results by plotting the logarithm of the normalized mean pixel level (e.g., $\log_{10}(\text{mean}(P)/255)$ for systems with 8-bits per color) against $\log_{10}(\text{exposure})$. This can be derived from the known density steps of the chart, most often 0.10 or 0.15. $\log_{10}(\text{exposure}) = -density + k$, where k is an arbitrary constant. This is a standard plot that is similar to traditional characteristic curves for film.

6. The dynamic range is the range of densities, or the density step multiplied by the number of steps, where 1) the contrast step $(\Delta(\log_{10}(\text{mean}(P)/255))/\Delta(\text{density}))$ is larger than 0.2 of the maximum contrast step; and 2) the scene referenced SNR (S/N_S, defined in Section C.4.2) is larger than a specified minimum level, typically 1 or larger. If you choose a scene referenced SNR level other than 1, include this level with the DR specification. Convert dynamic range in density to f-stops by multiplying by 3.32.

The following steps use the Imatest application[44] as an example to illustrate the direct method of measurement for DR:

1. Download and install the Imatest application.

2. Start the Imatest application, and click the **Stepchart** button in the main Imatest window.

3. Open the input image file.

4. Crop the image to minimize edge effects.

    The red rectangle in Figure 33 shows a typical crop.

Figure 33:   Example Crop of a Stouffer T4110 Chart

5. Make any necessary changes in the step chart input window (see Figure 34).

    The default selection is a reflective target with density steps of 0.10 (i.e., the Kodak Q-13 or Q-14). If you are using a transmission target (see Table 58), choose the correct target type from the drop-down list.

6. Click **OK** to continue.

    Figure 35 shows the strip chart image of Figure 33 after step chart processing.

---

44. Imatest camera, lens, scanner, and printer image performance measurement tool. Available from *http://www.imatest.com*. Cited June 14, 2006.

Figure 34:   Example Step Chart Input Selection
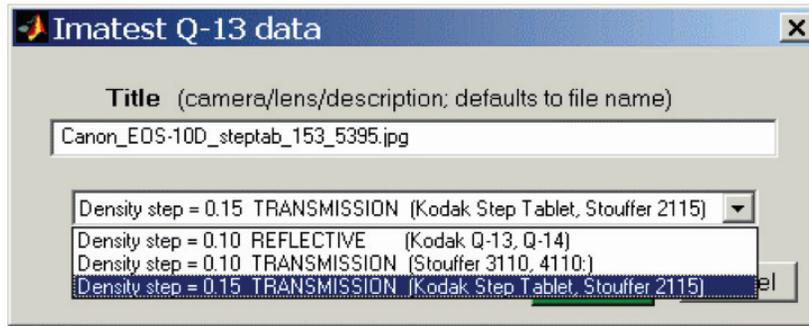
Figure 35:   Strip Chart Image of Figure 33 After Step Chart Processing

Imatest detects the chart zones using the smallest density step that results in uniformly spaced detected zones. For smaller steps, noise can be mistaken for zone boundaries. For larger steps, fewer zones are detected. The dynamic range is the difference in density between the zone where the pixel level is 98 percent of its maximum value (250 for 8-bits per color, where the maximum is 255), estimated by interpolation, and the darkest zone that meets the measurement criterion in step 6 of the preceding list of steps for manually measuring DR using the direct method. Figure 36 presents example DR results from the Imatest application. The measured DR is 8.34 f-stops.

Figure 36:   Example DR Measurement Results
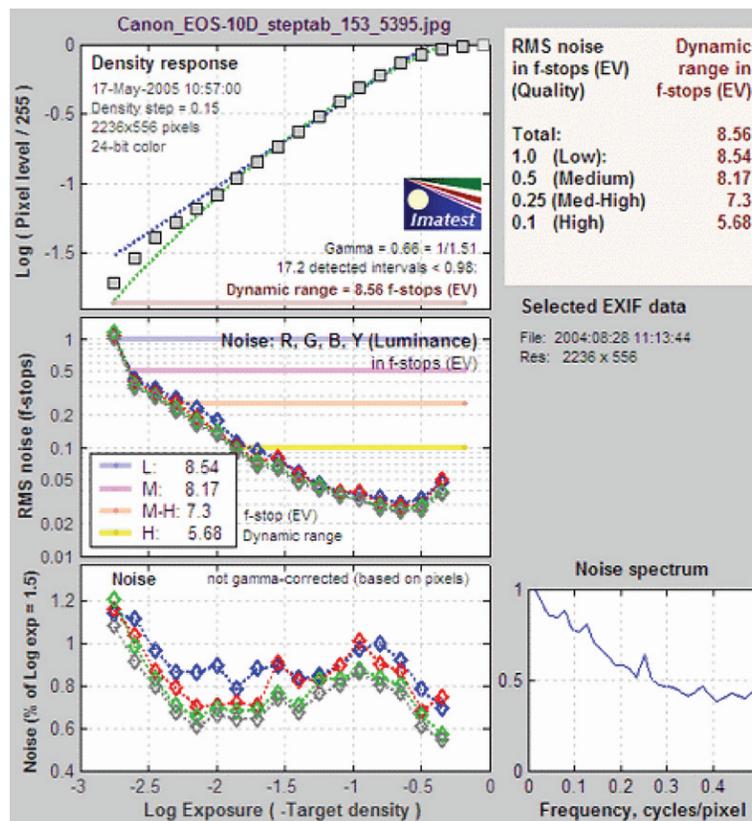
C.4.3.2   DR Indirect Method

The indirect dynamic range measurement is easier to perform than the direct measurement because it takes advantage of the same lighting setup used in the sharpness and color measurements (see Section C.3.2). It is based on a minimum detectable contrast step with a specified SNR in an image of a reflective step chart with a density range of 1.9: somewhat less than the expected total dynamic range, but very practical nonetheless.

Some of the following steps for the indirect dynamic range measurement are identical to the direct method in Section C.4.3.1.

1.   Photograph the Q-14 (or similar) reflective step chart, mounted as described in Section C.3.1.2, and light as described in Section C.3.2. Check the image carefully to make sure there is no glare or reflections on the target, which would ruin the measurements.

☐ You can photograph the chart slightly out of focus to minimize noise measurement errors due to texture in the test chart patches. We emphasize the word "slightly" because the boundaries between the patches must remain distinct.

☐ The distance to the test chart is not overly critical. For an accurate noise analysis, ensure the chart fills most of the image width for cameras with VGA (640 pixels wide) or lower resolution. Increasing the size improves the accuracy of the noise measurement, although in some cases it may increase light falloff (vignetting), which may affect the accuracy of the measurement.

☐ Capture the image from the camera in the highest quality format. If the camera employs data compression, use the highest quality (lowest compression) setting.

2.   Determine the logarithm of the normalized mean pixel level (e.g., $\log_{10}(\text{mean}(P)/255)$ for 8-bit systems) and scene-referenced SNR (S/N_S) of each patch in the chart image. (Section C.4.2 describes this process.)

3.   Visualize the results by plotting the logarithm of the normalized mean pixel level against $\log_{10}(\text{exposure})$, which can be derived from the known density steps of the chart, typically 0.10 or 0.15, using the equation, $\log_{10}(\text{exposure}) = -density + k$, where *density* is the patch density and *k* is an arbitrary constant. This is a standard plot that is similar to traditional characteristic curves for film.

4.   The dynamic range is the range of densities (the density step times the number of steps) where: 1) the contrast step $(\Delta(\log_{10}(\text{mean}(P)/255))/\Delta(\text{density}))$ is larger than 0.2 of the maximum contrast step; and 2) the scene referenced SNR (Section C.4.2 defines S/N_S) is larger than a specified minimum level, typically 1 or larger. If you choose a scene-referenced SNR level other than 1, include this level must with the DR specification. Choosing a scene referenced SNR level that is greater than one for this indirect DR measurement will allow a higher effective DR to be specified, provided all patches still fall within the criteria. Convert dynamic range in density to f-stops by multiplying by 3.32.

C.4.4   Color Accuracy

Color accuracy is dependent on a camera's sensor quality and signal processing, particularly its white balance (WB) algorithm. Measure color accuracy under both daylight and tungsten lighting, as Section C.2 describes.

Measure color accuracy by photographing the GretagMacbeth ColorChecker (see Section C.3.1.2), the widely used standard color chart consisting of 24 patches: 18 color and 6 grayscale. Using the color difference equations in the sections that follow, analyze the individual color patches for color error. These color difference equations are from the *Digital Color Imaging Handbook*.[45]

The ideal background for photographing the color chart is gray matte board of approximately 18 percent reflectance (*density* = 0.745): the reflectance of a standard gray card. This corresponds to zone 7 (M) on the Kodak Q-13 or Q-14 gray scale and to patch 22 (bottom row, fourth from the left) on the GretagMacbeth ColorChecker. The color and reflectance of the gray background does not have to be very accurate, as its only purpose is to influence the camera's automatic exposure and white balance.

C.4.4.1   Color Accuracy Measurement

Follow these steps to manually measure color accuracy:

1.   You can make the measurement for any specified combination of lighting intensity (standard, reduced, or dim) and color temperature (tungsten, or daylight) as Section C.2 specifies. Ensure you associate the lighting intensity and color temperature that was used with any measured values. Adjust the lighting and GretagMacbeth ColorChecker chart as Section Section C.3 specifies, and capture one video image from the GretagMacbeth ColorChecker chart.

2.   Measure the average color values for each patch in the ColorChecker chart, excluding areas near the boundaries. If the values are Red Green Blue (RGB), go to step 4 below. If they are YC_BC_R (common for many video cameras), use the equation in step 3, below, to convert to RGB.

---

45. Gaurav Sharma, 2003, *Digital Color Imaging Handbook*, CRC Press, Florida.

3. The conversion equation from YC_BC_R to RGB (scaled for maximum values of 255)[46] is

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 298.1 & 0 & 408.6 \\ 298.1 & -100.3 & -208.1 \\ 298.1 & 516.4 & 0 \end{bmatrix} \bullet \left(\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

RGB values from this equation that fall outside the range [0, 255] should be clipped at 0 and 255.

4. Convert the RGB color values into L*a*b* color values, using the equations in Section C.4.4.2.

5. The standard measurements of color error (or color difference) between colors 1 and 2 are ΔE^*_{ab} (which includes both color and luminance) and ΔC^*_{ab} (color only):

$$\Delta E^*_{ab} = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \qquad \text{chroma and luminance}$$

$$\Delta C^*_{ab} = \sqrt{(a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \qquad \text{chroma only}$$

ΔC^*_{ab} and ΔE^*_{ab} are the Euclidian distances in the CIELAB L*a*b* color space between the reference values from the table in Section C.4.4.2 and the measured sample values.

6. Alternatively, if greater accuracy is required, you can use the more accurate but less familiar CIE 1994 color difference formulas, ΔE^*_{94} and ΔC^*_{94}. These equations account for the eye's reduced sensitivity to chroma differences for highly saturated colors. In the equations that follow, subscript 1 represents the reference values from the table in Section C.4.4.2, and subscript 2 represents the measured sample values:

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L}{K_L S_L}\right)^2 + \left(\frac{\Delta C}{K_C S_C}\right)^2 + \left(\frac{\Delta H}{K_H S_H}\right)^2} \qquad \text{chroma and luminance}$$

$$\Delta C^*_{94} = \sqrt{\left(\frac{\Delta C}{K_C S_C}\right)^2 + \left(\frac{\Delta H}{K_H S_H}\right)^2} \qquad \text{chroma only}$$

$$\Delta L = L_1 - L_2 \qquad \Delta C = C_1 - C_2 \qquad \Delta a = a_1 - a_2 \qquad \Delta b = b_1 - b_2$$

$$C_1 = \sqrt{\Delta a_1^2 + \Delta b_2^2} \qquad C_2 = \sqrt{\Delta a_2^2 + \Delta b_2^2} \qquad \Delta H = \sqrt{\Delta a^2 + \Delta b^2 + \Delta C^2}$$

$$S_L = K_L = K_C = K_H = 1$$

$$S_C = 1 + 0.045 C_1 \qquad S_H = 1 + 0.015 C_1$$

ΔE^*_{94} and ΔC^*_{94} result in lower numbers than $\Delta E^*_{ab} E^*_{ab}$ and ΔC^*_{ab}, especially when strongly saturated colors (large values of C_1 and C_2), are compared.

7. For purposes of determining an overall measurement of color accuracy, the $\text{mean}(\Delta C^*_{ab})$ or $\text{mean}(\Delta C^*_{94})$ is computed over all 24 patches of the ColorChecker chart. ΔC^* is preferred over ΔE^* because it excludes luminance (exposure) error, which is dealt with separately in Section C.4.6.

Figure 37 shows example color accuracy measurement results as output by the Imatest application. The axis in this plot (i.e., a* and b*) are defined in step 4 above.

---

46. See "Color FAQ—Frequently Asked Questions about Color." Charles Poynton.
http://www.poynton.com/notes/colour_and_gamma/ColorFAQ.html. Cited June 14, 2006.

Figure 37:   Example Color Accuracy Measurement Results
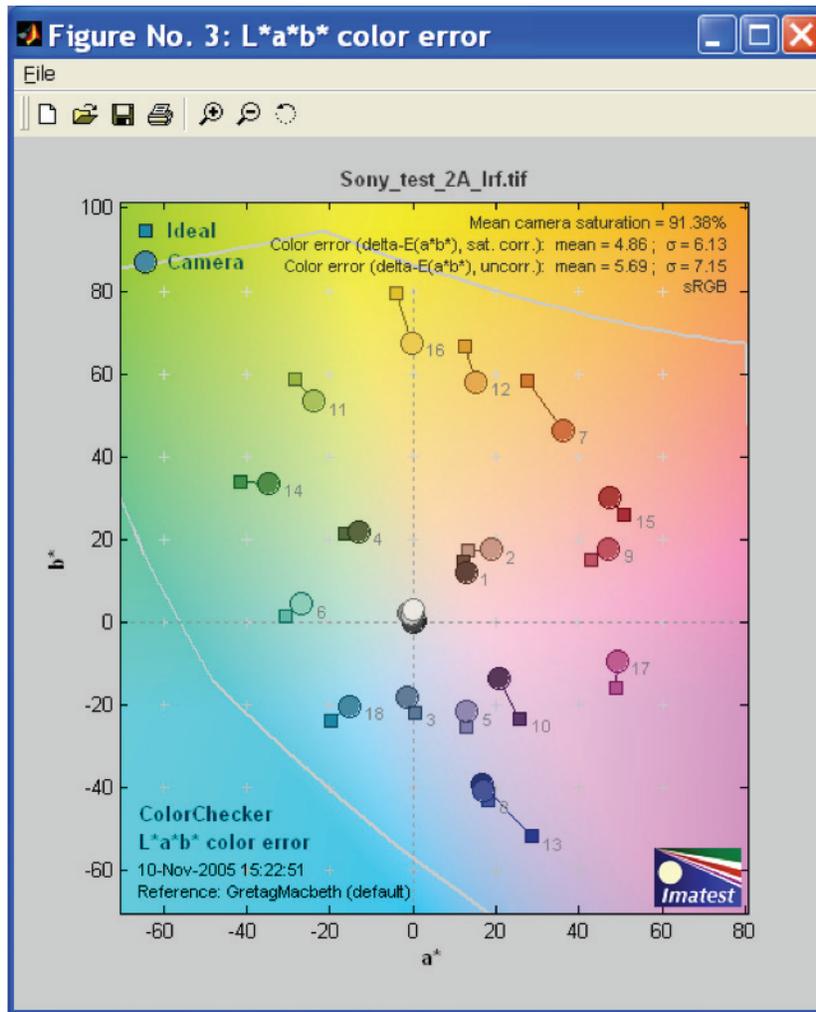
C.4.4.2   Converting RGB values to L*a*b*

To obtain ΔE^*_{ab} and related color difference values, it is necessary to convert the system-dependent RGB values into La*b* values. This is a two-step process: 1) Convert RGB into XYZ; 2) Convert XYZ to L*a*b*. The following equations and values are from brucelindbloom.com:[47]

1.   If the RGB values are in the range [0, 255], divide their values by 255. Given an RGB color whose components are in the nominal range [0.0, 1.0], compute:

$$[\, X\, Y\, Z\,] = [\, r\, g\, b\,]\, [\, M\,]$$

where, if the RGB system is **not** sRGB:

$$r = R^{\gamma}\,;\quad g = G^{\gamma}\,;\quad b = B^{\gamma}$$

and if it **is** sRGB:

---

47. "Useful Color Equations." Click "Math." Bruce Lindbloom. http://www.brucelindbloom.com. Cited June 14, 2006.

$$r = R/12.92 \qquad\qquad R \le 0.04045$$
$$ = ((R + 0.055)/1.055)^{2.4} \qquad R > 0.04045$$
$$g = G/12.92 \qquad\qquad G \le 0.04045$$
$$ = ((G + 0.055)/1.055)^{2.4} \qquad G > 0.04045$$
$$b = B/12.92 \qquad\qquad B \le 0.04045$$
$$ = ((B + 0.055)/1.055)^{2.4} \qquad B > 0.04045$$

sRGB is approximately (but not exactly) gamma $\gamma = 2.2$. Most video color spaces use gamma $\gamma = 2.2$. (See the Info section of brucelindbloom.com for the correct gamma γ values of various RGB color spaces.) For sRGB, the matrix [M] is:

$$[M] = \begin{bmatrix} 0.412424 & 0.212656 & 0.0193324 \\ 0.357579 & 0.717178 & 0.119193 \\ 0.180464 & 0.0721856 & 0.950444 \end{bmatrix}$$

(The Math section of brucelindbloom.com provides the matrix [M] for other RGB working spaces.)

2.   Convert XYZ from step 1 to L*a*b*. This conversion requires a reference white X_r, Y_r, Z_r. Since most color spaces in video cameras have a D65 (6,500K) white point, $X_r = 0.9505$, $Y_r = 1.0$, $Z_r = 1.0890$ are recommended. (Use $X_r = 0.9642$, $Y_r = 1.0$, $Z_r = 0.8252$ for color spaces that use a D50, or 5,000K illuminant.)

$$L* = 116\, f_y - 16\,; \quad a* = 500\,(f_x - f_y)\,; \quad b* = 200\,(f_y - f_z)\,, \text{ where:}$$

$$f_x = \sqrt[3]{x_r} \qquad\qquad x_r > \varepsilon \qquad \varepsilon = 0.008856$$
$$ = \frac{\kappa x_r + 16}{116} \qquad x_r \le \varepsilon \qquad \kappa = 903.3$$
$$f_y = \sqrt[3]{y_r} \qquad\qquad y_r > \varepsilon$$
$$ = \frac{\kappa y_r + 16}{116} \qquad y_r \le \varepsilon$$
$$f_z = \sqrt[3]{z_r} \qquad\qquad z_r > \varepsilon$$
$$ = \frac{\kappa z_r + 16}{116} \qquad z_r \le \varepsilon$$

and

$$x_r = \frac{X}{X_r} \qquad y_r = \frac{Y}{Y_r} \qquad z_r = \frac{Z}{Z_r}$$

Table 59 provides GretagMacbeth ColorChecker CIE L*a*b* reference values, measured with illuminant D65 and D50, 2 degree observer.

Table 59: GretagMacbeth ColorChecker CIE L*a*b* Reference Values

	2 Degree	**Illuminant**	**L***	**a***	**b***
1	CC1	D65	37.542	12.018	13.33
2	CC2	D65	65.2	14.821	17.545

Table 59:   GretagMacbeth ColorChecker CIE L*a*b* Reference Values (Continued)

	2 Degree	**Illuminant**	**L***	**a***	**b***
3	CC3	D65	50.366	-1.573	-21.431
4	CC4	D65	43.125	-14.63	22.12
5	CC5	D65	55.343	11.449	-25.289
6	CC6	D65	71.36	-32.718	1.636
7	CC7	D65	61.365	32.885	55.155
8	CC8	D65	40.712	16.908	-45.085
9	CC9	D65	49.86	45.934	13.876
10	CC10	D65	30.15	24.915	-22.606
11	CC11	D65	72.438	-27.464	58.469
12	CC12	D65	70.916	15.583	66.543
13	CC13	D65	29.624	21.425	-49.031
14	CC14	D65	55.643	-40.76	33.274
15	CC15	D65	40.554	49.972	25.46
16	CC16	D65	80.982	-1.037	80.03
17	CC17	D65	51.006	49.876	-16.93
18	CC18	D65	52.121	-24.61	-26.176
19	CC19	D65	96.536	-0.694	1.354
20	CC20	D65	81.274	-0.61	-0.24
21	CC21	D65	66.787	-0.647	-0.429
22	CC22	D65	50.872	-0.059	-0.247
23	CC23	D65	35.68	-0.22	-1.205
24	CC24	D65	20.475	0.049	-0.972
1	CC1	D50	37.986	13.555	14.059
2	CC2	D50	65.711	18.13	17.81
3	CC3	D50	49.927	-4.88	-21.925
4	CC4	D50	43.139	-13.095	21.905
5	CC5	D50	55.112	8.844	-25.399
6	CC6	D50	70.719	-33.397	-0.199
7	CC7	D50	62.661	36.067	57.096
8	CC8	D50	40.02	10.41	-45.964
9	CC9	D50	51.124	48.239	16.248
10	CC10	D50	30.325	22.976	-21.587

Table 59:   GretagMacbeth ColorChecker CIE L*a*b* Reference Values (Continued)

	2 Degree	**Illuminant**	**L***	**a***	**b***
11	CC11	D50	72.532	-23.709	57.255
12	CC12	D50	71.941	19.363	67.857
13	CC13	D50	28.778	14.179	-50.297
14	CC14	D50	55.261	-38.342	31.37
15	CC15	D50	42.101	53.378	28.19
16	CC16	D50	81.733	4.039	79.819
17	CC17	D50	51.935	49.986	-14.574
18	CC18	D50	51.038	-28.631	-28.638
19	CC19	D50	96.539	-0.425	1.186
20	CC20	D50	81.257	-0.638	-0.335
21	CC21	D50	66.766	-0.734	-0.504
22	CC22	D50	50.867	-0.153	-0.27
23	CC23	D50	35.656	-0.421	-1.231
24	CC24	D50	20.461	-0.079	-0.973

C.4.5   Capture Gamma

Charge-coupled device (CCD) image sensors are linear. But the output of still and video cameras is nonlinearly encoded for several reasons:

- Nonlinear encoding corresponds closely with the eye's response. Linear 8-bit coding would have more levels than necessary in the brightest regions and too few levels for smooth response in the darkest regions, resulting in banding.

- Historically, signals required for driving displays are non-linear.

- The file encoding standards for information interchange require nonlinear response.

A camera's response to light follows the approximate equation:

$$\text{Pixel level} = k \, \text{luminance}^\gamma$$

where:

the exponent γ is the camera or capture gamma and k is a constant related to exposure and bit depth.

The standard for video cameras and several still camera color spaces is $\gamma = 1 \div 2.2 = 0.45$. When pixel level vs. luminance is displayed logarithmically, γ is the slope of the curve:

$$\log_{10}(\text{pixel level}) = \gamma \log_{10}(\text{luminance}) + k_1$$

This curve resembles the classic characteristic curve for film, where response (density, in the case of film) is plotted against log exposure. Even when the characteristic curve for camera response deviates from the simple exponential equation, as it often does, the average response can still be fitted to the exponential. Figure 38 shows an example from the Imatest application, which illustrates the deviation from the straight line at Log Exposure < -1.5, which is apparently caused by glare (which can be minimized by careful lighting). As discussed previously, Log Exposure is equal to -1 × *density*.

Figure 38:   Density Response Plotted Against Log Exposure
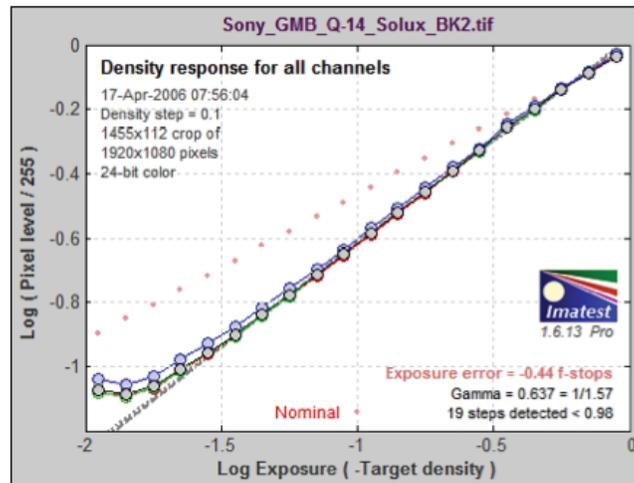
Measuring gamma requires photographing a target with patches of known density, d = -k $\log_{10}$(reflectance) = -k $\log_{10}$(Luminance), or Luminance = k 10^{-d}, where k is a constant with different values in different equations.

$$\text{Pixel level} = P = k\,10^{-d\,\gamma}$$

Solve for γ by measuring the average pixel level of two patches in the linear region, then solving:

$$P_1 = k10^{-d_1\gamma} \qquad P_2 = k10^{-d_2\gamma}$$

$$\frac{P_1}{P_2} = \frac{10^{-d_1\gamma}}{10^{-d_2\gamma}} = 10^{(d_2 - d_1)\gamma}$$

$$\log_{10}\left(\frac{P_1}{P_2}\right) = (d_2 - d_1)\gamma$$

$$\gamma = \frac{\log_{10}\left(\frac{P_1}{P_2}\right)}{d_2 - d_1}$$

Follow these steps to measure gamma:

1. Photograph a Kodak Q-13 or Q-14 step chart or a GretagMacbeth ColorChecker, illuminated using the standard lighting intensity as described in Section C.2 and Section C.3.2. Note that you can perform this measurement at the same time as other tests that use these charts.

2. Find the average measured pixel levels P_i of the patches i, excluding regions near boundaries that may not be representative of the interior. If it is convenient, plot $\log_{10}(P_i)$ as a function of $\log_{10}(\text{luminance})$, (i.e., $-1 \times density$).

3. Select two patches near the ends of the relatively linear region. Typically, this might be patches 4 and 10 of the Kodak Q-13 or Q-14 chart, or patches 2 and 5 on the bottom row of the GretagMacbeth ColorChecker chart. Label the selected patches 1 and 2 for the purpose of using the above equation for calculating γ (step 4 below).

4. As indicated in the above equation, $\gamma = \log_{10}(P_1/P_2)/(d_2 - d_1)$. Values of d_i for the Kodak Q-13 or Q-14 chart and GretagMacbeth ColorChecker chart are given in Section C.4.6. Example patches 2 and 5 for the GretagMacbeth ColorChecker chart (bottom row) have specified densities of $d_1 = 0.23$ and $d_2 = 1.05$, respectively. If the measured pixel values (from step 2) for these two patches were $P_1 = 200$ and $P_2 = 85$, then $\gamma = \log_{10}(200/85)/(1.05 - 0.23) = 0.4532$.

C.4.6 Exposure Accuracy

For a camera with a nominal capture gamma of γ, the pixel level for a correctly exposed test chart patch with *density* $d_i = -\log_{10}(\text{exposure})$ is:

$$P_i = 255\left(\frac{10^{-d_i}}{1.06}\right)^{\gamma}$$

where:

$\gamma = 1/2.2 = 0.4545$ for most video cameras and (approximately) for the sRGB color space.

For the Kodak Q-13 or Q-14 chart, $d_i = \{0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05\}$ for patches 1 to 11 (out of 20 total). The corresponding reference pixel levels for a camera with $\gamma = 0.4545$ are $P_{ri} = \{235.7, 212.3, 191.2, 172.2, 155.1, 139.6, 125.8, 113.3, 102.0, 91.9, 82.7\}$.

Table 60 lists sRGB reference values by GretagMacbeth.[48]

Table 60: GretagMacbeth ColorChecker sRGB Reference Values

Color	Name	R	G	B
1	dark skin	115	82	68
2	light skin	194	150	130
3	blue sky	98	122	157
4	foliage	87	108	67
5	blue flower	133	128	177
6	bluish green	103	189	170

Table 60:   GretagMacbeth ColorChecker sRGB Reference Values (Continued)

Color	Name	R	G	B
7	orange	214	126	44
8	purplish blue	80	91	166
9	moderate red	193	90	99
10	purple	94	60	108
11	yellow green	157	188	64
12	orange yellow	224	163	46
13	blue	56	61	150
14	green	70	148	73
15	red	175	54	60
16	yellow	231	199	31
17	magenta	187	86	149
18	cyan	8	133	161
19	white	243	243	242
20	n8	200	200	200
21	n6.5	160	160	160
22	n5	122	122	121
23	n3.5	85	85	85
24	n2	52	52	52

For the GretagMacbeth ColorChecker, the grayscale patch densities on the bottom row have nominal densities of d_i = {0.05, 0.23, 0.44, 0.70, 1.05, 1.50}. The corresponding sRGB reference pixel levels for a camera with $\gamma = 0.4545$ are P_{ri} = {243, 200, 160, 122, 85, 52}.[49]

For a set of measured pixel levels P_i, the mean exposure error (Err_{exp}) in f-stops is:

$$Err_{exp} = -3.32 \, \mathrm{mean}(\log(P_i) - \log(P_{ri}))/\gamma$$

where:

γ is the measured camera (capture) gamma, as described in Section C.4.5, and P_{ri} are the sRGB reference pixel levels for the patches (given above). For the Kodak Q-13 or Q-14 chart, take the mean over patches 4 to 11, which is the camera's probable linear region, excluding very dark and light regions. For the GretagMacbeth ColorChecker chart, take the mean over grayscale patches 2 to 5 on the bottom row

48. GretagMacbeth ColorChecker, February 2006. GretagMacbeth color checker charts. Available from http://www.gretagmacbeth.com/index/products/products_colorcheckers.htm. Cited June 14, 2006.

49. Values are from the file, *Lab & Spectral Data D65 & D50 Spectrolino.xls*, February 2006, supplied by GretagMacbeth.

(patches 20 to 23 on the chart as a whole). The following steps summarize the procedure for measuring exposure error:

1.   Photograph a Kodak Q-13 or Q-14 step chart or a GretagMacbeth ColorChecker chart, mounted on a neutral gray background as described in Section C.3.1.2, and illuminated with standard lighting intensity as described in Section C.2 and Section C.3.2.

2.   Find the average measured pixel levels P_i of the patches, excluding regions near boundaries that may not be representative of the interior.

3.   Apply the above equations, making use of measured capture gamma γ from Section C.4.5, to calculate exposure error, Err_{exp}.

4.   Exposure may be affected by history. It is worth testing the image after the lens has been covered for a few seconds and after a bright light has been reflected into the camera for a few seconds (long enough for the auto exposure to adapt).

5.   Exposure may be affected by light level and color temperature. This test may be repeated for reduced and dim lighting intensities and for light with daylight and tungsten color temperatures (see Section C.2 and Section C.3.4).

C.4.7   Vignetting

Vignetting is the falloff of light at the edges of the image. The following steps for measuring relative illumination are from the SMIA Camera Characterization Specification:[50]

1.   Photograph a flat, evenly lit surface. The surface can be composed of a matte gray or white material. Check that the surface illumination varies by no more than ±10 percent.

2.   Measure the average pixel levels in the center of the image and all four corners of the image, using small rectangular shaped sub-regions (square sub-regions preferred). Ensure the squares or rectangles are no larger than 1 percent of the total image area. That would make them about 10 percent (or slightly less) of the image dimensions on each side.

3.   The relative illumination is:

relative illumination (%) = 100% P_{center} / P_{corner}(worst case)

where:

P_{center} is the average pixel level of the center square, and P_{corner}(worst case) is the average pixel level of the darkest corner.

C.4.8   Lens Distortion

Lens distortion is the deformation of the image due to straight lines in the subject rendered as curved lines in the image from the camera.

Follow these steps to measure lens distortion:

1.   Capture a video image of the rectilinear grid test chart (see Section C.3.1.3). Ensure that the video image is framed so that A_1 and A_2 (see Figure 39) are approximately 98 percent of the image

---

50.   Section 5.17, "Relative Illumination." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

height. This does not have to be measured precisely but the rectilinear grid test chart should nearly fill the image.

2.   Apply the veiling glare formula from the Standard Mobile Imaging Architecture (SMIA) Camera Characterization Specification to the captured image.[51] The distances in the equation below (A_1, A_2, and B) can be measured electronically using image analysis software, or manually using a printed version of the image:

SMIA TV Distortion = 100($A\text{-}B$)/B ; $A = (A_1\text{+}A_2)/2$

3.   SMIA TV Distortion > 0 for pincushion distortion while SMIA TV distortion < 0 indicates barrel distortion.

Figure 39:   SMIA TV Distortion
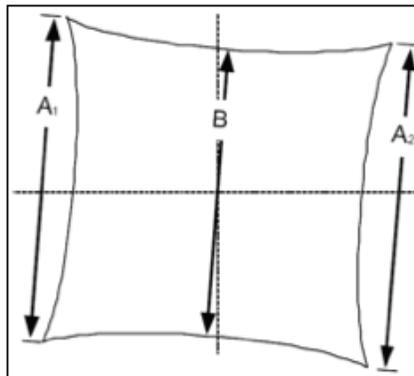
C.4.9   Reduced Light and Dim Light Measurements

Unlike still cameras, long exposures are not an option with video cameras. At low light levels the aperture is opened and the sensor gain is increased, degrading the SNR and dynamic range. Performance at low light levels cannot be inferred from measurements at high light levels. Many public safety applications involve working in reduced lighting conditions (i.e., tactical video, nighttime surveillance video). For these applications, video camera performance should be directly measured under reduced or dim lighting intensities (see "Reduced Lighting Intensity" and "Dim Lighting Intensity" in Table 57).

Select the color temperature of the reduced or dim lighting to match the color temperature that will be encountered (for example, see "Daylight Light" and "Tungsten Light" in Table 57). Figure C.3.4 presents a method for modifying lamp head assemblies to properly emulate reduced lighting with the proper color temperature. Dimmers cannot in general be used to obtain low light levels with the proper color temperature, since the color temperature drops to very low levels as the light is dimmed—well under 3,000K.

You can achieve reduced light by stacking neutral density and color-correction (CC) filters. For example, the 85B warming (amber) filter, +131 mireds, changes 4,700K to 2,900K, where 2,900K is typical of ordinary incandescent bulbs. The 85B warming filter also decreases exposure level by two thirds f-stop. Add neutral density (ND) filters to the stack to achieve additional light reduction. Typical ND filters have D = 0.3 (2x; 1 f-stop), 0.6 (4x; 2 f-stops), and 0.9 (8x; 3 f-stops). Stacking filters sums density. For example, to achieve dim lighting with two SoLux Task Lamps located 1 meter from the target

---

51. Section 5.24, "Veiling Glare." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

(approximately 250 lux at the target), 0.6 + 0.9 ND filters (density = 1.5 total) can reduce the illumination to 250/32 = 7.8 lux, well within the definition of dim lighting intensity.

You can perform the following measurements at reduced light levels (30 to 60 lux) and low light levels (5 to 10 lux), as appropriate for the application. The primary measurements are exposure accuracy, dynamic range, and color accuracy. Noise and gamma are measured to determine exposure accuracy and dynamic range:

- **Exposure Accuracy.** Underexposure is a frequent problem at low illumination. Exposure accuracy is important and easy to measure.

- **Dynamic Range.** Degraded at low light levels. A reduced dynamic range may be acceptable at low light levels.

- **Color Accuracy.** Affected by increased noise, changes in signal processing, and the tungsten light balance typical of low illumination. Reduced color accuracy may be acceptable at low light levels.

- **Noise.** Degraded at low light levels. Measured to find the dynamic range.

- **Gamma.** A secondary measurement at low light levels: measured to determine exposure accuracy.

The details of the tests depend on the specific application. Where low light performance is of interest, you would typically test a camera for exposure accuracy, dynamic range, and color accuracy at approximately 2,700 to 3,200K (typical of tungsten lighting), at reduced and dim lighting intensities, as defined in Table 57.

Note: *Resolution does not need to be tested at low light levels, unless there is suspicion that low light performance is achieved by techniques such as combining pixels, which reduce resolution .*

C.4.10   Flare Light Distortion (Under Study)

Flare light, also called spatial crosstalk, or veiling glare, is light that bounces between lens elements and off the interior barrel of the lens. Flare light can significantly reduce the dynamic range of a camera under adverse lighting conditions, for example, when there is a strong light source, such as a spotlight in the image or near the image frame. Photographers go to considerable lengths to control flare from various sources. For instance, they they use lens shades, "barn doors" on studio lights, etc., but flare light may not be controllable in public safety applications. Therefore, it is desirable to have a standard method of measurement for determining the reduction in camera DR due to flare light. This phenomenon is known as flare light distortion.

The International Electrotechnical Commission (IEC) 61966-8 has a standard for spatial crosstalk (somewhat equivalent to flare light).[52] The IEC standard was designed for scanners, and may be inadequate for cameras, where images can be degraded by light in the proximity of the image frame.

---

52. International Electrotechnical Commission, IEC 61966-8, 2001, *International Standard*. Introduction available at http://domino.iec.ch/preview/info_iec61966-8%7Bed1.0%7Den.pdf. Cited June 14, 2006. Full standard available at http://webstore.ansi.org/ansidocstore/product.asp?sku=IEC+61966-8+Ed.+1.0+en%3A2001. Cited June 14, 2006.

Flare light is called "veiling glare" in section 5.24 of the SMIA Camera Characterization Specification.[53] The SMIA measurement is satisfactory when: 1) a manual exposure setting is available; and 2) access to the raw or unprocessed output of the sensor is available. Unfortunately, these conditions rarely apply to video cameras. Veiling glare is "veiled" by the signal processing in the video camera. A more sensitive measurement is necessary.

There is no widely accepted measurement for flare light distortion at this time. Hence flare is usually described qualitatively in published lens test reports.

Measuring flare involves several difficulties. The assumption cannot be made that cameras respond linearly to light, or linearly with a gamma curve. Many cameras have "S" tonal response curves that improve pictorial quality by boosting contrast ($\Delta C = \Delta(\log_{10}(P))/\Delta d$, see Step 7 of Section C.4.3.2) of the middle tones at the expense of contrast in the shadows and highlights. It is important not to confuse the decrease in contrast near the toe of the "S" curve with flare. Furthermore, some cameras have adaptive signal processing, which means gain may be decreased or increased, depending on the scene contrast. This improves pictorial quality but complicates measurements.

One possible flare light distortion test takes advantage of the fact that flare tends to reduce step chart contrast in shadow areas. The measurement under study involves photographing a standard reflective step chart, such as the Kodak Q-13 or Q-14, against a large black-and-white background, for example, 32 by 40 inch (80 by 100 cm) foam board. The charts can be surrounded by a small neutral gray region, just large enough to influence the auto exposure setting, and no larger.

C.5   MAKEOECF.M

The following MATLAB m-file script creates an OECF look up table (LUT) for use by the sfrwin application. The sfrwin application assumes a gamma of 1.0 if no LUT is provided to it, and this will result in the MTF measurement being off by 10 to 20 percent. The output file generated by the makeoecf.m script will either contain a lookup table with one column (for black and white) or three columns (for color), depending upon how the script is called. See the instruction manual (user_guide.pdf) that comes bundled with the sfrwin application for details on how to input a user-generated LUT.

```
function makeoecf(varargin)
% makeoecf (gamma, ncol, filename)
% Create an OECF file for sfrwin.
% First argument is gamma. Default = 0.5
% Second argument is ncol: 1 (for Black & White), 3 (for color). Default = 3
% Third argument is the filename (no blanks). Default = lut_gamma_ncol.dat
%
% Example Calls: makeoecf ('0.5','3', 'Lut_0.5_3.dat')
%                makeoecf ('0.5')

% Set defaults
gamma = 0.5;
ncol = 3;

if (nargin>=1)     % gamma between 0.1 and 10
   gamma = str2num(varargin{1});
   if gamma<.1 | gamma>10
      disp('Gamma must be a number between 0.1 and 10.');  return;
   end
```

---

53. Section 5.24 "Veiling Glare." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

```
end
if (nargin>=2)     % 1 or 3 for B&W or color
   ncol = str2num(varargin{2});
   if ~(ncol==1 | ncol==3)
      disp('Number of colors (2nd argument) must be 1 or 3.');  return;
   end
end
if (nargin>=3)     % Output file name.
   fileout = varargin{3};
else
   fileout = ['lut_' num2str(gamma,2) '_' num2str(ncol) '.dat'];
end

foecf = 255*linspace(0,1,256).^(1/gamma);

fid = fopen(fileout, 'w');
if ncol==1  % B&W
   fprintf(fid, '%7.2f\n',foecf);
else        % ncol==3: color
   foecf = [foecf; foecf; foecf];  % 3 columns.
   fprintf(fid, '%7.2f %7.2f %7.2f\n',foecf);
end
fclose(fid);
disp(['End makeoecf. File written to ' fileout '.']);
return;
```

Appendix D   Video Quality Experiment PS1

This appendix describes a laboratory study to investigate the level of quality required for narrow field of view tactical video. The study involved a video quality questionnaire and a subjective video quality experiment. The video quality questionnaire was used to obtain an initial estimate for the minimum quality levels that are required for public safety video applications. After the questionnaire results were examined, the first public subjective video quality experiment, PS1, was conducted. This appendix describes the questionnaire and subjective experiments and presents the results from the data analysis.

D.1   Video Quality Questionnaire

Since the use of video in public safety applications is relatively new, a questionnaire was used to gather fundamental video quality information (e.g., frame rates, and luma image size) that could be used to design the first subjective video quality experiment for public safety applications (PS1). This questionnaire, conducted from July to September, 2005, was given to 18 public safety practitioners from around the United States.

Practitioners were provided with a list of public safety video applications, and asked to rank their importance. This ranking was used to decide the highest priority public safety video application that the PS1 experiment should address. For each application where the practitioner had relevant experience, application-specific questions were asked.

The definitions used in this questionnaire to differentiate between the various public safety video applications confused some practitioners. As a result, the video categories have since been refined and the definitions improved. To avoid potential confusion going forward, the original definitions and categories in the questionnaire will not be presented here.

For each specific application (e.g., tactical video, surveillance video), practitioners were asked questions concerning acceptable values for the following video characteristics:

- Video Delay
- Control Lag
- Image Size
- Frame Rate
- Lossless Video Quality Requirement
- Coding Impairment
- Response to Network Impairments
- Color Fidelity
- Lighting Requirements
- Focus Distance
- Transmission Distance (between source and destination ends)
- Camera and Monitor Mobility

For each characteristic, an example level of service was presented using a combination of text, still imagery, and video sequences. High-speed computer hardware was required to correctly play back the high resolution video sequences, which unfortunately prevented the questionnaire from being more widely distributed. Practitioners were asked to mark each level of service as either "desirable," "acceptable," or "unacceptable." This scale was well-received and understood.

The following preliminary quality levels were indicated for tactical video, which was the application considered as the very highest priority by those participating in the questionnaire. (The following quality levels listed as "borderline" were marked as "acceptable" or "desirable" for tactical video by 60 percent to 80 percent of responders, and "unacceptable" by 20 percent to 40 percent of responders.) All results obtained from the questionnaire were treated as general indications only, due to confusions regarding the definitions mentioned above and the small number of practitioners that were questioned:

- Image Size: HDTV is desired, NTSC/525-line is acceptable, and CIF is unacceptable. (The other decreasing image size that was included, QCIF, is unacceptable.)

- Video Delay: Real-time is desired, near real-time is acceptable, and from one to several seconds delay is borderline.

- Control Lag: Real-time is desired, near real-time is acceptable, and from one to several seconds delay is borderline.

- Frame Rate: 30 fps is desired, 10 fps is acceptable, and 5 fps is borderline.

- Lossless Video: Participants have been told they need lossless compression, but probably do not know what this means or why they need it.

- Video Quality: "Imperceptible" is desired, "Perceptible but Not Annoying" is acceptable, and "Slightly Annoying" is borderline. (The other two decreasing quality levels that were included, "Annoying" and "Very Annoying", are unacceptable.)

- Response to Errors: Transient drops in quality are undesirable. Consistency was preferred over quality. Video Quality is more important than delay.

- Color: Color is very important, both in terms of accurate color rendition and color display being preferred over black and white.

- Lighting and Distance: All lighting levels are important—indoors, daylight, night with dim light and no noticeable light. Distances less than 200 yards are important; distances farther away are less important.

As a result of the application rankings, the first application considered was narrow field of view tactical video. The PS1 experiment design focused on 525-line/NTSC video sampled according to ITU-R Recommendation BT.601 (Rec. 601), with a limited number of systems being included that operated at CIF and QCIF image sizes. This PS1 experiment was also limited to a minimum frame rate of 5 fps and to color video (i.e., no black and white). The experiment was designed to have an average quality level that was centered on the quality level indicated by the questionnaire responses.

D.2   Description of PS1 Subjective Video Quality Experiment

D.2.1   Overview

The subjective video quality experiment PS1 was conducted from September 2005, to February 2006. PS1 focused on determining the minimum acceptable video transmission quality that is required to support

narrow field of view tactical video applications. This would involve determining fundamental characteristics of the video transmission subsystem (see Figure 3 on page 14) such as image size, frame rate, coder type, coder bit rate, and packet loss ratio. High-quality source video sequences (point B in Figure 3) were used wherever possible so as to not influence the outcome of the experiment. Likewise, a studio-grade television monitor was used to display the resulting videos during the subjective test (points E to F in Figure 3), so that the display would not add any impairments.

During the subjective test, subjects watched a number of short video sequences and judged each of them for quality and acceptability. Subjects rated the quality of each sequence after they had seen it, using the following five-point grading scale: Excellent, Good, Fair, Poor, and Bad. Subjects were also asked to judge whether the video clip was acceptable or unacceptable for use as narrow field of view tactical video. Subjects had a short time to make their judgment of the video quality and acceptability and mark ratings on a printed score sheet. The first rating session lasted approximately 20 minutes, followed by a break. After the break, there was another similar session. Viewing sessions were held in a controlled viewing environment (see ITU-T Recommendation BT.500).

Before participating in the two viewing sessions, subjects were tested for visual acuity and color perception, and underwent a training session where they were reminded that public safety personnel use tactical video in real time during an incident to make decisions on how to respond to that incident.

The following examples were given to subjects:

- A camera carried by a public safety practitioner into a burning building to provide the incident commander with situation information.

- A body-worn camera.

- An aerial camera following a subject on foot.

D.2.2   Experiment Design

The PS1 experiment included 47 Hypothetical Reference Circuits (HRCs).[54] HRCs were selected for experiment PS1 to investigate the minimal requirements for the following attributes:

- Coder Bit Rate for H.264 without network impairments

- Coder Bit Rate for MPEG-2 without network impairments

- Image Size

- Frame Rate

- Packet Loss Ratio

- Error Concealment Strategies (including no error concealment)

Of these, packet loss ratio and error concealment are the most difficult to characterize using short video sequences. This is because the location of the lost packet within the video transmission stream can significantly impact the reduction in perceived video quality, and the perceptual impact of packet loss is significantly affected by the particular error concealment scheme being used. The large number of

---

54. Hypothetical Reference Circuit (HRC) is an industry-accepted term for a specific configuration of a video transmission subsystem (i.e., fixed configuration settings for the behavior of the video coder, the network, and the video decoder).

combinations for the above video quality variables limited what could be examined for any one particular variable.

Video was always displayed to subjects at Rec. 601 image size. Original video sequences produced with higher resolutions (e.g., high-definition) were down-sampled to Rec. 601 using professional-grade hardware. The video produced by HRCs that used a lower resolution (e.g., CIF) were up-sampled to Rec. 601 prior to use in the subjective test.

Sixteen source video sequences were selected, each 8 seconds in duration. These video sequences were split into two sets of eight sequences. The HRCs were likewise split into two sets of 25 HRCs.[55] Half of the video sequences were paired with half of the HRCs to produce 400 video sequences (i.e., 8 scenes × 25 HRCs × 2 sets = 400 video sequences). To prevent viewer fatigue, each viewer was asked to rate only half of the video sequences.

The 400 clips were randomized such that no HRC or scene appeared consecutively during a viewing session. The clips were divided into four session tapes of 100 clips each (1, 2, 3, and 4). Each viewer saw two of these tapes in a randomized order (e.g., 4 then 2, 2 then 1, 1 then 3, etc.). Every attempt was made to assure that the same number of viewers saw each of 12 randomized tape orderings.

D.2.3   Original Video Sequences

Original video sequences were selected from the following sources:

- Footage shot at a football game with a shoulder-mounted camera that followed police officers as they performed their duties. Footage was shot in HDTV 720p format, and then converted to Rec. 601.

- Footage shot at a firefighter training session using a shoulder-mounted camera. Footage was shot in HDTV 720p format and then converted to Rec. 601.

- In-car camera footage depicting simulated drunk driving stops. Video was shot in HDTV 1080i format, and then converted to Rec. 601.

- Footage of a SWAT team training session. Footage was received on a VHS tape and digitally sampled in accordance with Rec. 601.

- Footage of an underwater crime investigation. Footage was received on a VHS tape, and digitally sampled in accordance with Rec. 601.

Sixteen video sequences were selected, each 8 seconds in duration. These video sequences were split into two sets of eight sequences. Scene set A was matched with H.264 codec impairments and some simulated impairments. Scene set B was matched with MPEG-2 codec impairments, some simulated impairments, and several H.264 codec impairments. Video sequences were selected to meet the following needs:

- Match the definition of narrow field of view tactical video, so that subjects could envision using the system in real time during an actual incident.

- Provide a variety of visually different scene content.

- Span a wide range of scene coding difficulty, from easy-to-code (i.e., little motion, little spatial detail) to hard-to-code (i.e., high motion, abundant and intricate details).

---

55. Three of the 47 HRCs were present in both sets.

The video sequences within each set are described below. The original video format is listed in parenthesis. All video sequences were converted from their original format to Rec. 601 prior to use in the experiment.

Scene Set A

1.   Zoom out of a crowd in a football stadium. (720p)

2.   Underwater crime scene investigation where a gun is found on a boat. Water contains many floating particles. (VHS (Video Home System) recording media)

3.   Officer watching a dense crowd of people walk past. (720p)

4.   Firefighter trainee squeezing through a wall. (720p)

5.   Camera bouncing while following an officer from car to simulated traffic stop, daytime. (1080i)

6.   Nighttime simulated stop, zooming in on officer approaching a stopped car. The driver's face could be seen in the car's driver-side mirror, and license plate number was visible. (1080i)

7.   Walking up an outside fire escape, simulating a shoulder- or helmet-mounted camera. (720p)

8.   Scene focused on a fire, which is extinguished by water, forming smoke; then panning back and zooming out to show firefighters spraying water. The camera wobbles a bit. (720p)

Scene Set B

1.   Pan across a crowded stadium at a football game. (720p)

2.   Night-time footage of a SWAT team deployed around a door, about to knock it down. (VHS)

3.   Close-up of a woman undergoing a sobriety test. Camera zooms from showing her entire upper body to just her eyes. (1080i)

4.   Firefighter backing up through a large cement pipe. (720p)

5.   Camera outside a car window positioned low to the ground as the car drives along a road. (720p)

6.   Following officers escorting people out of a football game; some camera flare from bright sunlight. (720p)

7.   Inside a driving car with the camera pointed to the driver's side of the car, showing a policeman driving; moving scenery is visible outside the driver's window. (720p)

8.   Dumpster fire with background blurred by smoke, panning to show people through the smoke. (720p)

Obtaining high-quality original video footage with suitable content is difficult. Due to usage restrictions, sample video frames from most of the video sequences cannot be displayed in this report.

D.2.4   HRC Video Transmission Systems

Each video transmission system sample, referred to as a hypothetical reference circuit (HRC) by the video quality measurement community, involves taking the original video sequences and changing or distorting them in some way. The HRCs chosen for subjective experiment PS1 can be divided into three categories:

- First, the original video sequences were modified to reflect different image resolutions and frame rates. (These will be referred to as synthetic HRCs.)

- Second, the original sequences were MPEG-2 encoded, transmitted over an IP network, and then decoded. MPEG-2 was chosen due to the wide prevalence of this established coding scheme.

- Third, the original sequences were H.264 encoded, transmitted over an IP network, and then decoded. H.264 was chosen as a state of the art coding method that is widely regarded as requiring only one-half to one third the bit rate of MPEG-2 for the equivalent quality. Thus, within the next several years, H.264 can reasonably be expected to become the codec of choice for many new services being deployed.

Given the time constraints for completion of the PS1 experiment, locating MPEG-2 and H.264 video codecs with error concealment proved to be a challenge. As an unfortunate result, most of the video transmission systems included in the PS1 experiment do not implement error concealment. One hardware-based H.264 system with error concealment was included in the experiment. However, this system was limited to 384 kbps and below, and was limited to CIF image size. According to the questionnaire results, the video quality of H.264 at that bit rate and image size was expected to be unacceptable. (However, this did not prove to be the case—see Section D.3.) A high-priority item for study in future tests will be to include more H.264 codecs with advanced error concealment schemes.

Codecs without error concealment were paired with packet loss ratios of 0, 0.1, 0.5, and 1.5 percent. Significantly higher levels of packet loss ratios (3.3, 5.6, and 11.5 percent) were initially considered for these codecs. However, a visual examination of sample scenes transmitted at these levels of packet loss ratio indicated a level of quality significantly below the minimum quality indicated by the questionnaire responses. Higher packet loss levels (e.g., 2, 3, 6, and 12 percent) were retained for the H.264 codec with error concealment. To ensure that the minimum video quality required by practitioners was presented, the subjective test had to extend significantly beyond unacceptable (at the low-quality end) and acceptable (at the high-quality end).

Table 61 lists the synthetic HRCs that were created to explore the suitability of various frame rates and image sizes. The Scene Set column identifies the set of scenes that were used for each HRC.

Table 61:   Synthetic HRCs of Various Frame Rates and Image Sizes

HRC	Scene Set	Description
original	A, B	Original unimpaired video sequence.
sif	A	Down-sample by 2 horizontally and vertically to SIF resolution, then up-sampled back to Rec. 601 image size using pixel interpolation.
qsif	A	Down-sample by 4 horizontally and vertically to QSIF resolution, then up-sampled back to Rec. 601 image size using pixel interpolation.
fps15	B	15 fps video sequence, created by discarding every other frame, and replacing each discard with a duplicate of the previous frame.
fps10	B	10 fps video sequence, created by discarding two of every three frames, and replacing them with a duplicate of the previous frame.
fps5	B	5 fps video sequence, created by discarding five of every six frames, and replacing them with a duplicate of the previous frame.
fps10sif	A, B	First create a 10 fps video sequence by discarding two of every three frames, and replacing them with a duplicate of the previous frame. Then, down-sample by 2 horizontally and vertically to SIF resolution, then up-sample back to Rec. 601 image size using pixel interpolation.

Table 61: Synthetic HRCs of Various Frame Rates and Image Sizes (Continued)

HRC	Scene Set	Description
fps10qsif	A, B	First create a 10 fps video sequence by discarding two of every three frames, and replacing them with a duplicate of the previous frame. Then, down-sample by 4 horizontally and vertically to QSIF resolution, then up-sample back to Rec. 601 image size using pixel interpolation.

D.2.4.1   MPEG-2 Software Encoder

This MPEG-2 software encoder was operated at constant bit rate (CBR) encoding with a Group of Pictures (GOP) structure equal to "I_BB_P_BB_P_BB_P_BB_P_BB_". This produced two Intra-coded (I) frames each second. I-frames limit the propagation of errors since they encode only for spatial redundancy (i.e., no temporal redundancy). Coding was performed using Rec. 601 image size at a frame rate of 30 fps. The MPEG-2 program stream created by the encoder was encapsulated into a transport stream (*.ts), and streamed over IP using UDP/RTP multicast. A network impairment emulator was used to randomly drop packets, 1358 bytes in size. The MPEG-2 decoder did not implement error concealment (EC). Errors generally appeared as error blocks or horizontal strips, as Figure 40 shows (a packet loss ratio of 0.1 percent was used for this example).

Figure 40:   MPEG-2 HRCs with 0.1 Percent Packet Loss (No EC) Example

HRC names that utilized the MPEG-2 software codec begin with the character "M". The following additional abbreviations are used within the MPEG-2 software HRC names to denote the coder bit rate and packet loss ratio:

- 768k   — Coded at 786 kbps

- 1.5M   — Coded at 1.5 Mbps

- 3.1M   — Coded at 3.1 Mbps

- 6.1M   — Coded at 6.1 Mbps

- –0 percent   — 0 percent packet loss ratio

- –0.1 percent   — 0.1 percent packet loss ratio

- –0.5 percent   — 0.5percent packet loss ratio

- –1.5 percent   — 1.5 percent packet loss ratio

Table 62 provides a summary description of the MPEG-2 software HRCs that were used in the PS1 experiment, including the scene set used for each HRC.

Table 62:   MPEG-2 Software HRCs

HRC	Scene Set	Description
M768k-0 percent	B	Software MPEG-2, 768 kbps, 30fps, 0 percent packet loss ratio
M768k-0.1 percent	B	Software MPEG-2, 768 kbps, 30fps, 0.1 percent packet loss ratio
M768k-0.5 percent	B	Software MPEG-2, 768 kbps, 30fps, 0.5 percent packet loss ratio
M768k-1.5 percent	B	Software MPEG-2, 768 kbps, 30fps, 1.5 percent packet loss ratio
M1.5M-0 percent	B	Software MPEG-2, 1.5 Mbps, 30fps, 0 percent packet loss ratio
M1.5M-0.1 percent	B	Software MPEG-2, 1.5 Mbps, 30fps, 0.1 percent packet loss ratio
M1.5M-0.5 percent	B	Software MPEG-2, 1.5 Mbps, 30fps, 0.5 percent packet loss ratio
M1.5M-1.5 percent	B	Software MPEG-2, 1.5 Mbps, 30fps, 1.5 percent packet loss ratio
M3.1M-0 percent	B	Software MPEG-2, 3.1 Mbps, 30fps, 0 percent packet loss ratio
M3.1M-0.1 percent	B	Software MPEG-2, 3.1 Mbps, 30fps, 0.1 percent packet loss ratio
M3.1M-0.5 percent	B	Software MPEG-2, 3.1 Mbps, 30fps, 0.5 percent packet loss ratio
M3.1M-1.5 percent	B	Software MPEG-2, 3.1 Mbps, 30fps, 1.5 percent packet loss ratio
M6.1M-0 percent	B	Software MPEG-2, 6.1 Mbps, 30fps, 0 percent packet loss ratio
M6.1M-0.1 percent	B	Software MPEG-2, 6.1 Mbps, 30fps, 0.1 percent packet loss ratio
M6.1M-0.5 percent	B	Software MPEG-2, 6.1 Mbps, 30fps, 0.5 percent packet loss ratio

Table 62:   MPEG-2 Software HRCs (Continued)

HRC	Scene Set	Description
M6.1M-1.5 percent	B	Software MPEG-2, 6.1 Mbps, 30fps, 1.5 percent packet loss ratio

D.2.4.2   H.264 Software Encoder

The H.264 software encoder used version 10.1 of the Joint Model (JM), which was developed by the Joint Video Team, a collaborative effort between MPEG and the Video Coding Experts Group (VCEG). Coding was performed using Rec.601 image size at two different frame rates (15 and 30 fps), with one I-frame every second. Encapsulation was done at the VCL (video coding layer) level using the H.264 NAL (Network Abstraction Layer) header option for RTP streaming (see ITU-T Recommendation H.264). The RTP/UDP/IP streaming was based on the UCL (University College London) package. Packets 600 bytes in size were randomly dropped before decoding.[56] The decoder did not implement any error concealment. Errors appear as dropped blocks of video (i.e., black) as Figure 41 shows (a packet loss ratio of 0.1 percent was used for this example).

Figure 41:   H.264 HRCs with 0.1 Percent Packet Loss (No EC) Example
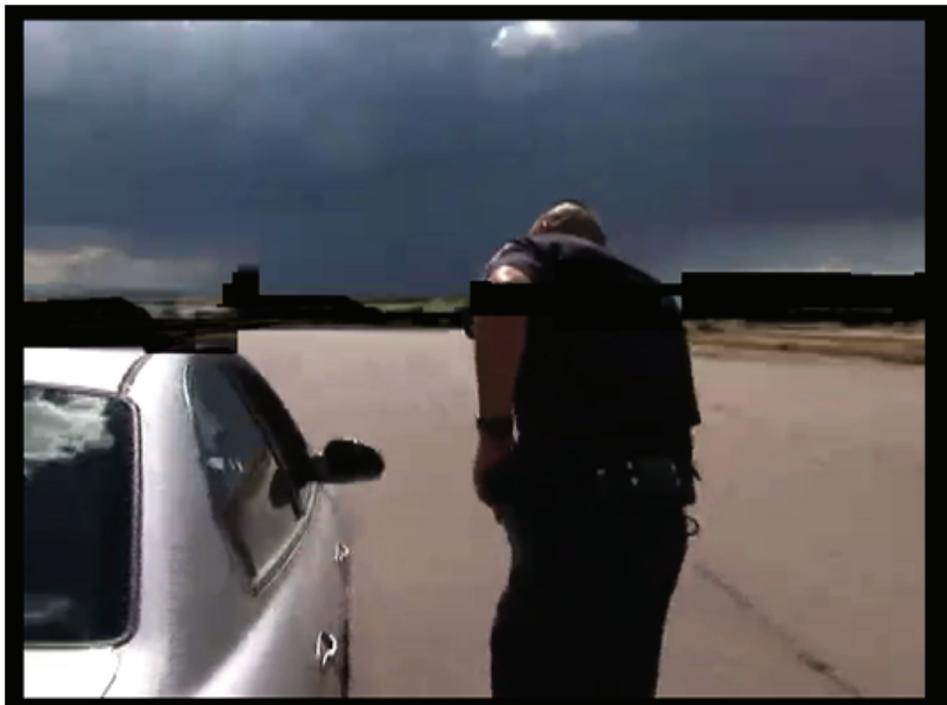
HRC names that utilized the H.264 software codec begin with the letter "S". The following additional abbreviations are used within the H.264 software HRC names to denote the coder bit rate, the frame rate

---

56. HRCs that utilized the H.264 software codec were generated by the Wireless Communications Technology Group at the National Institute of Standards and Technology (NIST).

(only used when two different frame rates were examined for the same coder bit rate), and the packet loss ratio:

- 384k   — Coded at 384 kbps
- 768k   — Coded at 786 kbps
- 1.5M   — Coded at 1.5 Mbps
- 3.1M   — Coded at 3.1 Mbps
- 6.1M   — Coded at 6.1 Mbps
- –0 percent   — 0 percent packet loss ratio
- –0.1 percent   — 0.1 percent packet loss ratio
- –0.5 percent   — 0.5percent packet loss ratio
- –1.5 percent   — 1.5 percent packet loss ratio

Table 63 provides a summary description of the H.264 software HRCs that were used in the PS1 experiment, including the scene set used for each HRC.

Table 63:   H.264 Software HRCs without Error Concealment

HRC	Clips Set	Description
S384k-0 percent	A	Software H.264, 384 kbps, 15 fps, 0 percent packet loss
S384k-0.1 percent	A	Software H.264, 384 kbps, 15 fps, 0.1 percent packet loss
S384k-1.5 percent	A	Software H.264, 384 kbps, 15 fps, 1.5 percent packet loss
S768kA-0 percent	A	Software H.264, 768 kbps, 15 fps, 0 percent packet loss
S768kA-0.1 percent	A	Software H.264, 768 kbps, 15 fps, 0.1 percent packet loss
S768kA-0.5 percent	A	Software H.264, 768 kbps, 15 fps, 0.5 percent packet loss
S768kA-1.5 percent	A	Software H.264, 768 kbps, 15 fps, 1.5 percent packet loss
S768kB-0 percent	A	Software H.264, 768 kbps, 30 fps, 0 percent packet loss
S768kB-0.1 percent	A	Software H.264, 768 kbps, 30 fps, 0.1 percent packet loss
S768kB-1.5 percent	A	Software H.264, 768 kbps, 30 fps, 1.5 percent packet loss
S1.5M-0 percent	A	Software H.264, 1.5 Mbps, 30 fps, 0 percent packet loss
S1.5M-0.1 percent	A	Software H.264, 1.5 Mbps, 30 fps, 0.1 percent packet loss
S1.5M-0.5 percent	A	Software H.264, 1.5 Mbps, 30 fps, 0.5 percent packet loss
S1.5M-1.5 percent	A	Software H.264, 1.5 Mbps, 30 fps, 1.5 percent packet loss
S3.1M-0 percent	A	Software H.264, 3.1 Mbps, 30 fps, 0 percent packet loss

Table 63:   H.264 Software HRCs without Error Concealment (Continued)

HRC	Clips Set	Description
S3.1M-0.1 percent	A	Software H.264, 3.1 Mbps, 30 fps, 0.1 percent packet loss
S3.1M-1.5 percent	A	Software H.264, 3.1 Mbps, 30 fps, 1.5 percent packet loss

D.2.4.3   Hardware H.264 Codec

The hardware H.264 codec was set up to optimize motion rather than detail. With this setting, the system down-sampled the image size to SIF prior to transmission and coding, and then up-sampled the image size back to Rec. 601 after decoding and error concealment. No other control was possible over the frame rate or coding parameters such as how often I-frames are sent. A network impairment emulator was used to randomly drop packets of variable packet size, with an average size of 460 bytes. The H.264 operating mode of this codec was limited to 384 kbps or lower (this codec was only operated at 384 kbps). The hardware H.264 HRCs implemented error concealment, which appeared as Figure 42 illustrates. (A packet loss ratio of 3 percent was used for this example.)

Figure 42:   H.264 HRCs Error Concealment Example

HRC names that utilized the H.264 hardware codec begin with the character "H". The following additional abbreviations are used within the H.264 hardware HRC names to denote the bit rate and packet loss ratio::

- 384k   —   Coded at 384 kbps

- –0 percent   —   0 percent packet loss ratio

- –1 percent   —   1.0 percent packet loss ratio

- –2 percent   —   2.0 percent packet loss ratio

■   384k   —   Coded at 384 kbps

■   –3 percent   —   3.0 percent packet loss ratio

■   –6 percent   —   6.0 percent packet loss ratio

■   –12 percent   —   12.0 percent packet loss ratio

Table 64 provides a summary description of the H.264 hardware HRCs (with error concealment) that were used in the PS1 experiment, including the scene set that was used for each HRC.

Table 64:   H.264 Software HRCs with Error Concealment

HRC	Scene Set	Description
H384k-0 percent	A	Hardware H.264, 384 kbps, error concealment, 0 percent packet loss
H384k-1 percent	A	Hardware H.264, 384 kbps, error concealment, 1 percent packet loss
H384k-2 percent	B	Hardware H.264, 384 kbps, error concealment, 2 percent packet loss
H384k-3 percent	A	Hardware H.264, 384 kbps, error concealment, 3 percent packet loss
H384k-6 percent	B	Hardware H.264, 384 kbps, error concealment, 6 percent packet loss
H384k-12 percent	A	Hardware H.264, 384 kbps, error concealment, 12 percent packet loss

D.2.5   Viewers

Thirty-five public safety practitioners were recruited to participate in the PS1 subjective video quality experiment. The practitioners came from across the country. Various local jurisdictions (29), state jurisdictions (6), and Federal jurisdictions (2) were represented. Disciplines represented included fire, law enforcement, and emergency medical services, with 18, 13, and 9 practitioners respectively.[57] Three practitioners had less than 10 years experience, 11 practitioners had 10 to 20 years experience, 14 practitioners had 20 to 30 years experience, and 7 practitioners had more than 30 years of experience. A total of 34 practitioners were males and 1 was female. Roughly 25 percent were in their thirties, about 50 percent were in their forties, and 25 percent were in their fifties.

Subjects were screened for color perception and visual acuity at a distance of 10 feet. Each subject participated in two viewing sessions of 100 video clips each. Six sessions of data were discarded (i.e., one subject's scores for one tape) due to missing data (e.g., due to a power outage or missed scores) or extremely low correlation to the overall mean of the other viewer's scores (e.g., possible fatigue or inattention). The remaining data provided 16 viewer scores for each of the 400 video clips.

D.3   PS1 Data Analysis

The PS1 video data set is comprised of 400 separate video clips, which can be further broken down into 16 separate scenes. Each scene has been processed through three different codecs, designated here as M (MPEG-2 software codec), S (H.264 software codec) and H (H.264 hardware codec). The last codec includes error concealment, and was only operated at a coded bit rate of 384 kbps. Each video codec, with

---

57. Some practitioners represented more than one jurisdiction and/or discipline.

the exception just noted, was operated at multiple coder bit rates, and all video streams were subjected to multiple degrees of packet loss. In addition, lossless video streams with different image resolutions and frame rates were also presented to each practitioner.

The data consists of 16 observations for each video clip. Each viewer scored each clip for Mean Opinion Score (MOS) on a 1 to 5 point scale (where 1 is "bad" and 5 is "excellent"), as well as acceptability on a 0 (non-acceptable) and 1 (acceptable) point scale. These data were aggregated in multiple ways to arrive at conclusions on viewer preferences in codec, coded bit rate, frame rate, image size, and packet loss tolerance. Some minor conclusions on error concealment are also evident from this experiment, although the paucity of data from the error concealing codec that was available means that such conclusions must be considered tentative at best.

The data presented here have been aggregated over all viewers and all scenes, so that each data point represents the performance of one HRC. Thus, the fraction acceptability values take into account variations among viewers' opinions, and variations due to changing scene content.

Figure 43 graphs a comparison of the correlation between the acceptability scale and the MOS scale. The two subjective scales are very well-correlated. Although both acceptability and MOS were measured for each HRC, only acceptability will be used for the remainder of this report.

Figure 43:   Acceptability Scale and MOS Scale Correlation Comparison
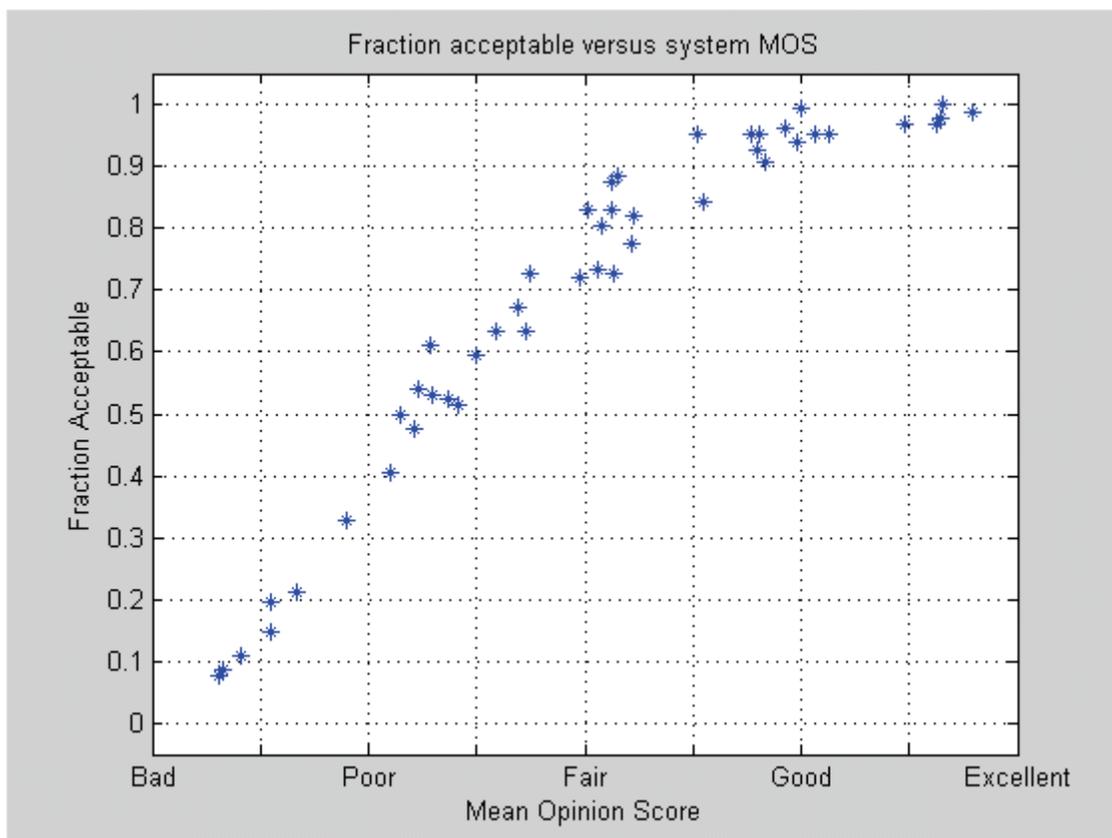
The results are grouped into four different sets, presented in the pages that follow:

- MPEG-2 software HRCs
- H.264 software HRCs

- H.264 hardware HRCs

- Synthetic HRCs

D.3.1   MPEG-2 Software HRCs

The graphs in Figure 44 through Figure 47 give the fraction of acceptable scores for each HRC with the given packet loss ratio. The horizontal axis of each graph represents the coded bit rate. The short red vertical lines that bisect the top edges of each bar graph give the extent or range of the 95 percent confidence interval for the estimate. A table also presents the results for each of the four sets of data.

The first set of graphs and the table describes the results for the MPEG-2 software HRCs. This MPEG-2 codec does not implement any error concealment (i.e., no EC), and represents one of the most commonly used video coding technologies.

Figure 44:   MPEG-2 Software HRCs with 0 Percent Packet Loss (No EC)
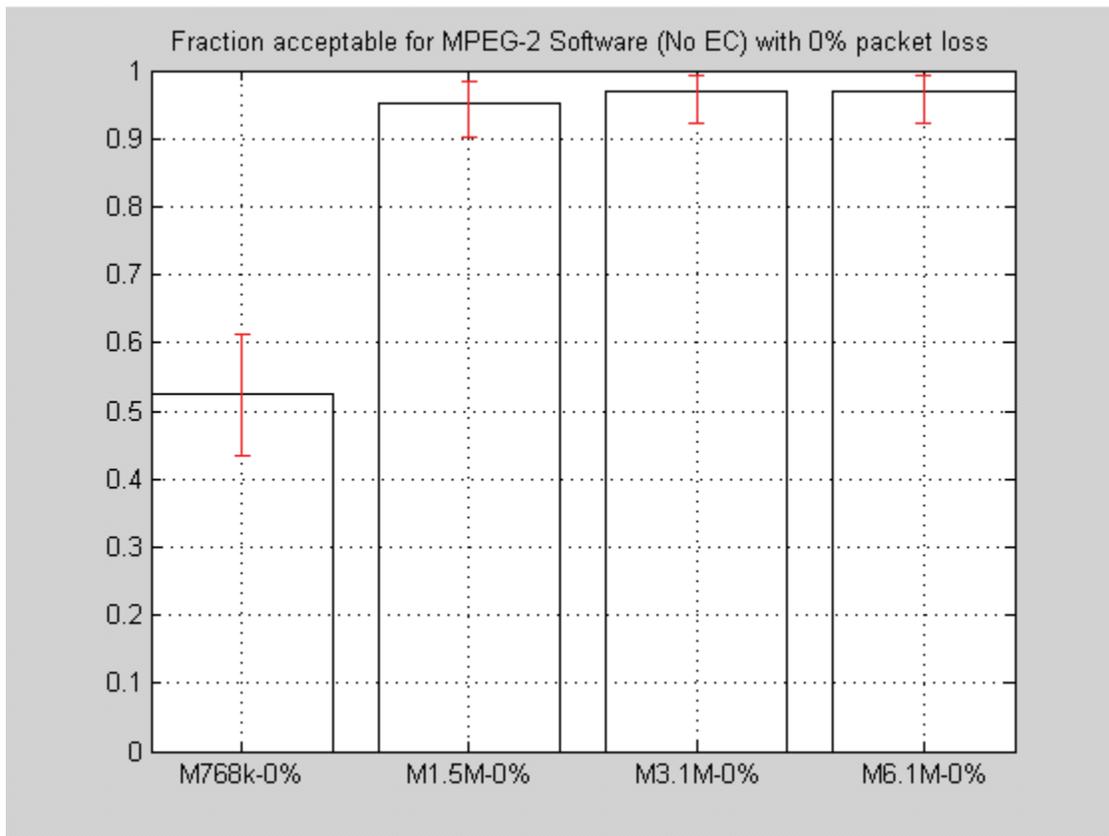
Figure 45:   MPEG-2 Software HRCs with 0.1 Percent Packet Loss (No EC)
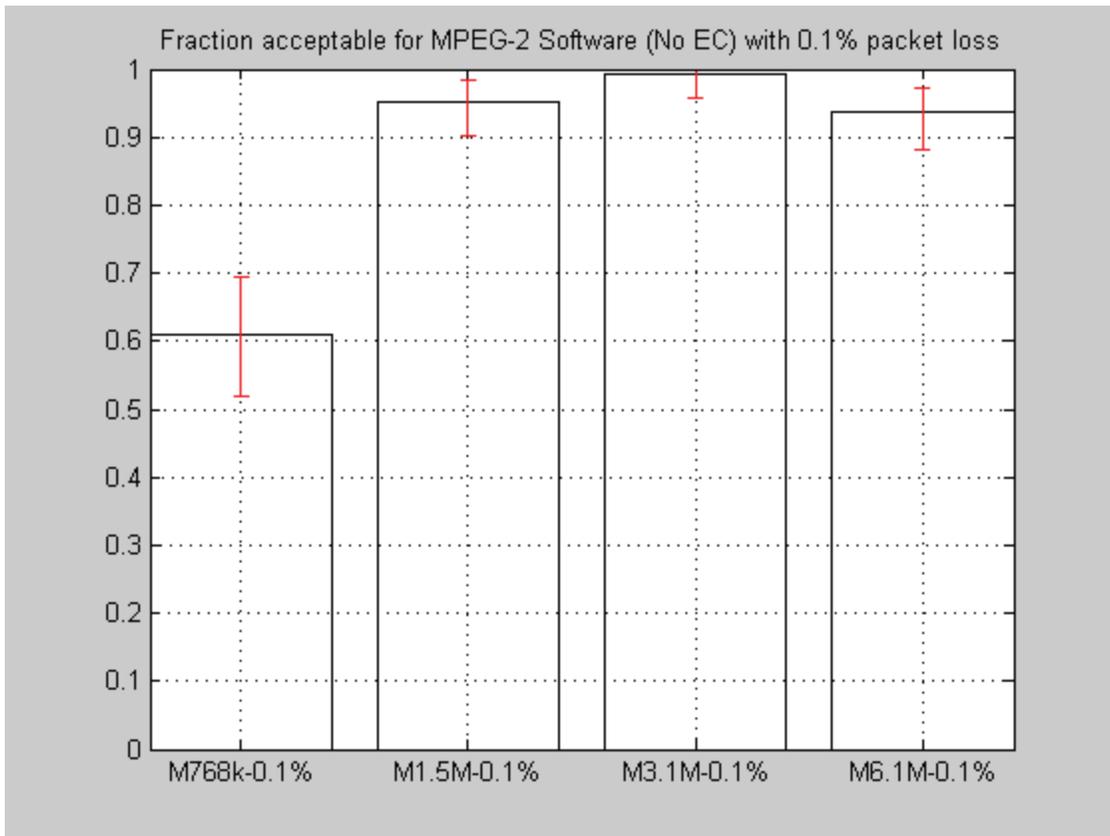
Figure 46:   MPEG-2 Software HRCs with 0.5 Percent Packet Loss (No EC)
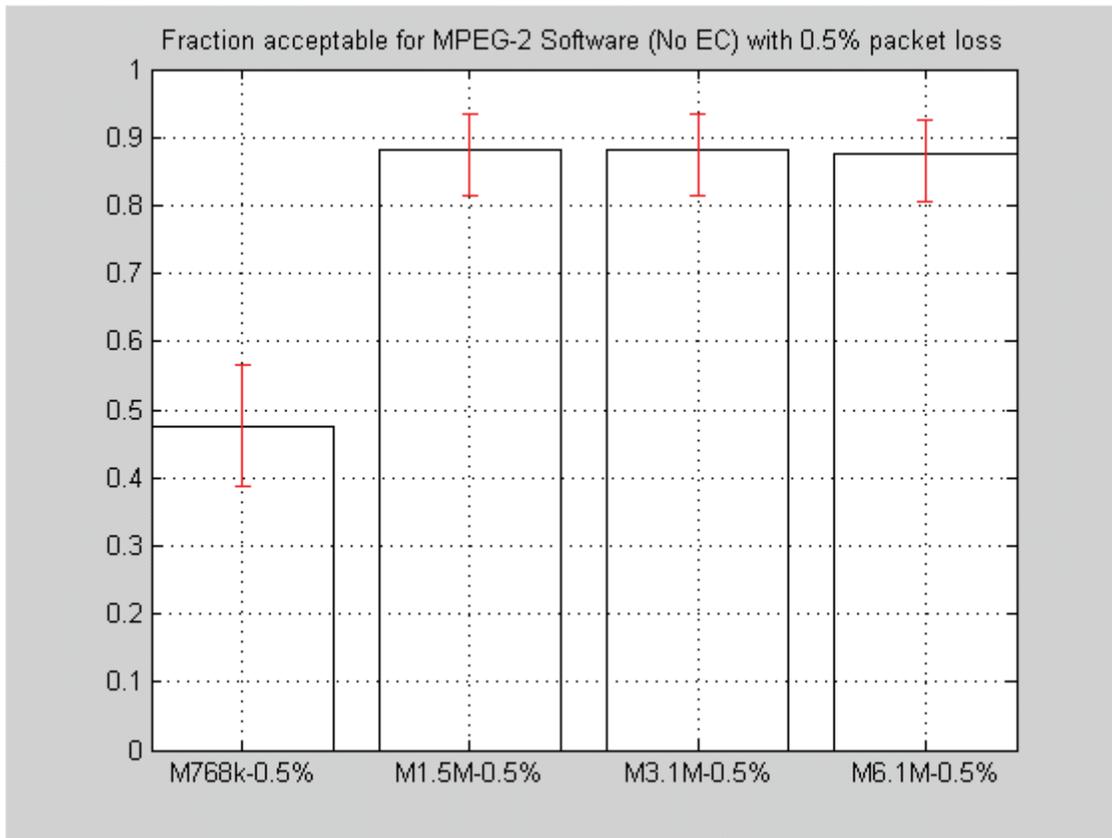
Figure 47:   MPEG-2 Software HRCs with 1.5 Percent Packet Loss (No EC)
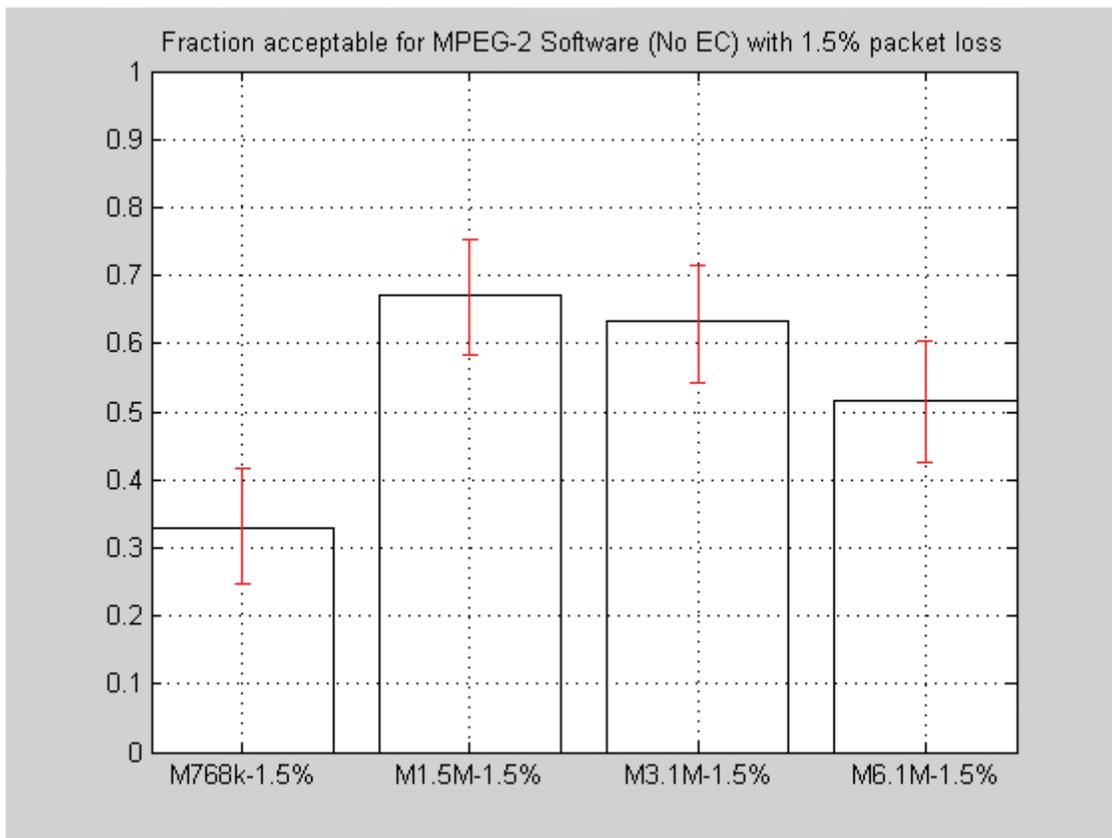
Table 65 summarizes the graphical results for the MPEG-2 software HRCs presented in Figure 44 through Figure 47. The table gives a mean value of the fraction acceptable for a given bit rate and packet loss ratio. Table 65 also gives the upper and lower bounds (indicated in right-side split cells) of the 95 percent confidence interval using a threshold of 0.7 (i.e., a given HRC is declared acceptable only if the lower bound of the confidence interval is greater than 0.7). Since the bottom of the 95% confidence bound must be greater than 0.7, the results indicate a requirement for an MPEG-2 coded bit rate of 1.5 Mbps or more, transmitted with a packet loss rate of 0.5 percent or less.

Table 65:   MPEG-2 Software HRCs Results Summary

Bit Rate	Fraction Acceptable (MPEG-2 Software HRCs) Packet Loss Ratio							
	0.0 Percent		0.1 Percent		0.5 Percent		1.5 Percent	
768 kbps	0.52	0.61	0.61	0.69	0.48	0.57	0.33	0.42
		0.43		0.52		0.39		0.25
1.5 Mbps	0.95	0.98	0.95	0.98	0.88	0.93	0.67	0.75
		0.90		0.90		0.81		0.58

Table 65:　MPEG-2 Software HRCs Results Summary (Continued)

	Fraction Acceptable (MPEG-2 Software HRCs) Packet Loss Ratio							
Bit Rate	**0.0 Percent**		**0.1 Percent**		**0.5 Percent**		**1.5 Percent**	
3.1 Mbps	0.97	0.99	0.99	1.00	0.88	0.93	0.63	0.72
		0.92		0.96		0.81		0.54
6.1 Mbps	0.97	0.99	0.94	0.97	0.88	0.93	0.52	0.60
		0.92		0.88		0.81		0.43

D.3.2　　H.264 Software HRCs

Figure 48 through Figure 51 and Table 66 give the results for the H.264 software HRCs. This H.264 software codec does not implement any error concealment (i.e., no EC), and no effort was made to spatially decorrelate errors due to packet loss (i.e., errors in the image may appear close to one another on the screen). However, this H.264 software codec represents one of the most advanced video coding technologies currently available.

Figure 48:　H.264 Software HRCs with 0 Percent Packet Loss (No EC)
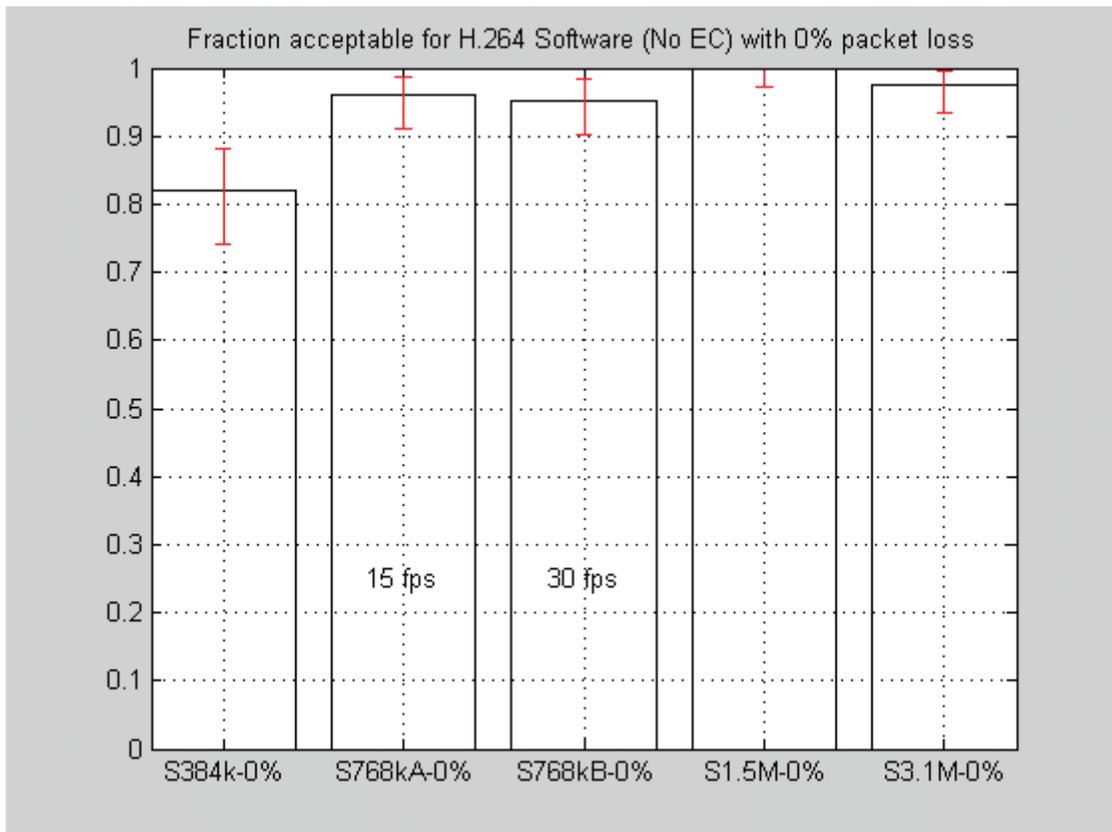
Figure 49: H.264 Software HRCs with 0.1 Percent Packet Loss (No EC)
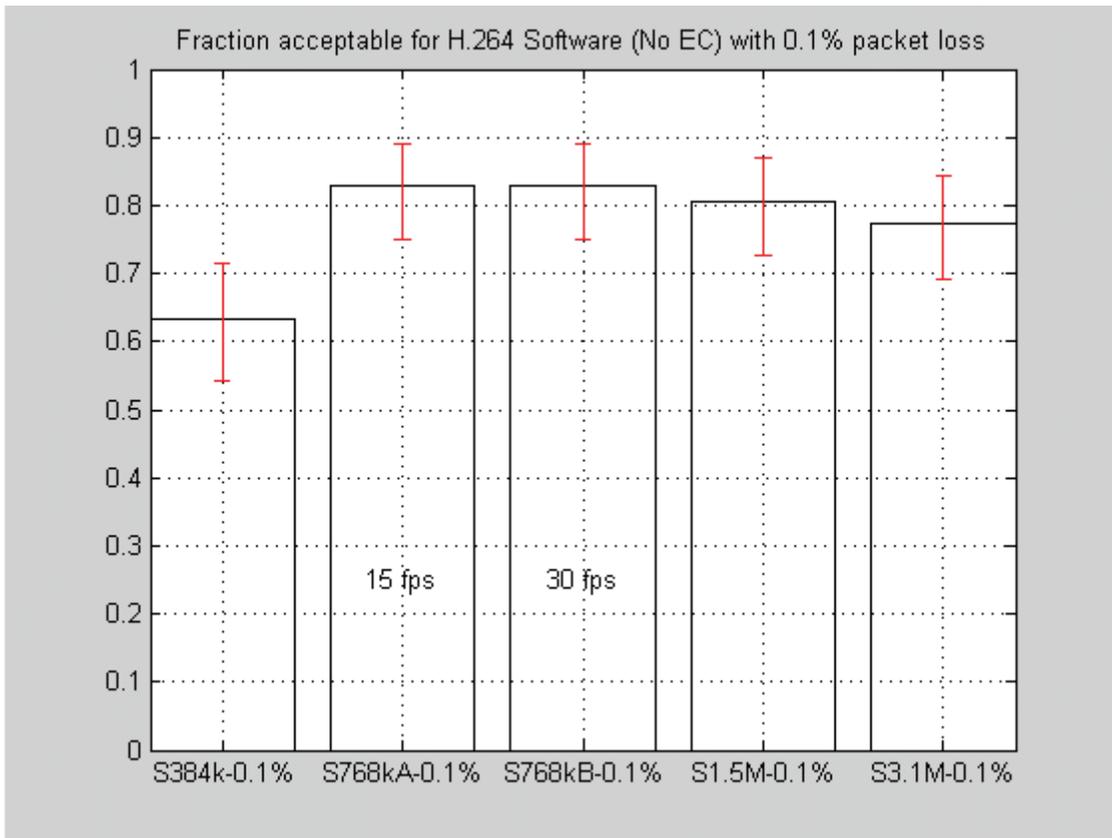
Figure 50:　H.264 Software HRCs with 0.5 Percent Packet Loss (No EC)
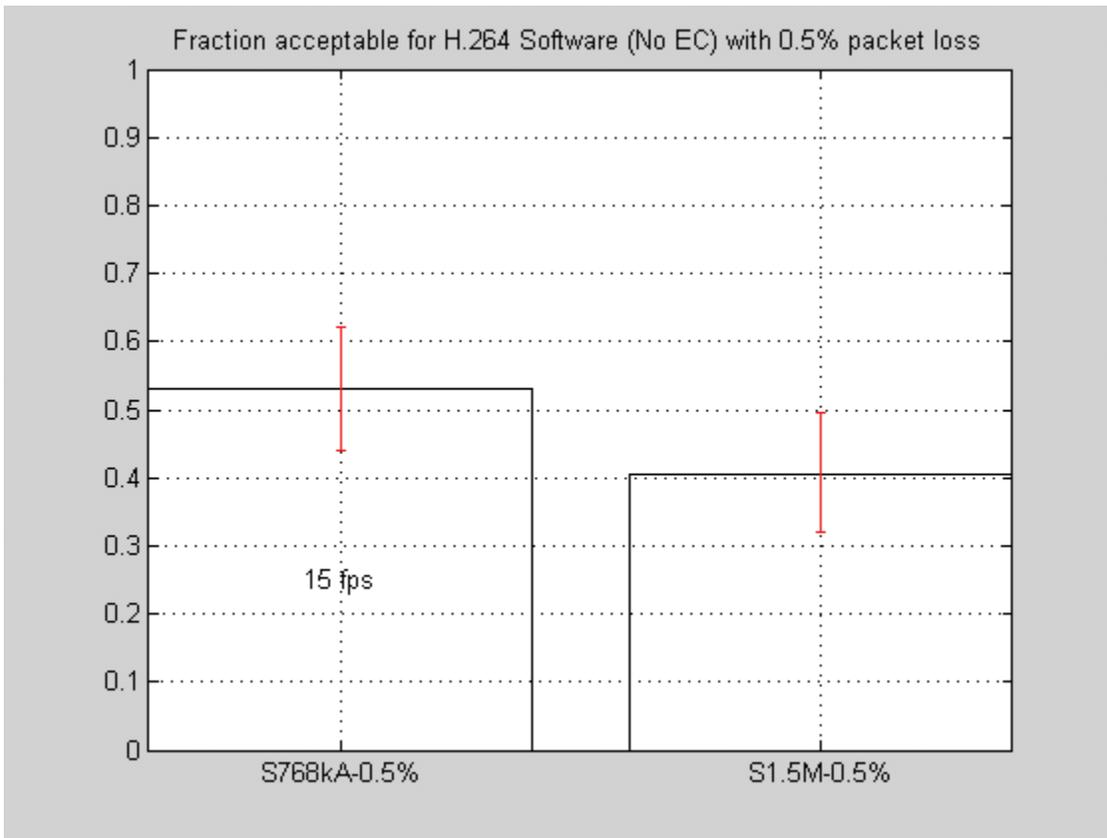
Figure 51:   H.264 Software HRCs with 1.5 Percent Packet Loss (No EC)
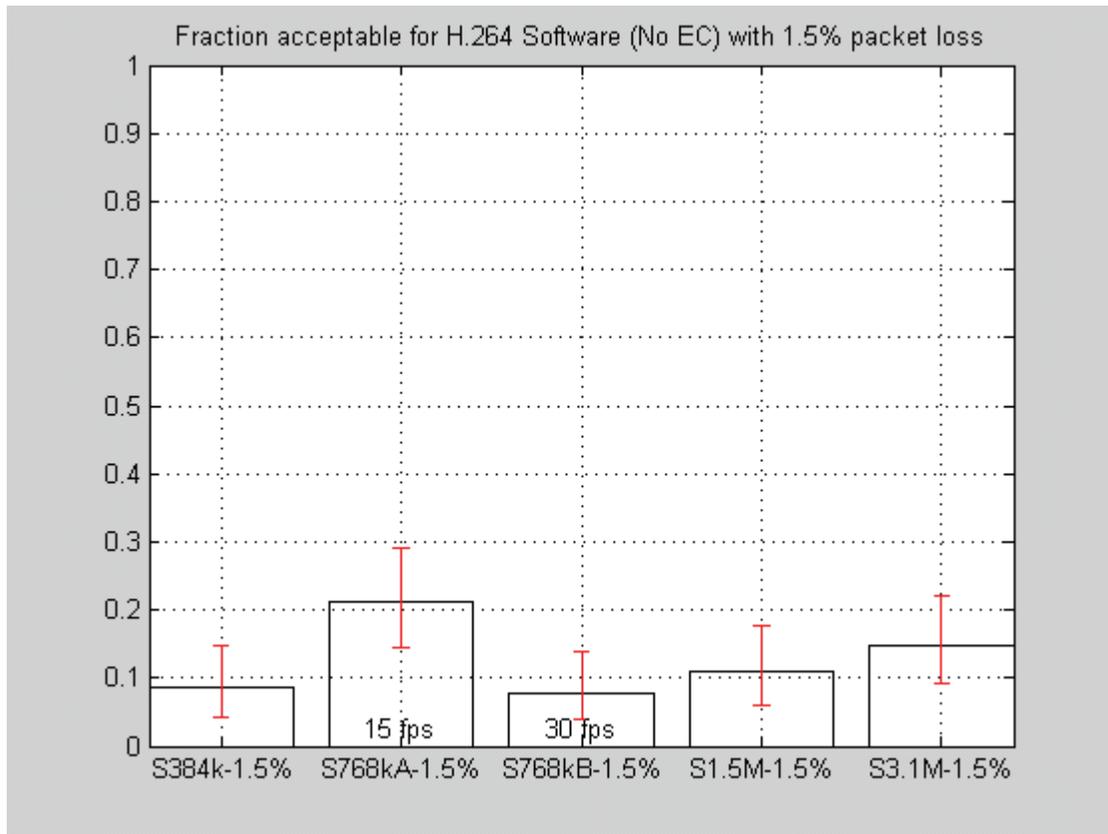
Table 66 summarizes the graphical results for the H.264 software HRCs presented in Figure 48 through Figure 51. The table gives a mean value of the fraction acceptable for a given bit rate and packet loss ratio. Table 66 also gives the upper and lower bounds (indicated in right-side split cells) of the 95 percent confidence interval using a threshold of 0.7 (i.e., a given HRC is declared acceptable only if the lower bound of the confidence interval is greater than 0.7). Since the bottom of the 95% confidence bound must be greater than 0.7, the results indicate that a coder bit rate of 384 kbps is acceptable provided the network has no packet loss; whereas coder bit rates of 768 kbps and higher are acceptable if the packet loss ratio is 0.1 percent or less.

Table 66:   H.264 Software HRCs Results Summary

Bit Rate Frame Rate	Fraction Acceptable (H.264 Software HRCs) Packet Loss Ratio							
	0.0 Percent		0.1 Percent		0.5 Percent		1.5 Percent	
384 kbps 15 fps	0.82	0.88	0.63	0.72	—	—	0.09	0.15
		0.74		0.54		—		0.04
768 kbps 15 fps	0.96	0.99	0.83	0.89	0.53	0.62	0.21	0.29
		0.91		0.75		0.44		0.14

Table 66:   H.264 Software HRCs Results Summary (Continued)

Bit Rate Frame Rate	Fraction Acceptable (H.264 Software HRCs) Packet Loss Ratio							
	0.0 Percent		**0.1 Percent**		**0.5 Percent**		**1.5 Percent**	
768 kbps 15 fps	0.95	0.98	0.83	0.89	—	—	0.08	0.14
		0.90		0.75		—		0.04
1.5 Mbps 30 fps	1.00	1.00	0.80	0.87	0.41	0.50	0.11	0.18
		0.97		0.73		0.32		0.06
3.1 Mbps 30 fps	0.98	1.00	0.77	0.84	—	—	0.15	0.22
		0.93		0.69		—		0.09

D.3.3   H.264 Hardware HRCs

Figure 52 and Table 67 give the results for the H.264 hardware HRCs. This H.264 hardware codec did implement error concealment (i.e., EC), so higher levels of packet loss ratio were considered.

Figure 52:   H.264 Hardware HRCs with Packet Loss (EC)
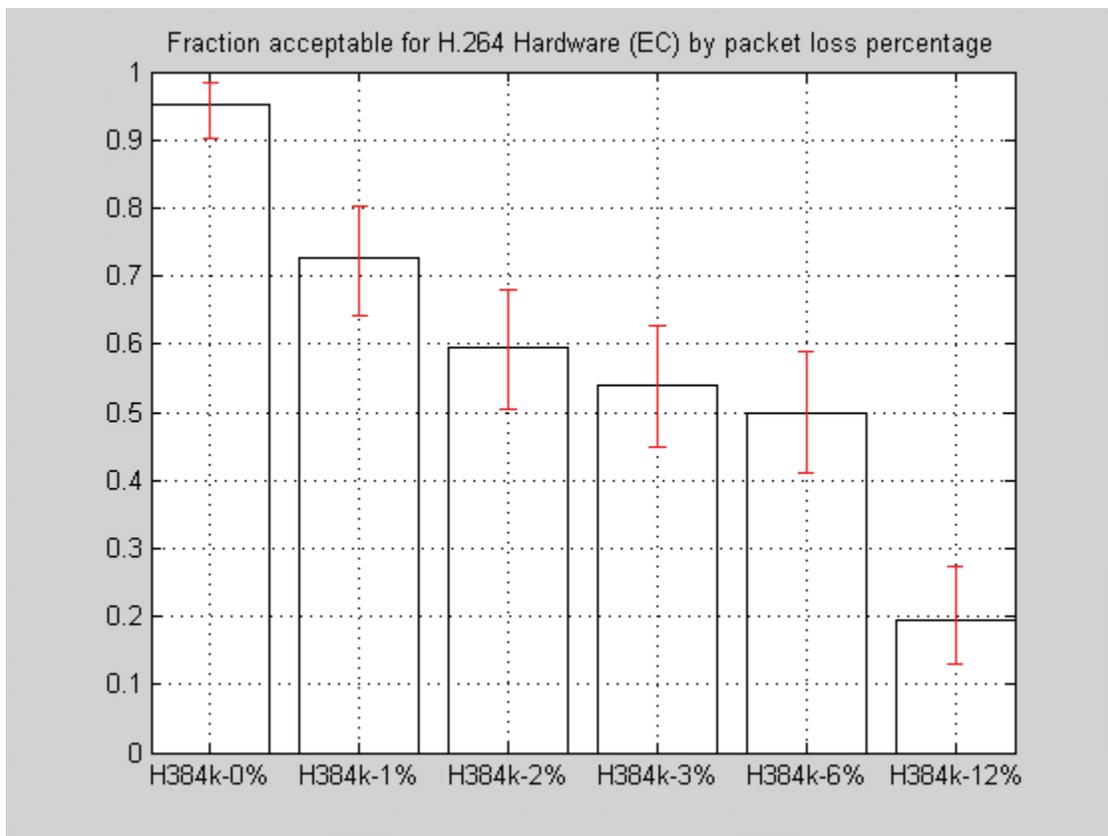
Table 67 summarizes the graphical results for the H.264 hardware HRCs presented in Figure 52. The table gives a mean value of the fraction acceptable for a given bit rate and packet loss ratio. Table 67 also gives the upper and lower bounds (indicated in right-side split cells) of the 95 percent confidence interval using a threshold of 0.7 (i.e., a given HRC is declared acceptable only if the lower bound of the confidence interval is greater than 0.7). Note that the fraction acceptable decreases much more gracefully with increasing packet loss ratio versus the H.264 software HRCs that did not implement any error concealment. Since the bottom of the 95% confidence bound must be greater than 0.7, the results suggest that the packet loss ratios must be held below about 1 percent. Further study using higher coder bit rates and error concealment schemes are required.

Table 67:   H.264 Hardware HRCs Results Summary

	Fraction Acceptable (H.264 Hardware HRCs) Packet Loss Ratio										
Bit Rate	0 Percent		1 Percent		2 Percent		3 Percent		6 Percent		12 Percent	
384 kbps	0.95	0.98	0.73	0.80	0.59	0.68	0.54	0.63	0.50	0.59	0.20	0.27
		0.90		0.64		0.50		0.45		0.41		0.13

D.3.4   Synthetic HRCs

Figure 53 and Table 68 give the results for the synthetic HRCs. These results can be used to establish values for fundamental image quality parameters such as frame rate and resolution.

Figure 53:   Synthetic HRCs for Frame Rate and Resolution
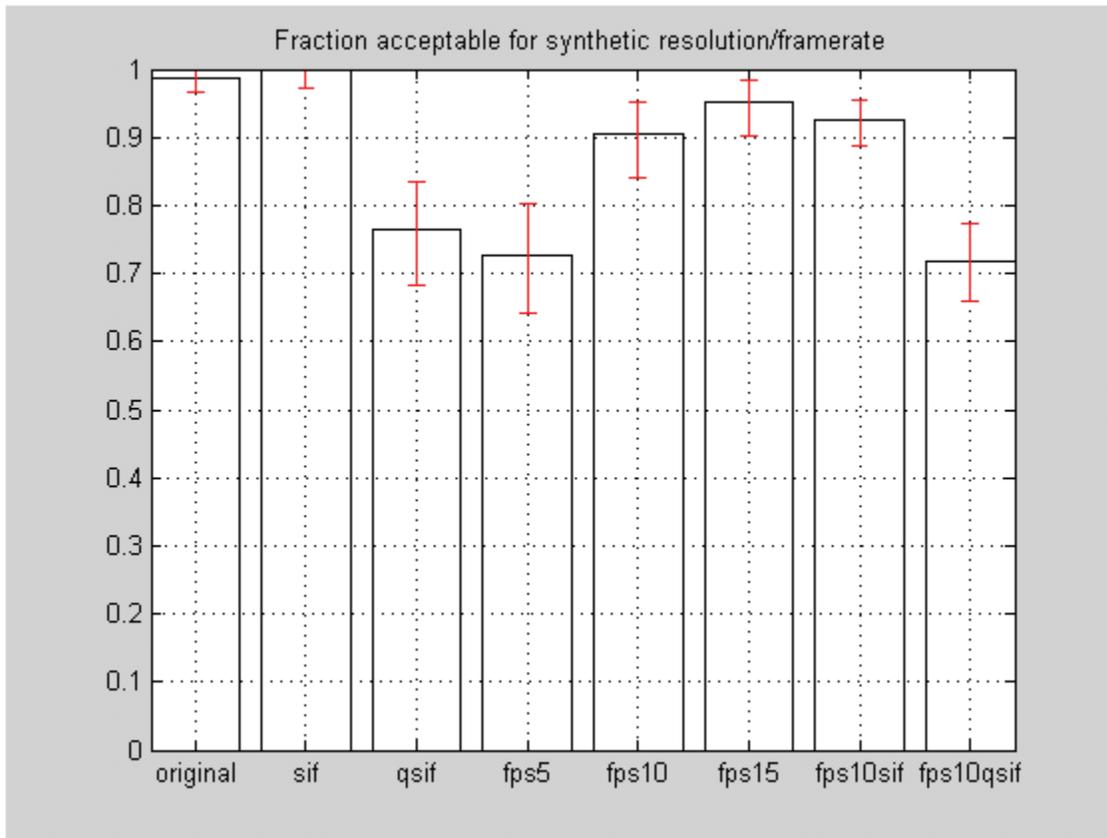
Table 68 summarizes the graphical results for the synthetic HRCs presented in Figure 53. The table gives a mean value of the fraction acceptable for each synthetic HRC. Table 68 also gives the upper and lower bounds (indicated in right-side split cells) of the 95 percent confidence interval using a threshold of 0.7 (i.e., a given HRC is declared acceptable only if the lower bound of the confidence interval is greater than 0.7). Since the bottom of the 95% confidence bound must be greater than 0.7, the results indicate that the frame rate should be at least 10 fps and the image resolution should be at least SIF.

Table 68:   Synthetic HRCs Summary

Fraction Acceptable (H.264 Hardware HRCs) Packet Loss Ratio															
Original		**SIF**		**QSIF**		**5 fps**		**10 fps**		**15 fps**		**10 fps SIF**		**10 fps QSIF**	
	1.00		1.00		0.84		0.80		0.95		0.98		0.95		0.77
0.99		1.00		0.77		0.73		0.91		0.95		0.93		0.72	
	0.97		0.97		0.68		0.64		0.84		0.90		0.89		0.66

D.3.5   Fraction Acceptable Versus Lossy Impairment Metric

Figure 54 gives a plot of Fraction Acceptable versus the Lossy Impairment metric (see Section 3.5.2.2 on page 23) for the 47 HRCs in the PS1 experiment. If an Acceptability Threshold (see Section 3.3.7 on page 18) of 0.7 is used, that is, the pink horizontal line in the plot, the 47 HRCs can be categorized as either

acceptable (the red circles) or unacceptable (the blue asterisks). Please recall that the Acceptability Threshold is applied to the lower bound of the 95 percent confidence interval. This is so that an HRC can have a Fraction Acceptable that is greater than 0.7, but still not be acceptable because the lower bound of the 95 percent confidence falls below 0.7.

The vertical green line represents a measured Lossy Impairment value of 0.41. HRCs that fall to the left of the green line would be acceptable, whereas HRCs that fall to the right of the green line would be unacceptable.

Figure 54:   Fraction Acceptable Versus the Lossy Impairment Metric
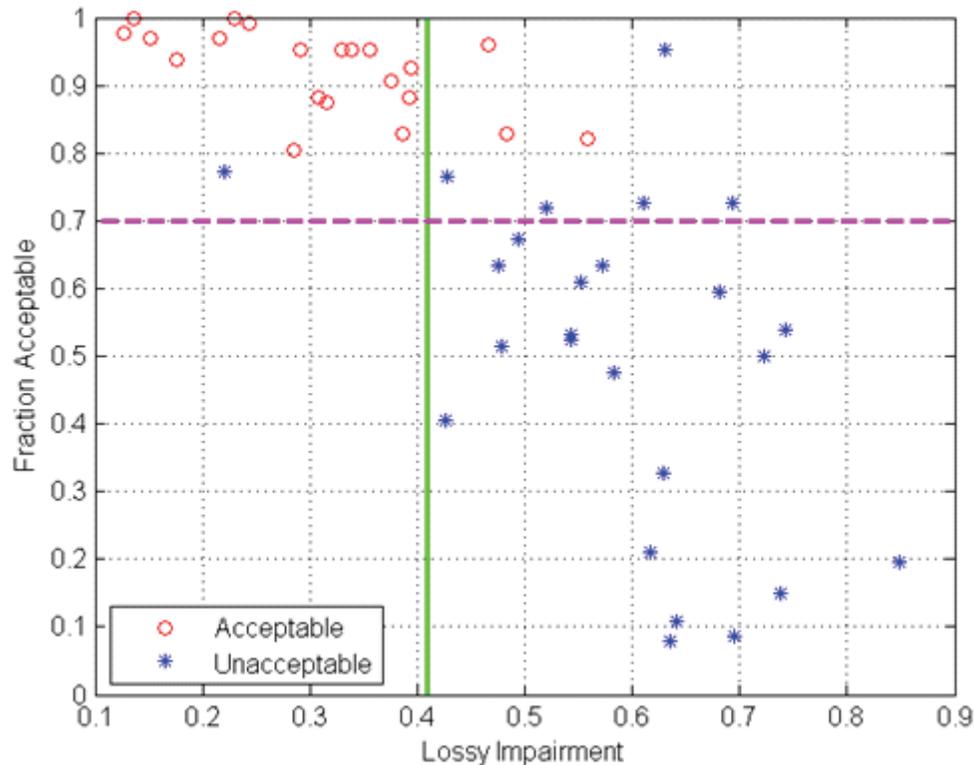
The Lossy Impairment metric misclassifies four of the 47 HRCs. The first of these errors is S3.1M-0.1 percent located at (Lossy Impairment = 0.22, Fraction Acceptable = 0.77). The Fraction Acceptable indicates this HRC as being unacceptable, while the Lossy Impairment metric indicates it being acceptable. The lower confidence bound for HRC S3.1M-0.1 percent is 0.69, which is just below the 0.7 acceptability threshold. Thus, S3.1M-0.1 percent was almost categorized as acceptable by subjective testing, and this particular error is benign. Generally speaking, however, this type of error is to be avoided, because errors of this sort would lead public safety practitioners to purchase inadequate equipment. Thus, the recommended value for the Lossy Impairment metric should be chosen conservatively (i.e., lower) to avoid this type of error.

For the other three errors, the Fraction Acceptable indicates acceptable, while the Lossy Impairment metric indicates unacceptable. The first is S384k-0 percent located at (0.56, 0.82), where the lower confidence bound was 0.74. The second is S768kA-0 percent located at (0.47, 0.96), and the lower confidence bound was 0.91. The third is S768kA-0.1 percent located at (0.48, 0.83), and the lower confidence bound was 0.75. Thus, two of these errors involve a lower confidence bound that is fairly close to the 0.7 acceptability threshold. Errors of this sort will cause public safety practitioners to avoid purchasing equipment that may

have adequate video quality for their application. These errors are considered less problematic than the previous type of errors.

As a note, the Lossy Impairment metric has been approved by ANSI for general use in objectively quantifying the perceptual impact of errors due to video codecs. That is why this particular metric has been placed in Section 3.5.2 on page 22. The Lossy Impairment metric has not been generally approved for use when network transmission errors (like packet loss) are present. However, the Lossy Impairment metric is currently one of the best available metrics for objectively measuring the perceptual video quality of digital video systems and the results presented above provide some support that the metric could have value in more general settings (e.g., as in the presence of network transmission errors).

D.4   Conclusions

The results presented in this appendix suggest that SIF image resolution should be the minimum specification for narrow field of view tactical video applications, particularly if any coding or transmission loss is present. The minimum frame rate should be at least 10 fps. The choice of codec is dependant upon available network capacities.

If capacities of 1.5 Mbps or higher are available, then an MPEG-2 codec (with no error concealment) could be chosen, provided the packet loss ratio is kept to 0.5 percent. For networks with lower capacities (768 kbps), an H.264 codec is the better choice as long as the packet loss ratio can be held to 0.1 percent or less (for H.264 codecs with no error concealment). H.264 systems operating at 384 kbps were acceptable for a packet loss ratio of 0 percent (i.e., no packet loss).

Preliminary results from this experiment indicate that acceptable packet loss ratios for H.264 codecs that perform error concealment are on the order of 1 percent, but further study is required in this area. Since error concealment algorithms are constantly being improved, these further studies should include state-of-the-art error concealment algorithms.

The proposed value for the Lossy Impairment metric (0.41) produces a fairly clean classification of the acceptable and unacceptable HRCs for an Acceptability Threshold of 0.7.

Results from the aforementioned questionnaire indicate that the one-way video delay for narrow field of view tactical video systems be kept to a maximum of 1 second.

Appendix E   Network Measurement Methods

This appendix presents the results obtained using the network model described in Section 5.3.12, which describes example speech and video applications that could be carried over the network. The graphs in this appendix plot one of the following in the network:

- Packet loss probability versus the number of PSCDs sharing the network

- Expected delay versus the number of PSCDs sharing the access network

Each graph assumes a particular value for the channel data rate and, in the case of the video application, a particular source compression algorithm (ITU-T H.264 or MPEG-2). Loss curves include results for Slotted Aloha but not TDMA. The TDMA worst-case loss probability is zero if the offered load is less than the channel data rate, and 1 if it is greater; the delay graphs show this cutoff point. When a graph (i.e., paths C, D, F, and G—see Figure 7 on page 33) includes curves with FRVs, the x-axis contains the number of PSCDs in IAN (i.e., per FRV). For example, a curve labeled FRV=20 at a point equal to 10 for the number of PSCDs in IAN means that there are 10 PSCDs for each of the 20 FRVs. Thus these 20 FRVs are generating to the JAN (jurisdictional communication tower) the aggregation of 200 PSCDs. For paths that include a JAN, we plot an ensemble of curves; each curve is associated with a different value for the number of FRVs attempting to communicate on the JAN. Two sets of curves plot speech application performance: one set plots speech packets containing 80 PCM (pulse code modulation) samples, the other set plots 320 samples per packet. A legend to the right of each graph identifies the parameters associated with each plot curve.

E.1   Speech Application

This section graphs the performance of the example PCM speech application using G.711 encoding described in Section 5.3.12.1. We examine the application's performance as a function of the number of PSCDs and FRVs in the networks, as characterized by the loss and delay metrics described in Section 6, for each of the path types listed in Section 5. This section also describes the effect of the channel data rate on the loss and delay performance.

E.1.1   Path Comparison

This section plots, for path types A through G, packet loss probability and average delay vs. the number of PSCDs, including the source PSCD, that share the source PSCD's area network. When PSCDs, FRVs, and a JAN are on the path, the graphs in Figure 55 through Figure 68 plot performance as a function of the number of FRVs competing for access to the JAN tower.

You can make several observations from the graphs in this section (Figure 55 through Figure 68). First, note that the loss curves in Figure 55, Figure 57, and Figure 63 are identical, indicating that the loss performance for Paths A, B, and E is the same. This follows from the fact that each of these paths features two 11 Mbps links that use either TDMA or Slotted Aloha. The difference between Paths A, B, and E arises from the delay performance, which is plotted in Figure 56, Figure 58, and Figure 64, respectively. The access delay to the three paths are identical, but the link and node delays become progressively larger from Path A with its single FRV to Path E, which incorporates two JAN towers and an EAN. You can expect the towers to be more distant from the PSCDs than an FRV would be, and thus the links to them

would have greater propagation delays. The EAN adds processing delays at various routers and also features potentially long transit times between its internal nodes.

You can observe the same kind of relationship between Paths C and D, and between Paths F and G. By looking at Figure 59 and Figure 61 we see that the loss performance does not change if we add an EAN to Path C to get Path D. This is because the EAN uses dedicated resources, which eliminates packet loss due to contention. The addition of the EAN has a significant effect on the application's delay performance, however. Figure 60 and Figure 62 show an increase of about 50 ms in the end-to-end delay. Note that the number of PSCDs beyond which Pr{loss} = 1 and the end-to-end delay is infinite, and a plot of the links using TDMA (labeled as dedicated in the figures) is the same in Figure 60 and Figure 62; this is because the IAN bandwidth is much less than that of the EAN and it is the limiting factor.

Figure 65 and Figure 67 illustrate the loss performance for the asymmetric paths F and G. Path G, which is essentially Path F plus an EAN, exhibits the same performance as Path F, for the same reason that Path D's loss performance is identical to that of Path C. As was the case with Paths C and D, the delay for Path G is about 50 ms greater than the delay for Path F, for all combinations of number of FRVs, PSCDs, and packet sizes.

Figure 55:   Packet Loss for Speech on Path A at 11 Mbps
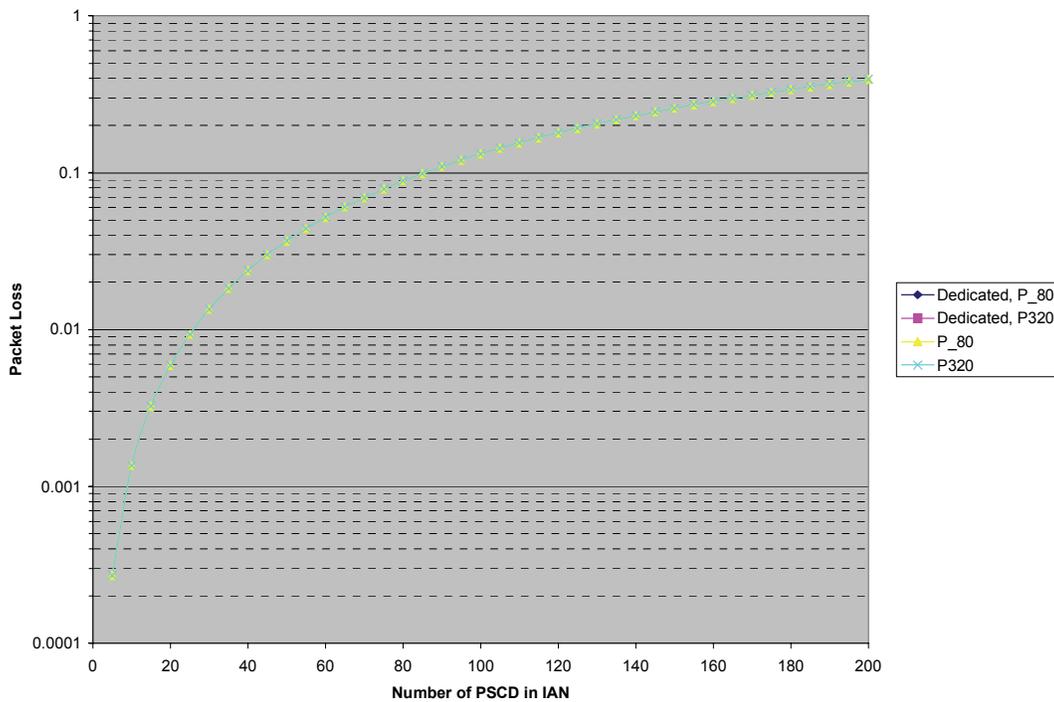
Figure 56:   Delay for Speech on Path A at 11 Mbps
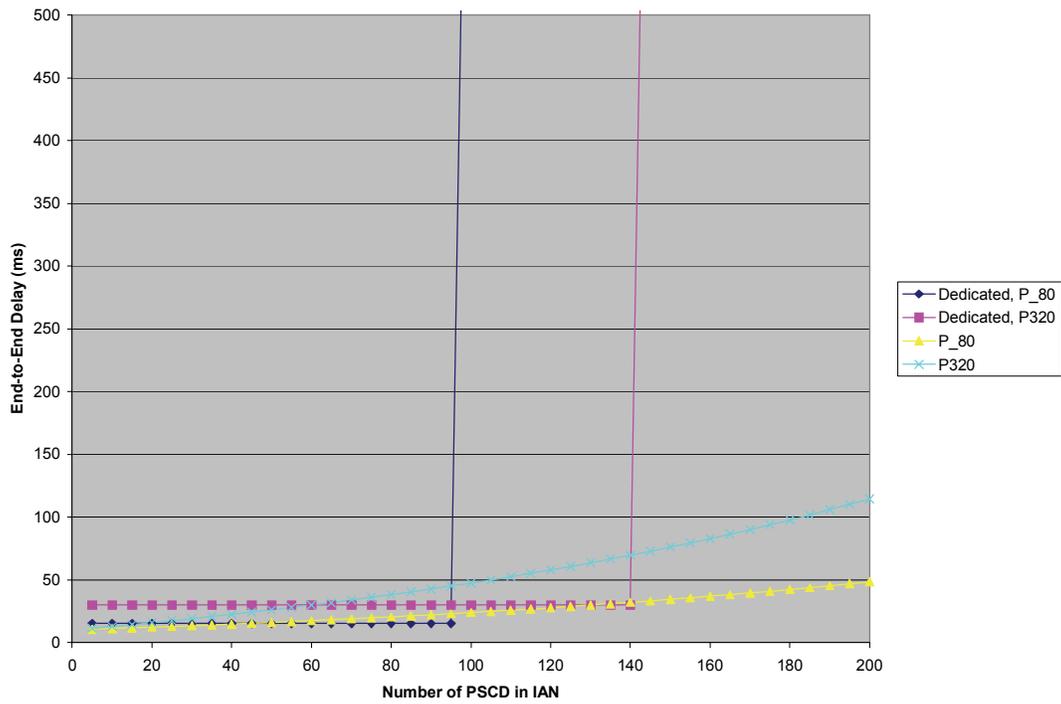
Figure 57:   Packet Loss for Speech on Path B at 11 Mbps
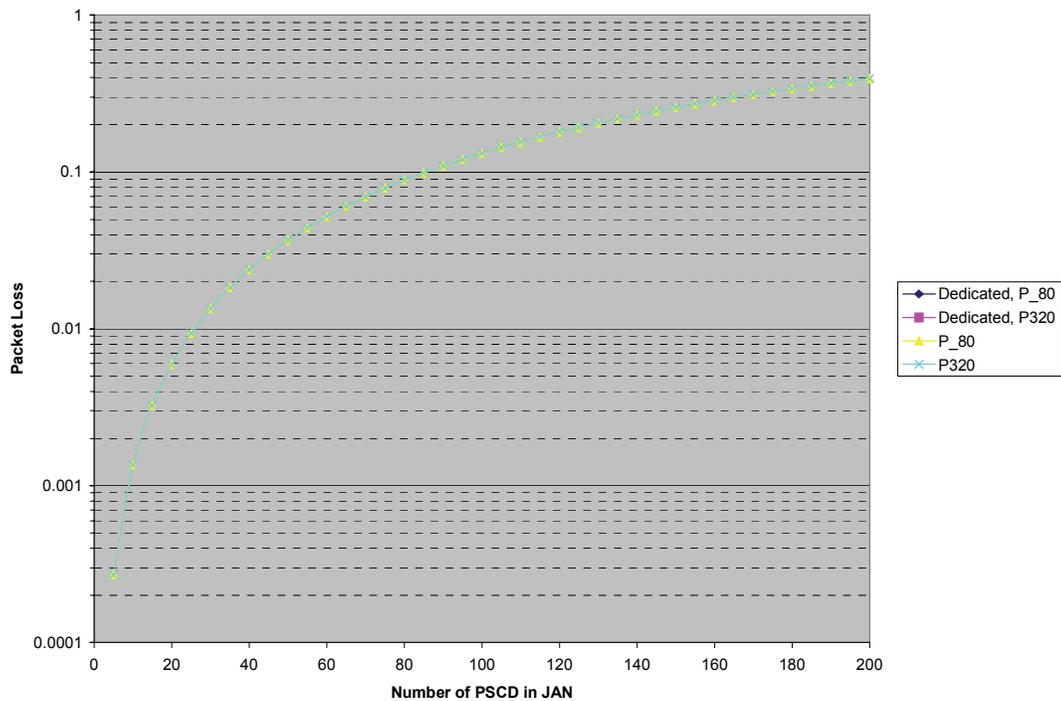
Figure 58:   Delay for Speech on Path B at 11 Mbps
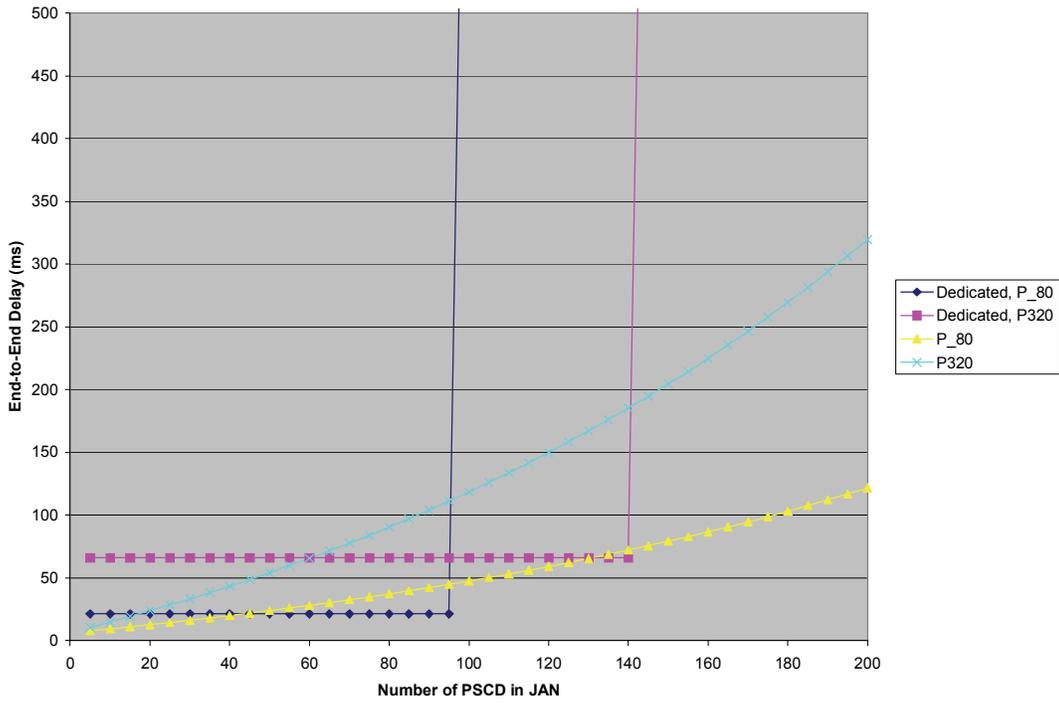
Figure 59:   Packet Loss for Speech on Path C at 11 Mbps
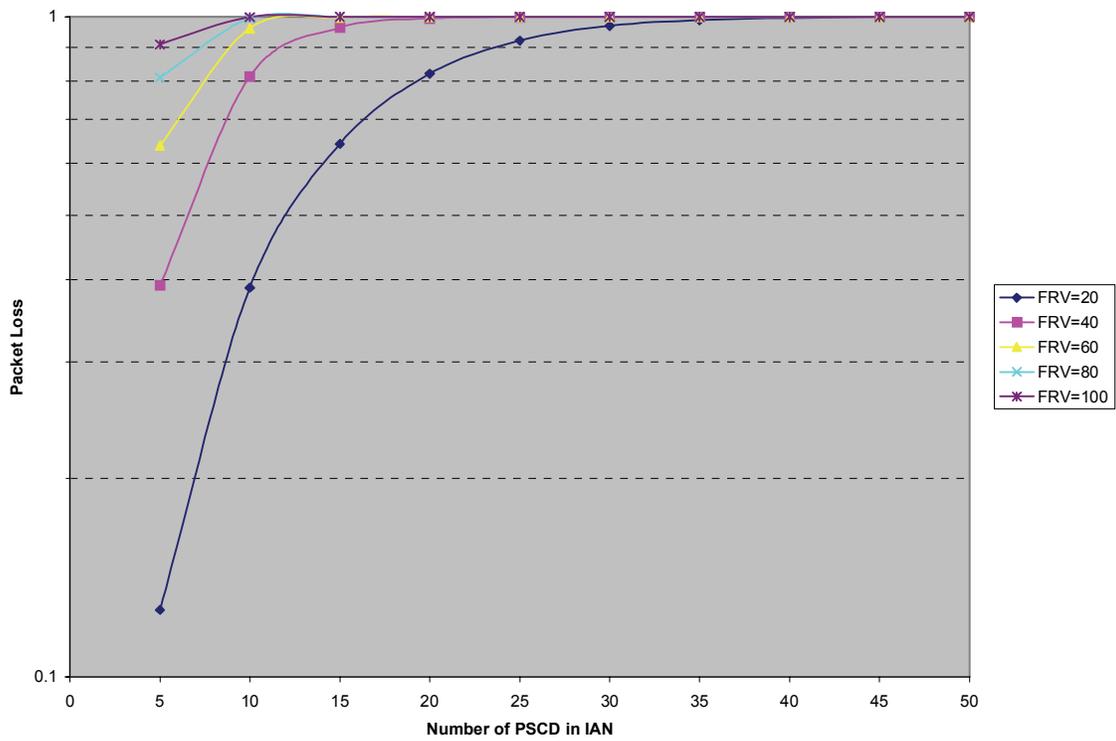
Figure 60:   Delay for Speech on Path C at 11 Mbps
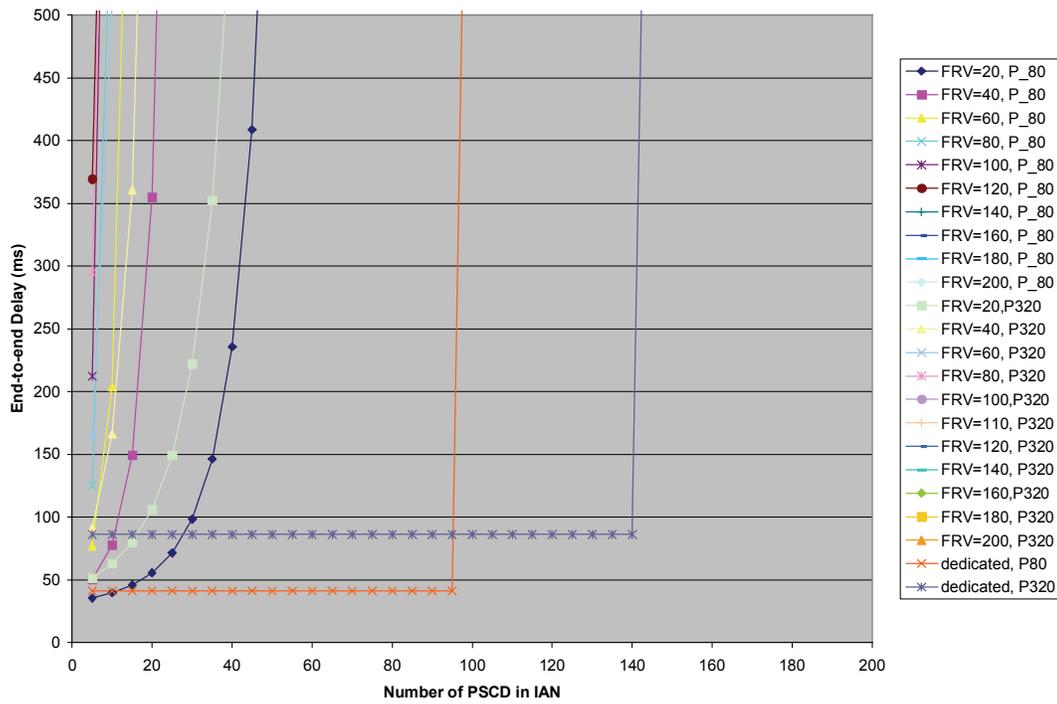
Figure 61:   Packet Loss for Speech on Path D at 11 Mbps
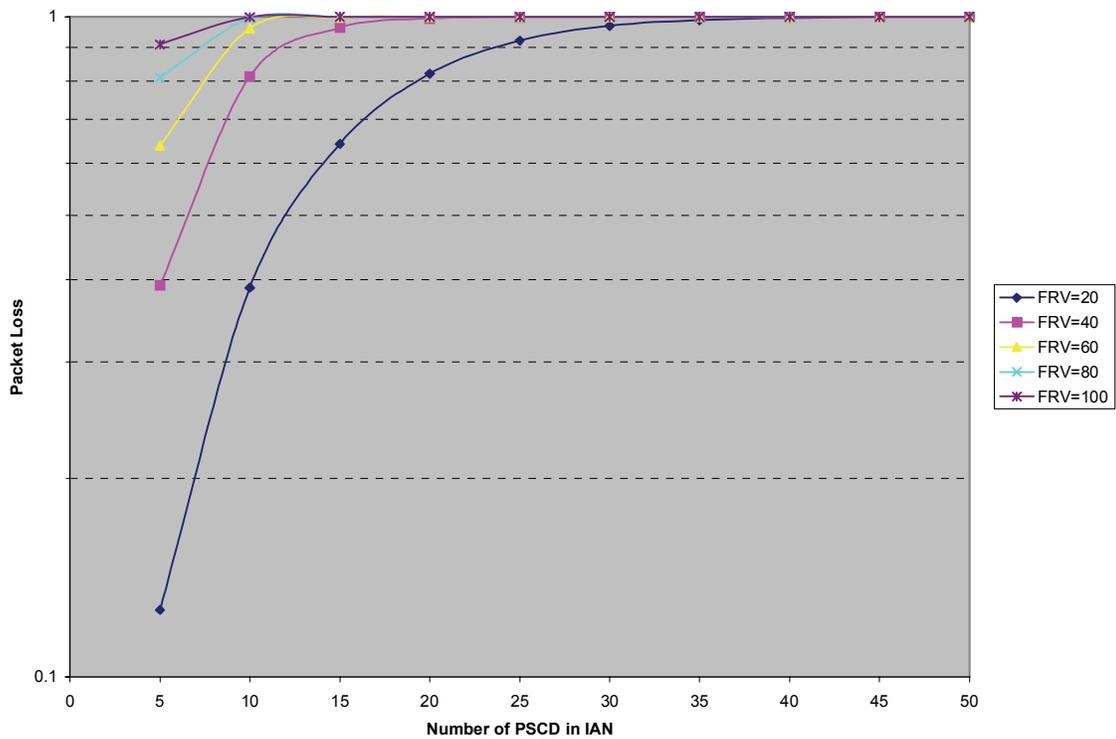
Figure 62:   Delay for Speech on Path D at 11 Mbps
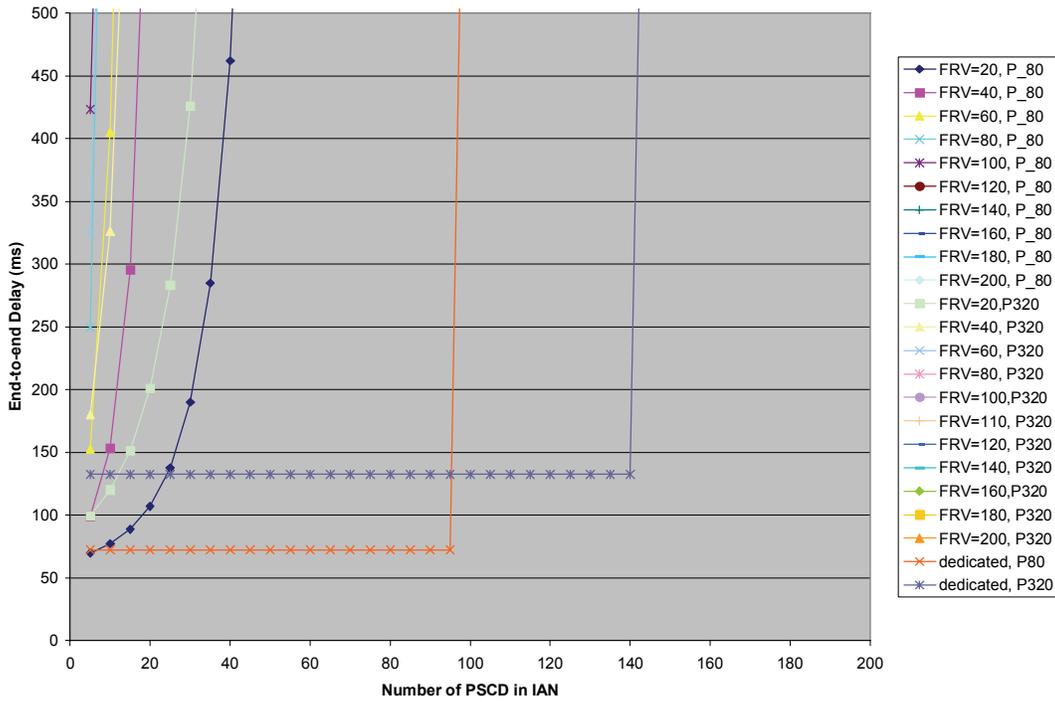
Figure 63:   Packet Loss for Speech on Path E at 11 Mbps
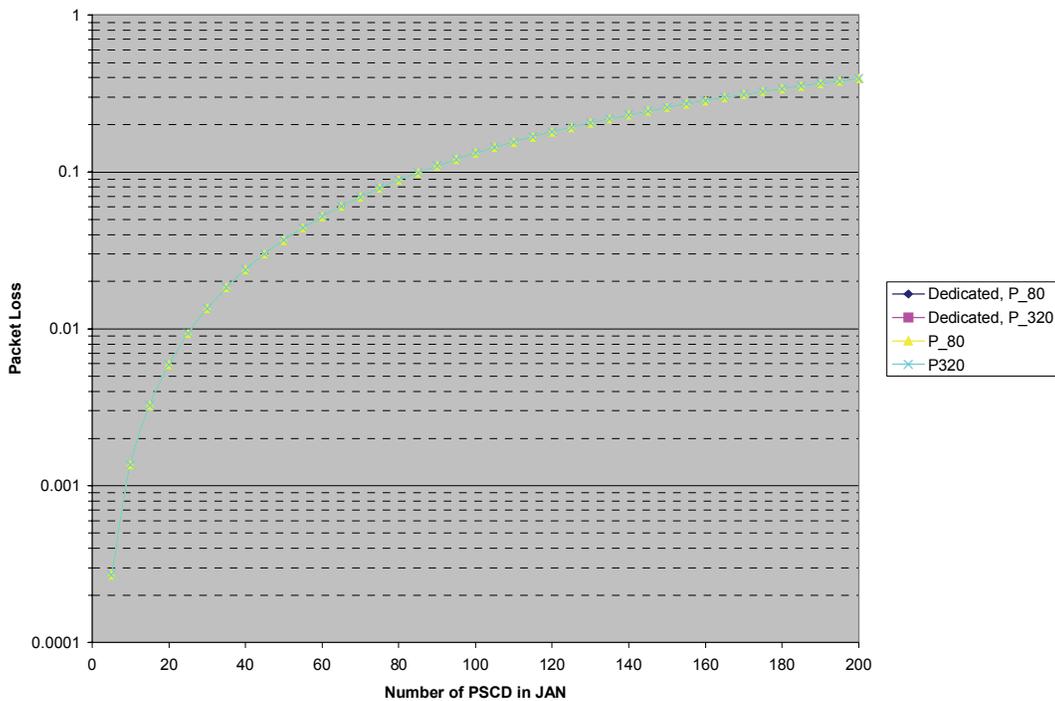
Figure 64:   Delay for Speech on Path E at 11 Mbps
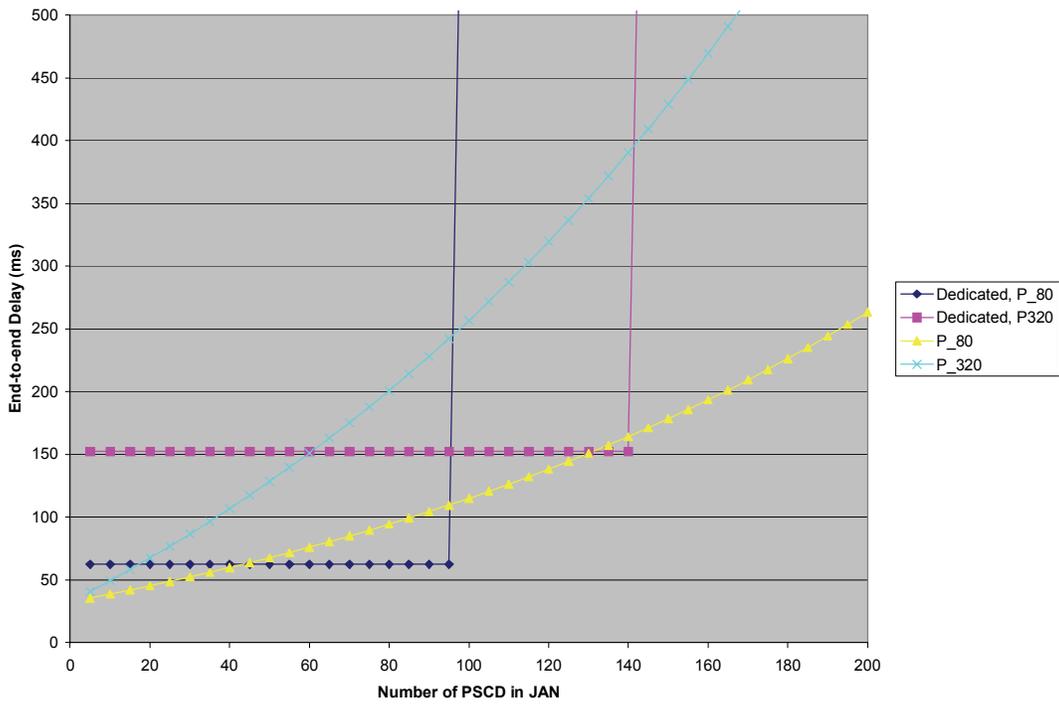
Figure 65:   Packet Loss for Speech on Path F at 11 Mbps
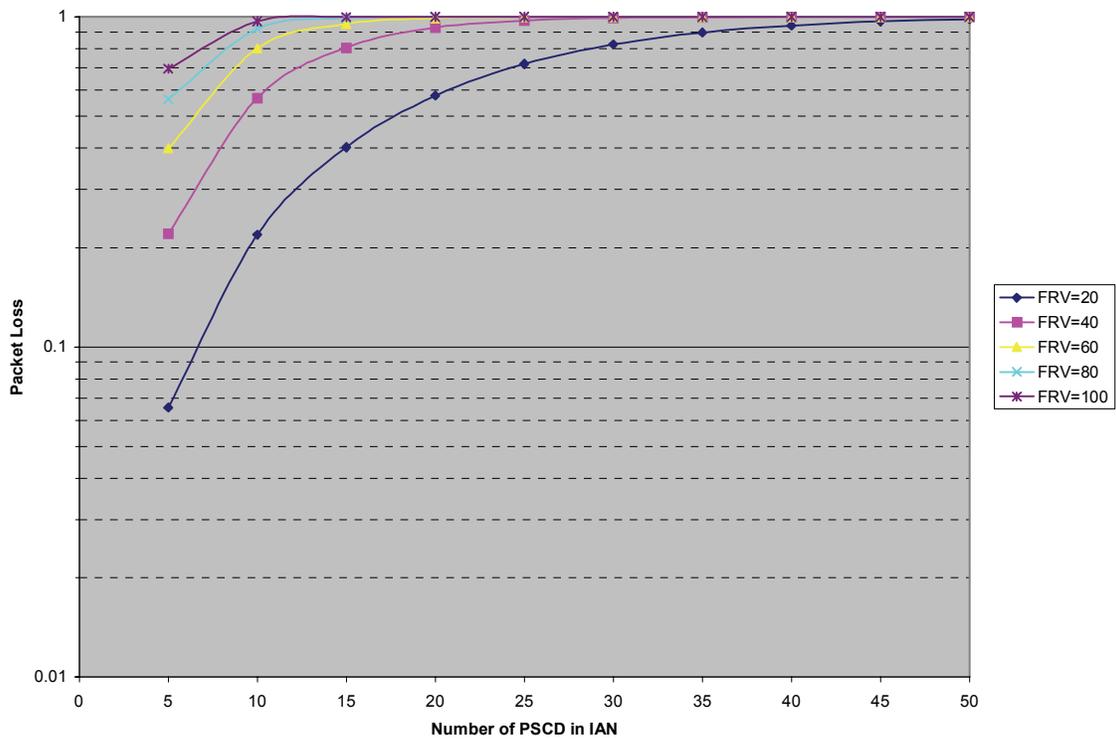
Figure 66:   Delay for Speech on Path F at 11 Mbps
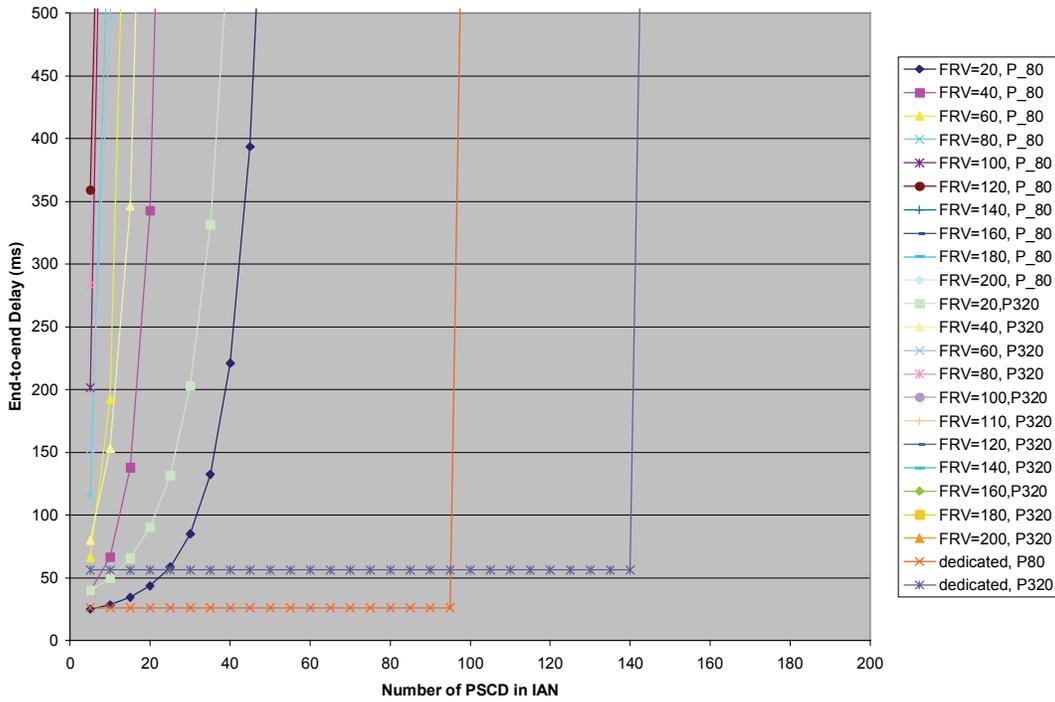
Figure 67:   Packet Loss for Speech on Path G at 11 Mbps
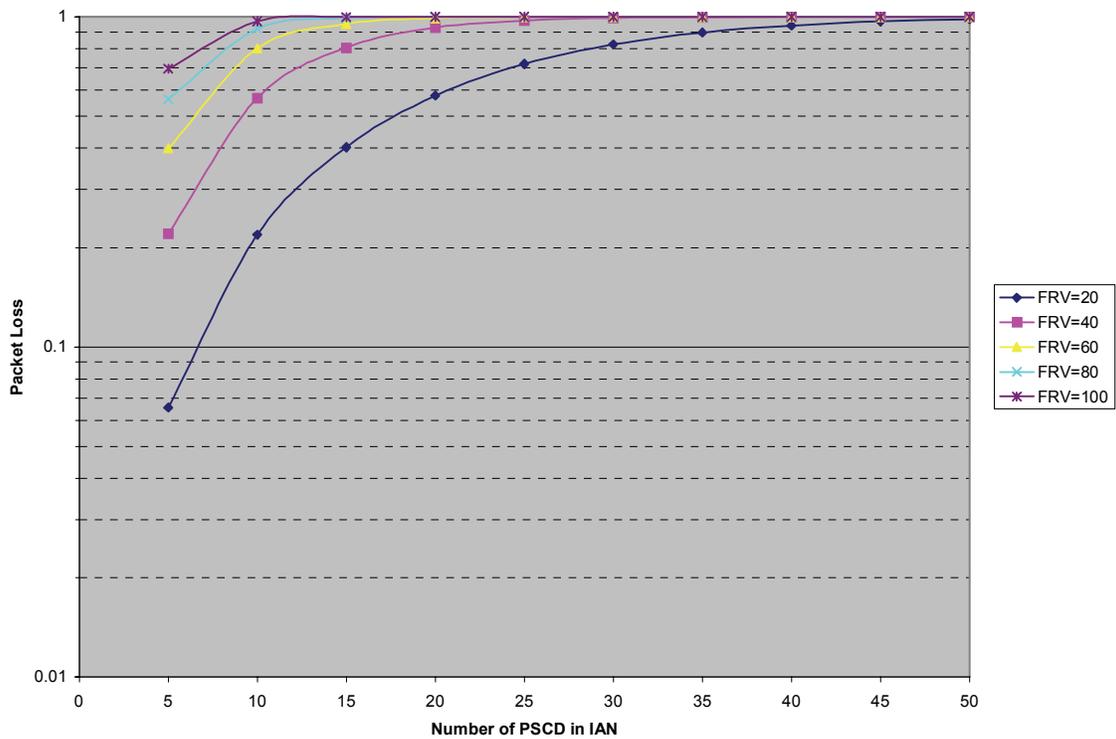
Figure 68:   Delay for Speech on Path G at 11 Mbps
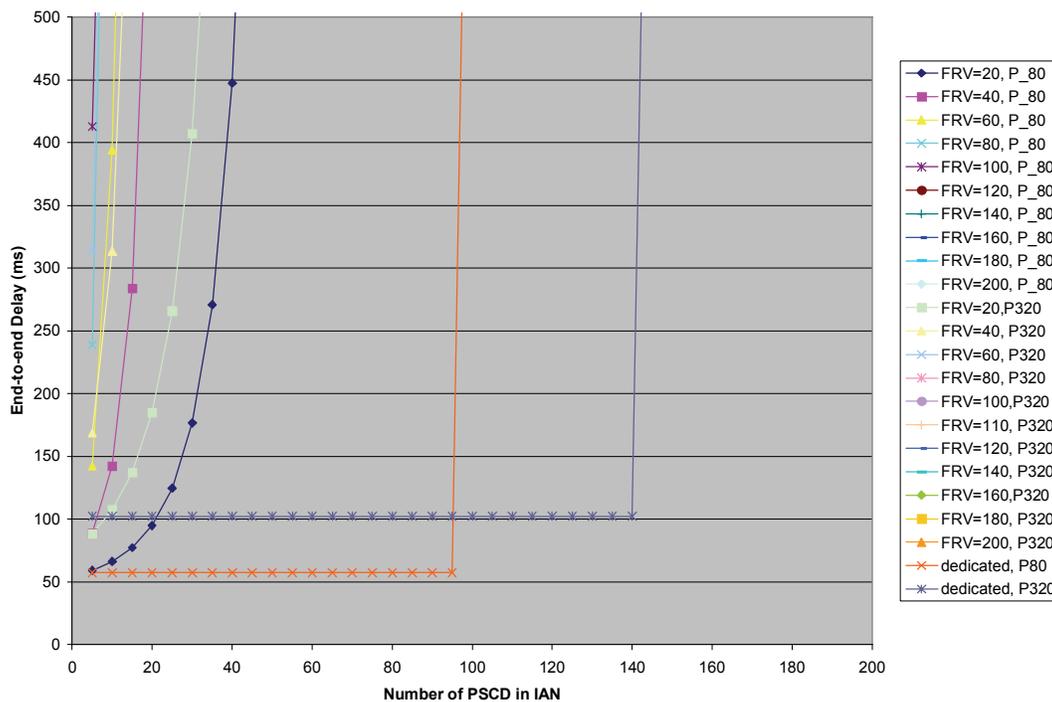
E.1.2   Effect of Channel Data Rate

This section describes the effect of increasing the channel data rate on the links traversed by the path. We choose 54 Mbps as an example data rate; it could potentially be used in the future network because it is part of the IEEE 802.11g high-speed wireless LAN standard. However, the network model does not depend on any particular value for the channel data rate.

Using Path C as a point of comparison, we increased the data rate on the IAN-JAN and JAN-IAN links only, while keeping the data rate on the PAN-IAN and IAN-PAN links at 11 Mbps. Similar effects occur with the other paths (A, B, D, E, F, G). Comparing Figure 59 and Figure 69 illustrates that where the wireless links use Slotted Aloha, the loss performance improves due to the additional available bandwidth. If the network is lightly loaded (20 FRVs and 5 PSCDs), a loss rate of less than 1 percent is possible.

Similarly, Figure 60 and Figure 70 show the reduction in the end-to-end delay resulting from the nearly five-fold increase in the available bandwidth. For example, the curves associated with 20 FRVs and an 80-sample packet show that the maximum number of PSCDs the network can support while keeping the end-to-end delay less than 100 ms, rises from 30 when all links run at 11 Mbps, to nearly 160 when the data rate of the JAN is increased to 54 Mbps.

Figure 69:   Packet Loss for Speech on Path C at 11 Mbps on IAN and 54 Mbps on JAN
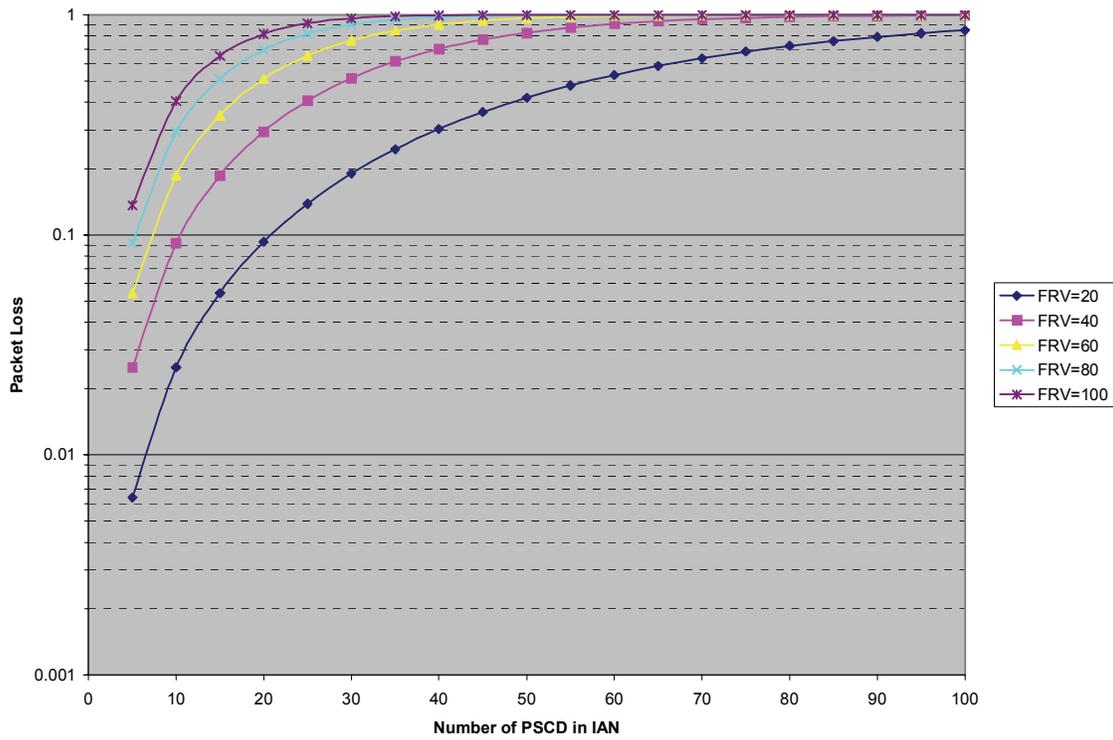
Figure 70:   Delay for Speech on Path C at 11 Mbps on IAN and 54 Mbps on JAN
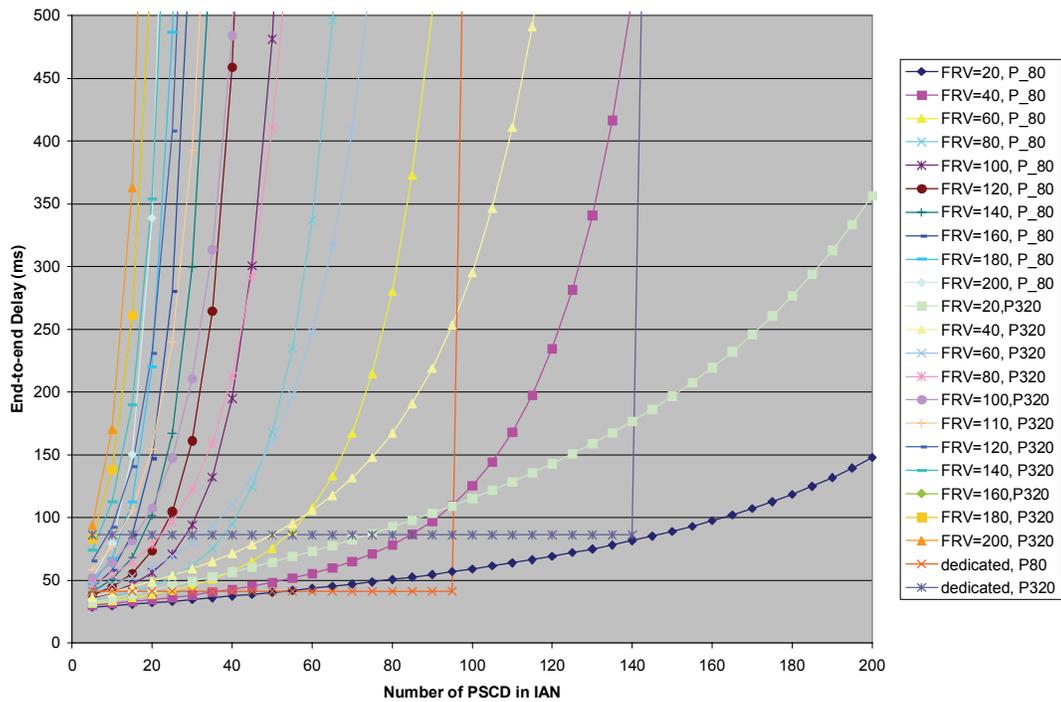
E.2   Video Application

This section graphs the example video application described in Section 5.3.12.2 to examine network behavior where the application has greater bandwidth requirements. Network performance depends on the type of video encoder and the channel data rate. As in Section E.1.2, this section compares network performance over Path C.

E.2.1   Effect of Coding Scheme

Graphs in this section plot the loss or delay performance of H.264 and MPEG-2 video coding where available bandwidth on every link of Path C is 11 Mbps. Figure 71 and Figure 72 show the packet loss and delay performance, respectively, of the H.264 encoder. When there are as few as 10 FRVs in the JAN, the loss performance is unacceptable for any number of PSCDs larger than or equal to 5 in the IAN. This shows the unsuitability of a simple contention resolution protocol like Slotted Aloha when the channel utilization is high. Delay performance is similarly poor. Steep rises in all the delay curves indicate that acceptable performance is possible only with dedicated links or with small numbers of PSCDs and FRVs.

Figure 71:   Packet Loss for H.264 Video on Path C at 11 Mbps
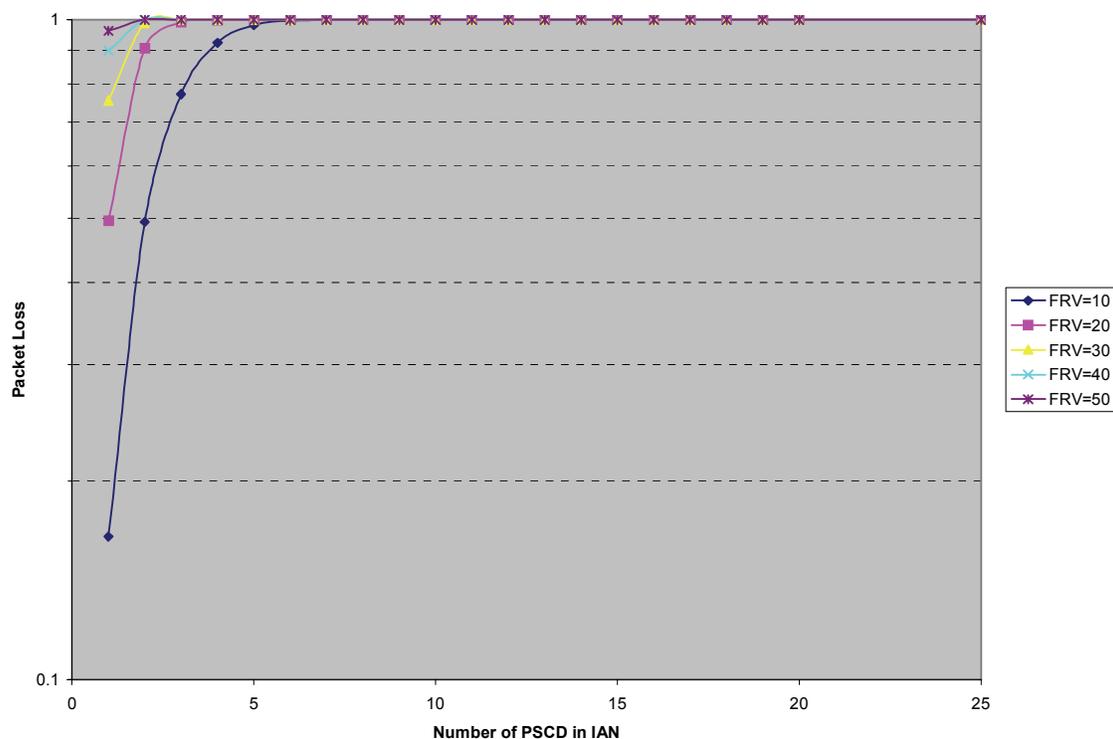
Figure 72: Delay for H.264 Video on Path C at 11 Mbps
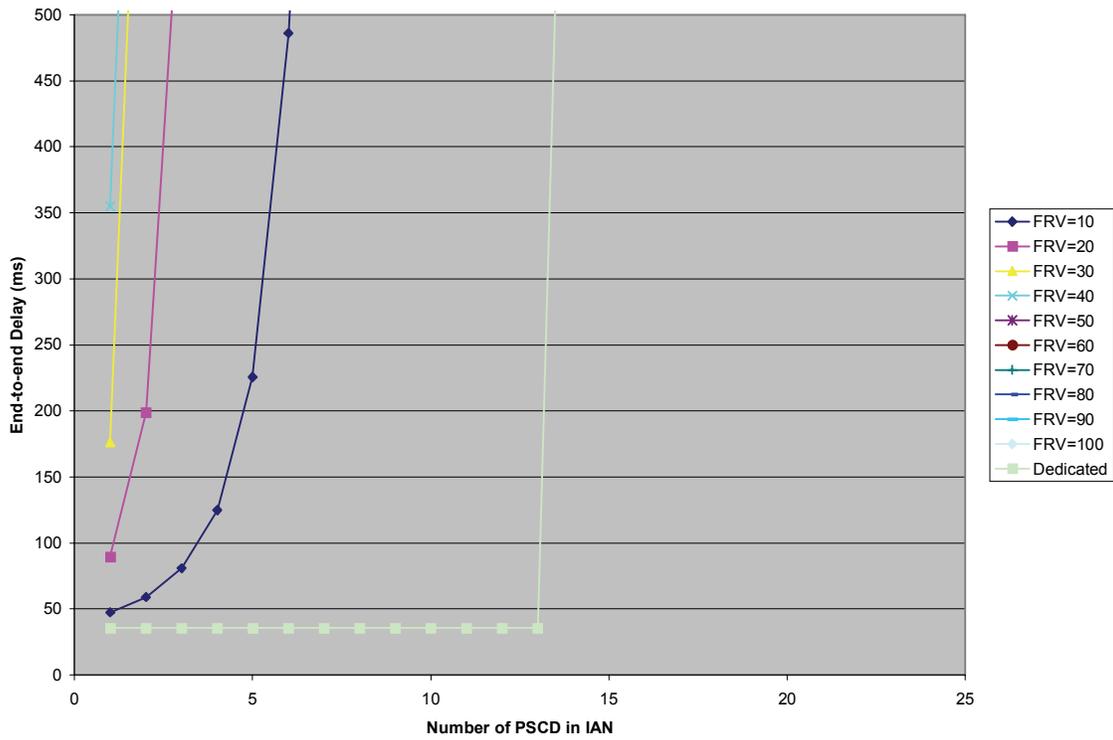
The MPEG-2 plots in Figure 73 (loss) and Figure 74 (delay) show that MPEG-2 loss and delay performance are worse than the performance of H.264 video because of the greater overhead associated with MPEG-2. For example, Figure 73 shows a packet loss probability of approximately 50 percent for the lowest load scenario we considered (5 PSCDs per IAN and 10 FRVs per JAN). That same scenario was the only one that yeilded an average end-to-end delay that was less than 200 ms, as Figure 74 shows.

Figure 73:   Packet Loss for MPEG-2 Video on Path C at 11 Mbps
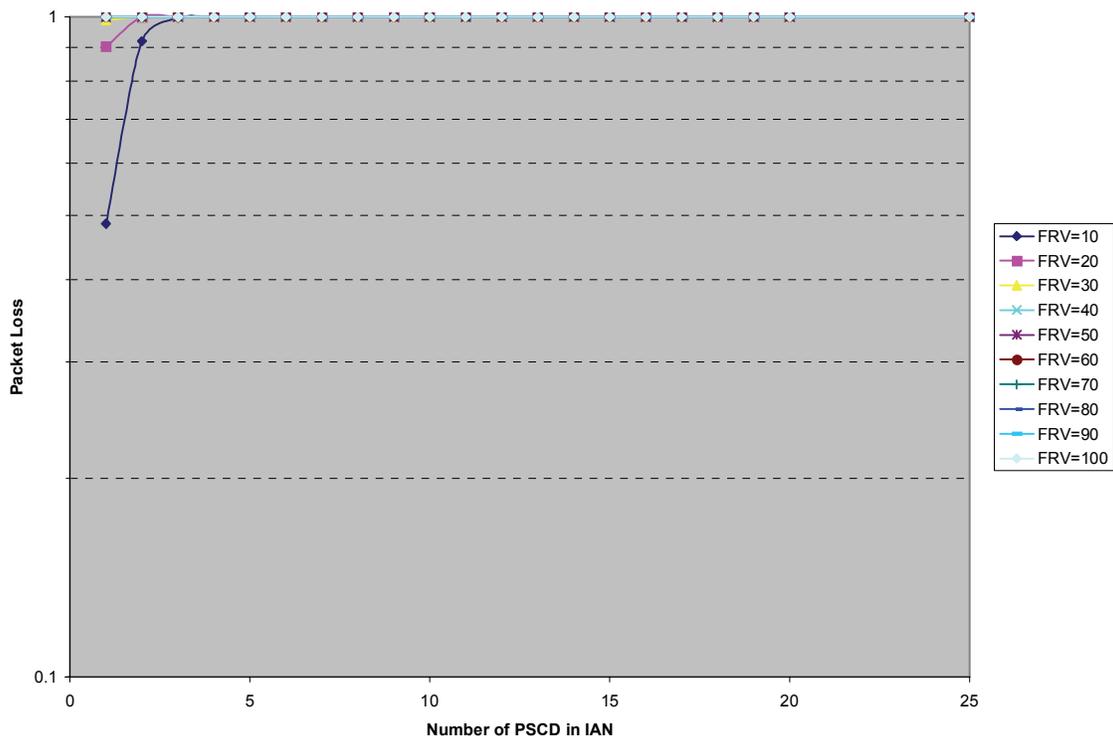
Figure 74:   Delay for MPEG-2 Video on Path C at 11 Mbps
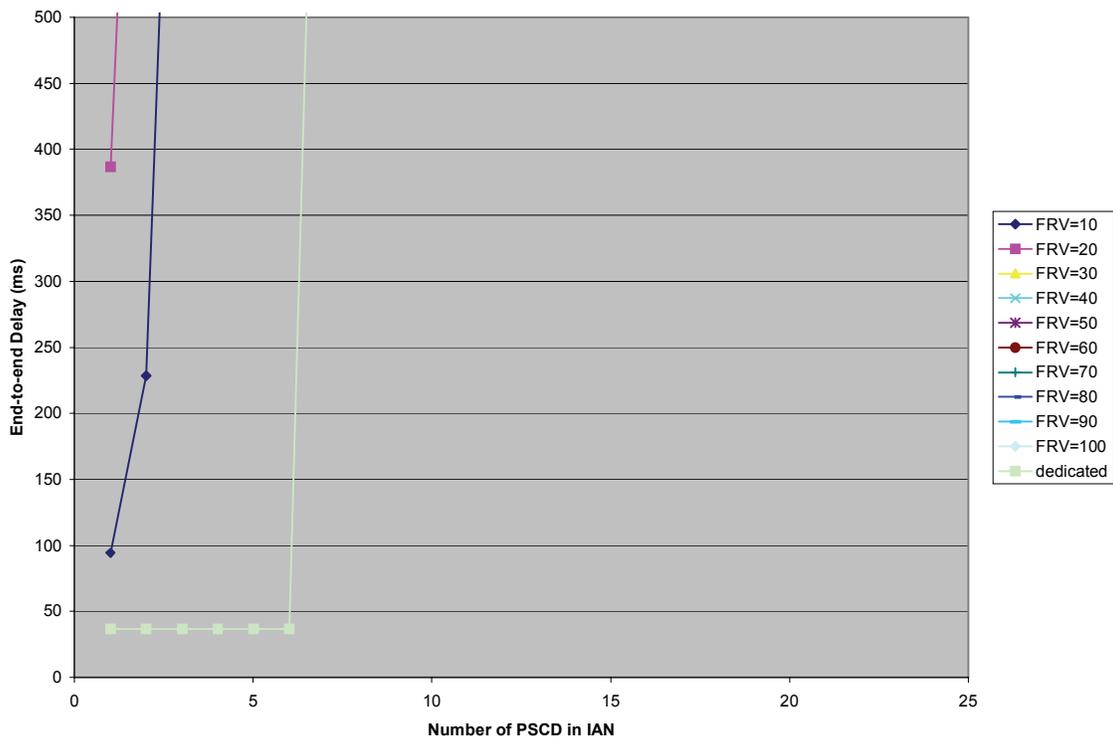
E.2.2   Effect of Channel Data Rate

Similar to Section E.1.2, this section describes how an increase in the channel data rate can produce significant improvements in the application performance. The H.264 plots in Figure 75 (loss) and Figure 76 (delay) show the performance for H.264 video over Path C where the data rate on all links has been increased to 54 Mbps.

Although many of the usage scenarios in the graphs illustrate unacceptable performance, reasonable performance results are still obtainable when the network is lightly loaded This means, (again, that the number of FRVs per IAN is 10, and the number of PSCDs per JAN is 5.) When the number of FRVS in the JAN is 10, it is possible to deploy nearly 20 PSCDs per IAN and still have a packet loss rate of less than 10 percent using Slotted Aloha. You can observe a similar degree of improvement in the delay performance, such that end-to-end delays of less than 100 ms are possible with 20 FRVs per JAN and 13 PSCDs per IAN.

Figure 75:   Packet Loss for H.264 Video on Path C at 54 Mbps
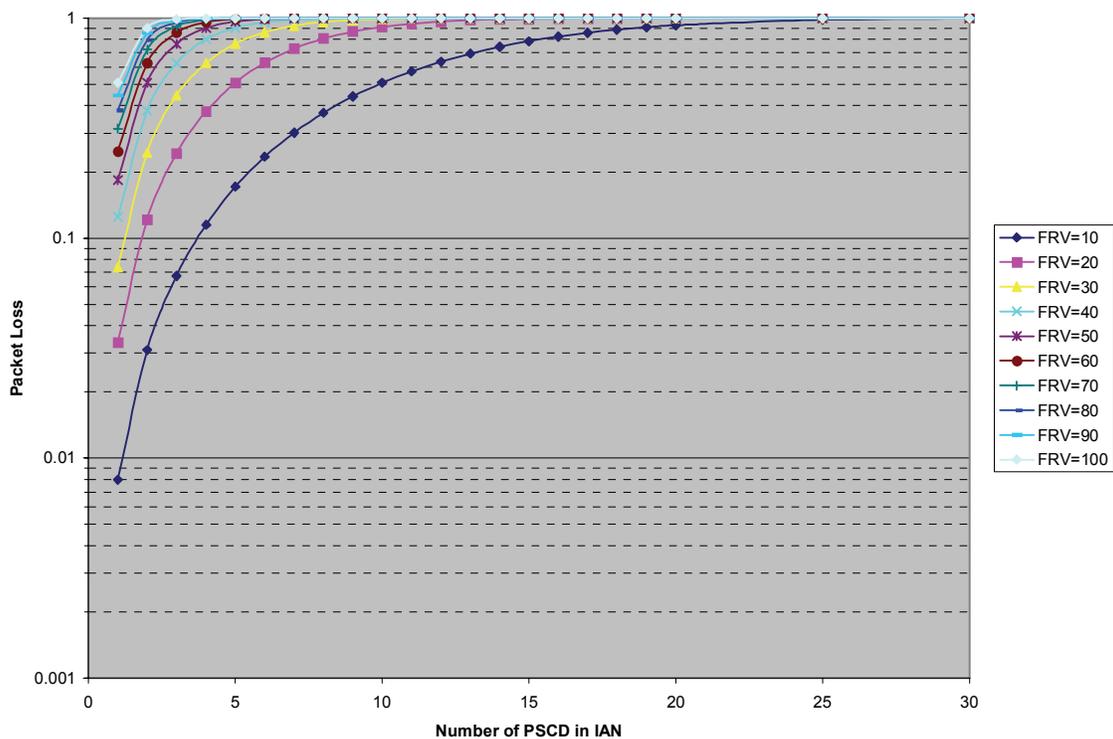
Figure 76:   Delay for H.264 Video on Path C at 54 Mbps
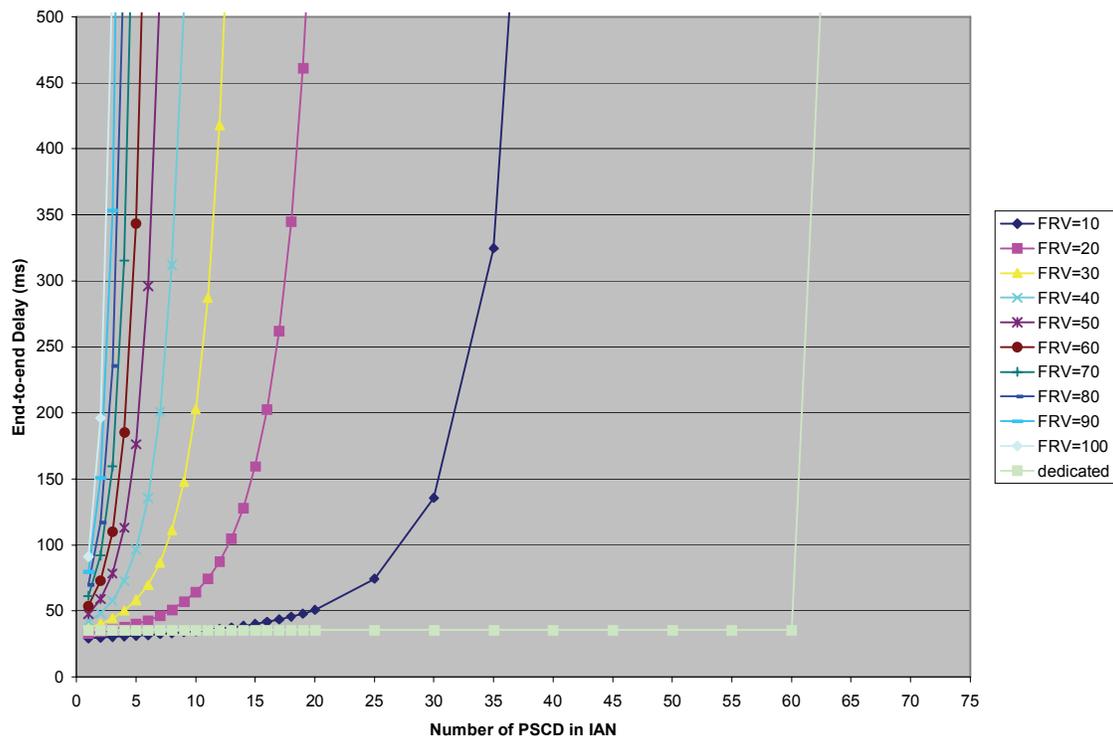
An increase in the channel data rate can produce improvements in MPEG-2 encoder application performance, although again the performance is not as good as what we can achieve with H.264. Figure 77 illustrates poor MPEG-2 encoder loss performance, but marginal performance is attainable using Slotted Aloha, although only when the network is very lightly loaded. Figure 78 shows the primary area of MPEG-2 improvement—delay. It is possible to achieve reasonable application performance using a MPEG-2 encoder even when the number of FRVs per JAN is on the order of 40, as long as the number of PSCDs per IAN remains at a reasonably small number.

Figure 77:   Packet Loss for MPEG-2 Video on Path C at 54 Mbps
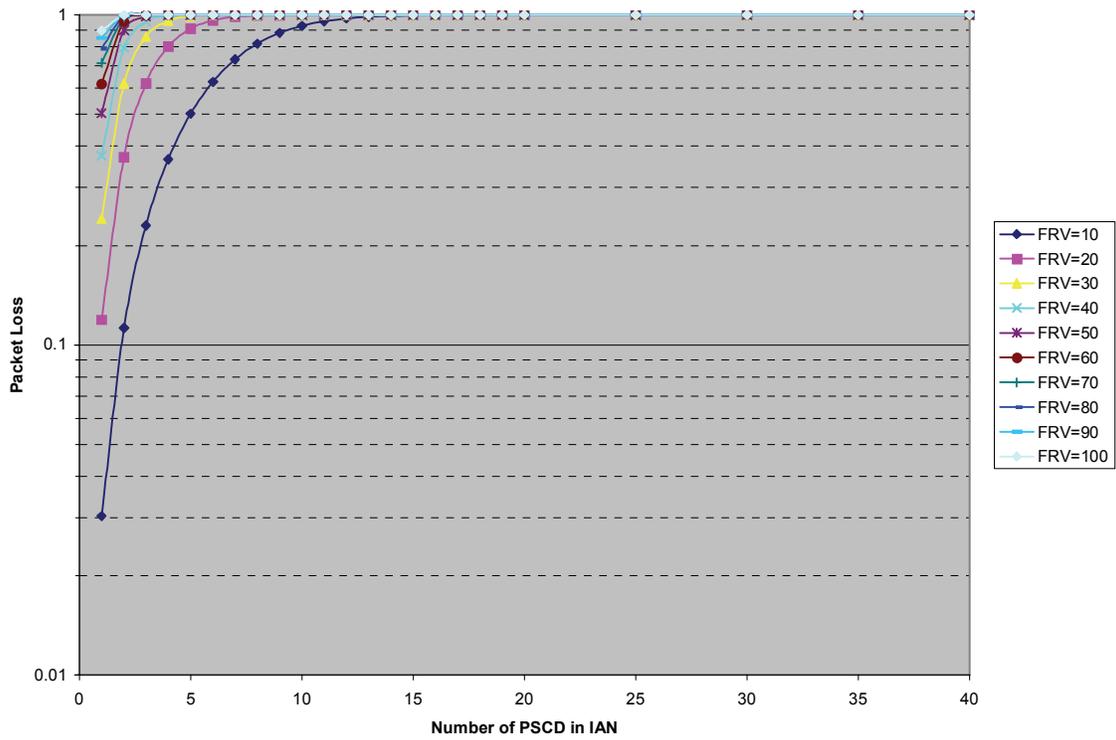
Figure 78:   Delay for MPEG-2 Video on Path C at 54 Mbps
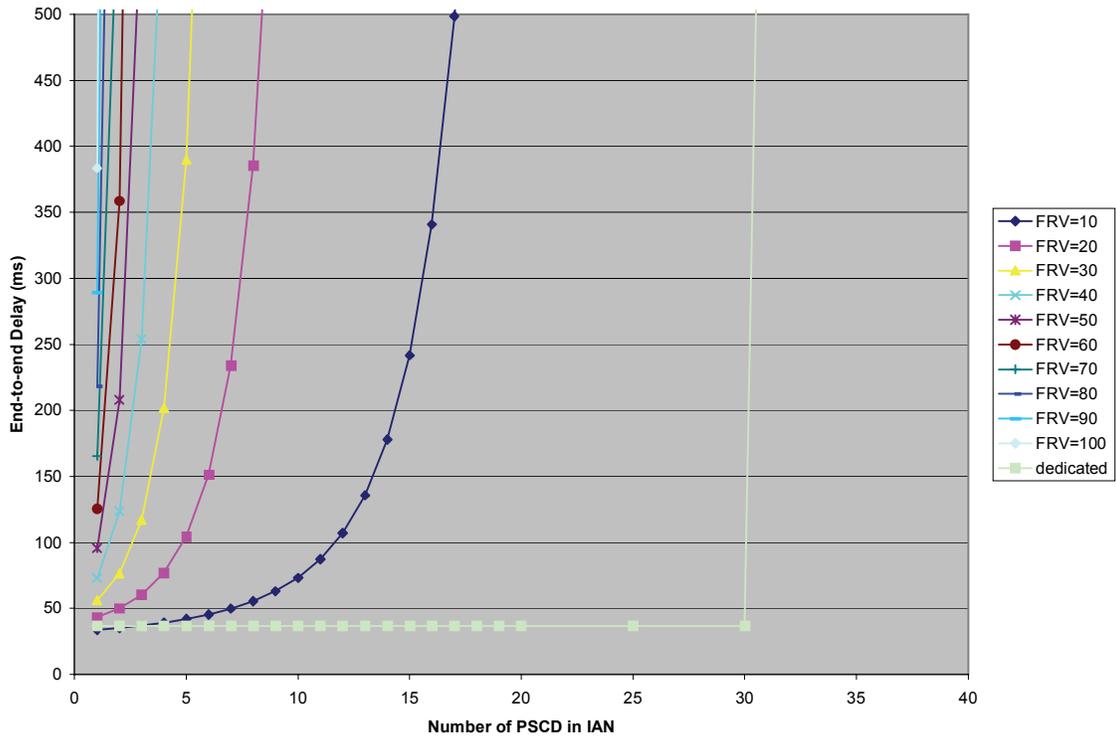
Appendix F   References

F.1   Print and Standards References

(ANSI, 2003)   American National Standards Institute, ANSI T1.801.03-2003, "American National Standard for Telecommunications—Digital Transport of One-Way Video Signals—Parameters for Objective Performance Assessment."

(ATIS, 2001)   Alliance for Telecommunications Industry Solutions, ATIS T1.TR.74-2001, "Technical Report on Objective Video Quality Measurement Using a Peak Signal-to-Noise Ratio (PSNR) Full Reference Technique."

(Bertsekas, 1992)   D. Bertsekas and R. Gallager, *Data Networks*, second edition, Prentice Hall, Inc., Upper Saddle River, NJ, 1992.

(FCC, 2004)   Federal Communications Commission, "FCC Adopts Solution to Interoperability Problem Faced by 800 MHz Public Safety Radio Systems."

(IEC, 2001)   International Electrotechnical Commission, IEC 61966-8, 2001, *International Standard*.

(IEEE, 1999)   Institute of Electrical and Electronics Engineers, Standard for Local and Metropolitan Area Networks, IEEE Std. 802.11, 1999 (Reaffirmed June 12, 2003), "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications."

(IEEE, 2004)   Institute of Electrical and Electronics Engineers, Standard for Local and Metropolitan Area Networks, IEEE Std. 802.16, 2004 (Including IEEE Std. 802.16-2001, IEEE Std. 802.16c-2002, and IEEE Std. 802.16a-2003), "Part 16: Air Interface for Fixed Broadband Wireless Access Systems."

(IEEE, 2005)   Institute of Electrical and Electronics Engineers, Standard for Local and Metropolitan Area Networks, IEEE Std. 802.3, 2005 (Revision of IEEE Std. 802.3-2002, including all approved amendments), "Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications."

(IETF, 1980)   The Internet Engineering Task Force, IETF Standard 6/RFC 768, "User Datagram Protocol."

(IETF, 1998)   The Internet Engineering Task Force, IETF RFC 2460, "Internet Protocol, Version 6 (IPv6) Specification."

(IETF, 1999)   The Internet Engineering Task Force, IETF RFC2680, "A One-Way Packet Loss Metric for IPPM."

(IETF, 2003)   The Internet Engineering Task Force, IETF RFC 3550, "RTP: A Transport Protocol for Real-Time Applications."

(ISO, 2000)   International Organization for Standardization, ISO 12233, "Photography—Electronic Still-Picture Cameras—Resolution Measurements."

(ISO, 2003)   International Organization for Standardization, ISO 15739, "Photography—Electronic Still-Picture Cameras—Noise Measurements."

(ISO/IEC, 2000)   International Organization for Standardization, ISO/IEC 13818 (commonly known as MPEG-2), "Information technology—Generic Coding of Moving Pictures and Associated Audio Information."

(ITU-R, 2002)   Recommendations of the International Telecommunication Union, Radiocommunication Sector, ITU-R Recommendation BT.500, "Methodology for the Subjective Assessment of the Quality of Television Pictures."

(ITU-R, 1995)   Recommendations of the International Telecommunication Union, Radiocommunication Sector, ITU-R Recommendation BT.601, "Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios."

(ITU-R, 2004)   Recommendations of the International Telecommunication Union, Radiocommunication Sector, ITU-R Recommendation BT.1683, "Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference."

(ITU-T, 1988)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.711, "Pulse Code Modulation (PCM) of Voice Frequencies."

(ITU-T, 1990)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.726, 1990, "40, 32, 24, 16 kbps Adaptive Differential Pulse Code Modulation (ADPCM)."

(ITU-T, 1996)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality."

(ITU-T, 1996)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation P.830, "Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs."

(ITU-T, 1998)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation P.931, "Multimedia Communications Delay, Synchronization, and Frame Rate Measurement."

(ITU-T, 1999)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.108, "Application of the E-model: A Planning Guide."

(ITU-T, 1999)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.711 Appendix I, "A High-Quality Low-Complexity Algorithm for Packet Loss Concealment with G.711."

(ITU-T, 2000)   Recommendations of the International Telecommunication Union, Telecommunication Sector, ITU¬T E.671, "Post-Selection Delay in PSTN/ISDN Networks Using Internet Telephony for a Portion of the Connection."

(ITU-T, 2001)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.712, "Transmission Performance Characteristics of Pulse Code Modulation Channels."

(ITU-T, 2001)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.1010, "End-User Multimedia QoS Categories."

(ITU-T, 2002)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation E.651, "Reference Connections for Traffic Engineering of IP Access Networks."

(ITU-T, 2002)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.113 Appendix I, "Provisional Planning Values for the Equipment Impairment Factor Ie and Packet-Loss Robustness Factor Bpl."

(ITU-T, 2002)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation Y.1541, "Network Performance Objectives for IP-Based Services."

(ITU-T, 2002)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation Y.1541, Appendix X, "Speech Quality Calculations for Y.1541 Hypothetical Reference Paths."

(ITU-T, 2002)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation Y.1541, Amendment 1, Revised Appendix VI, "Applicability of the Y.1221 Transfer Capabilities and IETF Differentiated Services to IP QoS Classes."

(ITU-T, 2002)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation Y.1541, Amendment 2, New Appendix XI, "Concatenating QoS Values."

(ITU-T, 2003)   Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.114, "One-Way Transmission Time."

(ITU-T, 2004)   Recommendations of the International Telecommunication Union, Radiocommunication Sector, ITU-T Recommendation J.144, "Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference."

(ITU-T, 2005) Recommendations of the International Telecommunication Union, Telecommunication Standardization Sector, ITU-T Recommendation G.107, "The E-Model, a Computational Model for Use in Transmission Planning."

(ITU-T, 2005) Recommendations of the International Telecommunication Union, Radiocommunication Sector, ITU-T Recommendation H.264, "Advanced Video Coding for Generic Audiovisual Services."

(Johnson, 1992) N. Johnson, S. Kotz, and A. Kemp, *Univariate Discrete Distributions,* second edition. New York: Wiley, 1992.

(MMCN, 2000) H. Sanneck and N. Le, "Speech Property-Based FEC for Internet Telephony Applications," *Proc. SPIE/ACM SIGMM Multimedia Computing and Networking Conference*.

(NFPA, 2003) National Fire Protection Association, NFPA 1901, "Standard for Automotive Fire Apparatus."

(NFPA, 2005) National Fire Protection Association, NFPA 1561, "Standard on Emergency Services Incident Management System."

(PS SoR Volume I, version 1.2, 2006) "Public Safety Statement of Requirements for Communications & Interoperability," Volume I: Qualitative, Version 1.2, August 18, 2006.

(Rom, 1990) R. Rom. and M. Sidi, *Multiple Access Protocols: Performance and Analysis*, Springer-Verlag, New York, 1990.

(Sharma, 2002) G. Sharma, *Digital Color Imaging Handbook*, CRC Press, December 23, 2002.

(SMIA, 2005) Standard Mobile Imaging Architecture, SMIA Specification 1.0 Part 5, *Camera Characterization Specification*, March 4, 2004, jointly developed by Nokia and STMicroelectronics for mobile imaging (camera phone) industry, available to members: http://www.smia-forum.org. Cited June 14, 2006.

(SMPTE, 1997) Society of Motion Picture and Television Engineers, SMPTE 259M, "Television—10-Bit 4:2:2 Component and 4fsc Composite Digital Signals—Serial Digital Interface."

(Strobel, 2000) Strobel, Compton, Current, and Zakia. 2000, *Basic Photographic Materials and Processes*, second edition., Focal Press, Massachusetts, p. 101 (the section on Log Exposure Range in the chapter on Photographic Sensitometry).

(TIA, 2004) Telecommunications Industry Association, TIA Project 25, "Public Safety Communications Interoperability – Frequently Asked Questions (FAQ)."

(Ziemer, 1985) R. E. Ziemer and W. H. Tranter, *Principles of Communications: Systems, Modulation, and Noise,* Houghton Mifflin Co., Boston.

F.2   Online References

Applications

Imatest camera, lens, scanner, and printer image performance measurement tool. Available from http://www.imatest.com. Cited June 14, 2006.

Infran image format conversion tool. Available from http://www.irfanview.com. Cited June 14, 2006.

Slant edge analysis tools—sfrmat2 and sfrwin. Available from http://www.i3a.org/downloads_iso_tools.html. Cited June 14, 2006.

Charts

GretagMacbeth color checker charts. Available from http://www.gretagmacbeth.com/index/products/products_colorcheckers.htm. Cited June 14, 2006. Corresponding sRGB reference pixel values are available from the file, *Lab & Spectral Data D65 & D50 Spectrolino.xls*, February 2006, supplied by GretagMacbeth.

Lamps

Dedolight DLH200D Sundance Halogen Metal Iodide (HMI). Available from http://www.dedolight.com, http://www.bhphotovideo.com, and other sources. Cited June 14, 2006.

GretagMacbeth Sol-Source daylight desk lamp with weighted base. Available from http://usa.gretagmacbethstore.com/, http://www.colorhq.com, and other sources. Cited June 14, 2006.

North Light Ceramic High Intensity Discharge (HID) Copy Light. Available from http://www.northlightproducts.com, http://www.betterlight.com/hid_copylights.html, and other sources. Cited June 14, 2006.

SoLux Task Lamp. Available from http://www.solux.net/, https://www.nleinternet.net/solux/cgi-bin/tlistore/index.html, and other sources. Cited June 14, 2006.

Online Specifications and Standards

International Electrotechnical Commission, IEC 61966-8, 2001, *International Standard*. Introduction available at http://domino.iec.ch/preview/info_iec61966-8%7Bed1.0%7Den.pdf. Cited June 14, 2006. Full standard available at http://webstore.ansi.org/ansidocstore/product.asp?sku=IEC+61966-8+Ed.+1.0+en%3A2001. Cited June 14, 2006.

SMIA 1.0 Part 5: Camera Characterisation Specification, Rev A March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

Figure 3 "Default Darkroom Set Up." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

Section 5.17 "Relative Illumination." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

Section 5.24 "Veiling Glare." *SMIA 1.0 Part 5: Camera Characterisation Specification*, 1.0 Rev A, March 10, 2005. Available at http://www.smia-forum.org/. Cited June 14, 2006.

Standards source for measuring digital camera performance. International Organization for Standardization. http://www.iso.org/. Cited June 14, 2006.

Web Pages

"Gamma FAQ—Frequently Asked Questions about Gamma." Charles Poynton. http://www.poynton.com/notes/colour_and_gamma/GammaFAQ.html. Cited June 14, 2006.

"Color FAQ—Frequently Asked Questions about Color." Charles Poynton. http://www.poynton.com/notes/colour_and_gamma/ColorFAQ.html. Cited June 14, 2006.

"Useful Color Equations." Click "Math." Bruce Lindbloom. http://www.brucelindbloom.com. Cited June 14, 2006.