

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: October 2, 2015

C. Gunther, Ed.
HARMAN
E. Grossman, Ed.
DOLBY
March 31, 2015

Deterministic Networking Professional Audio Requirements
draft-gunther-detnet-proaudio-req-01

Abstract

This draft documents the needs in the professional audio and video industry to establish multi-hop paths and optional redundant paths for characterized flows with deterministic properties. In this context deterministic implies that streams can be established which provide guaranteed bandwidth and latency which can be established from a Layer 3 (IP) interface.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 2, 2015.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Requirements Language	3
3.	Fundamental Stream Requirements	3
3.1.	Guaranteed Bandwidth	4
3.2.	Bounded and Consistent Latency	4
3.2.1.	Optimizations	5
4.	Additional Stream Requirements	6
4.1.	Deterministic Time to Establish Streaming	6
4.2.	Use of Unused Reservations by Best-Effort Traffic	6
4.3.	Layer 3 Interconnecting Layer 2 Islands	7
4.4.	Secure Transmission	7
4.5.	Redundant Paths	7
4.6.	Link Aggregation	8
4.7.	Traffic Segregation	8
4.7.1.	Packet Forwarding Rules, VLANs and Subnets	8
4.7.2.	Multicast Addressing (IPv4 and IPv6)	8
5.	Integration of Reserved Streams into IT Networks	9
6.	Security Considerations	9
6.1.	Denial of Service	9
6.2.	Control Protocols	9
7.	A State-of-the-Art Broadcast Installation Hits Technology Limits	10
8.	Acknowledgements	10
9.	IANA Considerations	10
10.	References	10
10.1.	Normative References	10
10.2.	Informative References	11
	Authors' Addresses	11

1. Introduction

The professional audio and video industry includes music and film content creation, broadcast, cinema, and live exposition as well as public address, media and emergency systems at large venues (airports, stadiums, churches, theme parks). These industries have already gone through the transition of audio and video signals from analog to digital, however the interconnect systems remain primarily point-to-point with a single (or small number of) signals per link, interconnected with purpose-built hardware.

These industries are now attempting to transition to packet based infrastructure for distributing audio and video in order to reduce

cost, increase routing flexibility, and integrate with existing IT infrastructure.

However, there are several requirements for making a network the primary infrastructure for audio and video which are not met by today's networks and these are our concern in this draft.

The principal requirement is that pro audio and video applications become able to establish streams that provide guaranteed (bounded) bandwidth and latency from the Layer 3 (IP) interface. Such streams can be created today within standards-based layer 2 islands however these are not sufficient to enable effective distribution over wider areas (for example broadcast events that span wide geographical areas).

Some proprietary systems have been created which enable deterministic streams at layer 3 however they are engineered networks in that they require careful configuration to operate, often require that the system be over designed, and it is implied that all devices on the network voluntarily play by the rules of that network. To enable these industries to successfully transition to an interoperable multi-vendor packet-based infrastructure requires effective open standards, and we believe that establishing relevant IETF standards is a crucial factor.

It would be highly desirable if such streams could be routed over the open Internet, however even intermediate solutions with more limited scope (such as enterprise networks) can provide a substantial improvement over today's networks, and a solution that only provides for the enterprise network scenario is an acceptable first step.

We also present more fine grained requirements of the audio and video industries such as safety and security, redundant paths, devices with limited computing resources on the network, and that reserved stream bandwidth is available for use by other best-effort traffic when that stream is not currently in use.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Fundamental Stream Requirements

The fundamental stream properties are guaranteed bandwidth and deterministic latency as described in this section. Additional stream requirements are described in a subsequent section.

3.1. Guaranteed Bandwidth

Transmitting audio and video streams is unlike common file transfer activities because guaranteed delivery cannot be achieved by re-trying the transmission; by the time the missing or corrupt packet has been identified it is too late to execute a re-try operation and stream playback is interrupted, which is unacceptable in for example a live concert. In some contexts large amounts of buffering can be used to provide enough delay to allow time for one or more retries, however this is not an effective solution when live interaction is involved, and is not considered an acceptable general solution for pro audio and video. (Have you ever tried speaking into a microphone through a sound system that has an echo coming back at you? It makes it almost impossible to speak clearly).

Providing a way to reserve a specific amount of bandwidth for a given stream is a key requirement.

3.2. Bounded and Consistent Latency

Latency in this context means the amount of time that passes between when a signal is sent over a stream and when it is received, for example the amount of time delay between when you speak into a microphone and when your voice emerges from the speaker. Any delay longer than about 10-15 milliseconds is noticeable by most live performers, and greater latency makes the system unusable because it prevents them from playing in time with the other players (see slide 6 of [SRP_LATENCY]).

The 15ms latency bound is made even more challenging because it is often the case in network based music production with live electric instruments that multiple stages of signal processing are used, connected in series (i.e. from one to the other for example from guitar through a series of digital effects processors) in which case the latencies add, so the latencies of each individual stage must all together remain less than 15ms.

In some situations it is acceptable at the local location for content from the live remote site to be delayed to allow for a statistically acceptable amount of latency in order to reduce jitter. However, once the content begins playing in the local location any audio artifacts caused by the local network are unacceptable, especially in those situations where a live local performer is mixed into the feed from the remote location.

In addition to being bounded to within some predictable and acceptable amount of time (which may be 15 milliseconds or more or less depending on the application) the latency also has to be

consistent. For example when playing a film consisting of a video stream and audio stream over a network, those two streams must be synchronized so that the voice and the picture match up. A common tolerance for audio/video sync is one NTSC video frame (about 33ms) and to maintain the audience perception of correct lip sync the latency needs to be consistent within some reasonable tolerance, for example 10%.

A common architecture for synchronizing multiple streams that have different paths through the network (and thus potentially different latencies) is to enable measurement of the latency of each path, and have the data sinks (for example speakers) buffer (delay) all packets on all but the slowest path. Each packet of each stream is assigned a presentation time which is based on the longest required delay. This implies that all sinks must maintain a common time reference of sufficient accuracy, which can be achieved by any of various techniques.

This type of architecture is commonly implemented using a central controller that determines path delays and arbitrates buffering delays.

3.2.1. Optimizations

The controller might also perform optimizations based on the individual path delays, for example sinks that are closer to the source can inform the controller that they can accept greater latency since they will be buffering packets to match presentation times of farther away sinks. The controller might then move a stream reservation on a short path to a longer path in order to free up bandwidth for other critical streams on that short path. See slides 3-5 of [SRP_LATENCY].

Additional optimization can be achieved in cases where sinks have differing latency requirements, for example in a live outdoor concert the speaker sinks have stricter latency requirements than the recording hardware sinks. See slide 7 of [SRP_LATENCY].

Device cost can be reduced in a system with guaranteed reservations with a small bounded latency due to the reduced requirements for buffering (i.e. memory) on sink devices. For example, a theme park might broadcast a live event across the globe via a layer 3 protocol; in such cases the size of the buffers required is proportional to the latency bounds and jitter caused by delivery, which depends on the worst case segment of the end-to-end network path. For example on today's open internet the latency is typically unacceptable for audio and video streaming without many seconds of buffering. In such scenarios a single gateway device at the local network that receives

the feed from the remote site would provide the expensive buffering required to mask the latency and jitter issues associated with long distance delivery. Sink devices in the local location would have no additional buffering requirements, and thus no additional costs, beyond those required for delivery of local content. The sink device would be receiving the identical packets as those sent by the source and would be unaware that there were any latency or jitter issues along the path.

4. Additional Stream Requirements

The requirements in this section are more specific yet are common to multiple audio and video industry applications.

4.1. Deterministic Time to Establish Streaming

Some audio systems installed in public environments (airports, hospitals) have unique requirements with regards to health, safety and fire concerns. One such requirement is a maximum of 3 seconds for a system to respond to an emergency detection and begin sending appropriate warning signals and alarms without human intervention. For this requirement to be met, the system must support a bounded and acceptable time from a notification signal to specific stream establishment. For further details see [ISO7240-16].

Similar requirements apply when the system is restarted after a power cycle, cable re-connection, or system reconfiguration.

In many cases such re-establishment of streaming state must be achieved by the peer devices themselves, i.e. without a central controller (since such a controller may only be present during initial network configuration).

Video systems introduce related requirements, for example when transitioning from one camera feed to another. Such systems currently use purpose-built hardware to switch feeds smoothly, however there is a current initiative in the broadcast industry to switch to a packet-based infrastructure (see [STUDIO_IP] and the ESPN DC2 use case described below).

4.2. Use of Unused Reservations by Best-Effort Traffic

In cases where stream bandwidth is reserved but not currently used (or is under-utilized) that bandwidth must be available to best-effort (i.e. non-time-sensitive) traffic. For example a single stream may be nailed up (reserved) for specific media content that needs to be presented at different times of the day, ensuring timely delivery of that content, yet in between those times the full

bandwidth of the network can be utilized for best-effort tasks such as file transfers.

This also addresses a concern of IT network administrators that are considering adding reserved bandwidth traffic to their networks that users will just reserve a ton of bandwidth and then never un-reserve it even though they are not using it, and soon they will have no bandwidth left.

4.3. Layer 3 Interconnecting Layer 2 Islands

As an intermediate step (short of providing guaranteed bandwidth across the open internet) it would be valuable to provide a way to connect multiple Layer 2 networks. For example layer 2 techniques could be used to create a LAN for a single broadcast studio, and several such studios could be interconnected via layer 3 links.

4.4. Secure Transmission

Digital Rights Management (DRM) is very important to the audio and video industries. Any time protected content is introduced into a network there are DRM concerns that must be maintained (see [CONTENT_PROTECTION]). Many aspects of DRM are outside the scope of network technology, however there are cases when a secure link supporting authentication and encryption is required by content owners to carry their audio or video content when it is outside their own secure environment (for example see [DCI]).

As an example, two techniques are Digital Transmission Content Protection (DTCP) and High-Bandwidth Digital Content Protection (HDCP). HDCP content is not approved for retransmission within any other type of DRM, while DTCP may be retransmitted under HDCP. Therefore if the source of a stream is outside of the network and it uses HDCP protection it is only allowed to be placed on the network with that same HDCP protection.

4.5. Redundant Paths

On-air and other live media streams must be backed up with redundant links that seamlessly act to deliver the content when the primary link fails for any reason. In point-to-point systems this is provided by an additional point-to-point link; the analogous requirement in a packet-based system is to provide an alternate path through the network such that no individual link can bring down the system.

4.6. Link Aggregation

For transmitting streams that require more bandwidth than a single link in the target network can support, link aggregation is a technique for combining (aggregating) the bandwidth available on multiple physical links to create a single logical link of the required bandwidth. However, if aggregation is to be used, the network controller (or equivalent) must be able to determine the maximum latency of any path through the aggregate link (see Bounded and Consistent Latency section above).

4.7. Traffic Segregation

Sink devices may be low cost devices with limited processing power. In order to not overwhelm the CPUs in these devices it is important to limit the amount of traffic that these devices must process.

As an example, consider the use of individual seat speakers in a cinema. These speakers are typically required to be cost reduced since the quantities in a single theater can reach hundreds of seats. Discovery protocols alone in a one thousand seat theater can generate enough broadcast traffic to overwhelm a low powered CPU. Thus an installation like this will benefit greatly from some type of traffic segregation that can define groups of seats to reduce traffic within each group. All seats in the theater must still be able to communicate with a central controller.

There are many techniques that can be used to support this requirement including (but not limited to) the following examples.

4.7.1. Packet Forwarding Rules, VLANs and Subnets

Packet forwarding rules can be used to eliminate some extraneous streaming traffic from reaching potentially low powered sink devices, however there may be other types of broadcast traffic that should be eliminated using other means for example VLANs or IP subnets.

4.7.2. Multicast Addressing (IPv4 and IPv6)

Multicast addressing is commonly used to keep bandwidth utilization of shared links to a minimum.

Because of the MAC Address forwarding nature of Layer 2 bridges it is important that a multicast MAC address is only associated with one stream. This will prevent reservations from forwarding packets from one stream down a path that has no interested sinks simply because there is another stream on that same path that shares the same multicast MAC address.

Since each multicast MAC Address can represent 32 different IPv4 multicast addresses there must be a process put in place to make sure this does not occur. Requiring use of IPv6 address can achieve this, however due to their continued prevalence, solutions that are effective for IPv4 installations are also required.

5. Integration of Reserved Streams into IT Networks

A commonly cited goal of moving to a packet based media infrastructure is that costs can be reduced by using off the shelf, commodity network hardware. In addition, economy of scale can be realized by combining media infrastructure with IT infrastructure. In keeping with these goals, stream reservation technology should be compatible with existing protocols, and not compromise use of the network for best effort (non-time-sensitive) traffic.

6. Security Considerations

Many industries that are moving from the point-to-point world to the digital network world have little understanding of the pitfalls that they can create for themselves with improperly implemented network infrastructure. DetNet should consider ways to provide security against DoS attacks in solutions directed at these markets. Some considerations are given here as examples of ways that we can help new users avoid common pitfalls.

6.1. Denial of Service

One security pitfall that this author is aware of involves the use of technology that allows a presenter to throw the content from their tablet or smart phone onto the A/V system that is then viewed by all those in attendance. The facility introducing this technology was quite excited to allow such modern flexibility to those who came to speak. One thing they hadn't realized was that since no security was put in place around this technology it left a hole in the system that allowed other attendees to "throw" their own content onto the A/V system.

6.2. Control Protocols

Professional audio systems can include amplifiers that are capable of generating hundreds or thousands of watts of audio power which if used incorrectly can cause hearing damage to those in the vicinity. Apart from the usual care required by the systems operators to prevent such incidents, the network traffic that controls these devices must be secured (as with any sensitive application traffic). In addition, it would be desirable if the configuration protocols that are used to create the network paths used by the professional

audio traffic could be designed to protect devices that are not meant to receive high-amplitude content from having such potentially damaging signals routed to them.

7. A State-of-the-Art Broadcast Installation Hits Technology Limits

ESPN recently constructed a state-of-the-art 194,000 sq ft, \$125 million broadcast studio called DC2. The DC2 network is capable of handling 46 Tbps of throughput with 60,000 simultaneous signals. Inside the facility are 1,100 miles of fiber feeding four audio control rooms. (See details at [ESPN_DC2]).

In designing DC2 they replaced as much point-to-point technology as they possibly could with packet-based technology. They constructed seven individual studios using layer 2 LANS (using IEEE 802.1 AVB) that were entirely effective at routing audio within the LANs, and they were very happy with the results, however to interconnect these layer 2 LAN islands together they ended up using dedicated links because there is no standards-based routing solution available.

This is the kind of motivation we have to develop these standards because customers are ready and able to use them.

8. Acknowledgements

The editors would like to acknowledge the help of the following individuals and the companies they represent:

Jeff Koftinoff, Meyer Sound

Jouni Korhonen, Associate Technical Director, Broadcom

Pascal Thubert, CTAO, Cisco

Kieran Tyrrell, Sienda New Media Technologies GmbH

9. IANA Considerations

This memo includes no request to IANA.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

10.2. Informative References

[CONTENT_PROTECTION]

Olsen, D., "1722a Content Protection", 2012,
<[http://grouper.ieee.org/groups/1722/contributions/2012/
avtp_dolsen_1722a_content_protection.pdf](http://grouper.ieee.org/groups/1722/contributions/2012/avtp_dolsen_1722a_content_protection.pdf)>.

[DCI]

Digital Cinema Initiatives, LLC, "DCI Specification,
Version 1.2", 2012, <<http://www.dcinovies.com/>>.

[ESPN_DC2]

Daley, D., "ESPN's DC2 Scales AVB Large", 2014,
<[http://sportsvideo.org/main/blog/2014/06/
espns-dc2-scales-avb-large](http://sportsvideo.org/main/blog/2014/06/espns-dc2-scales-avb-large)>.

[ISO7240-16]

ISO, "ISO 7240-16:2007 Fire detection and alarm systems --
Part 16: Sound system control and indicating equipment",
2007, <[http://www.iso.org/iso/
catalogue_detail.htm?csnumber=42978](http://www.iso.org/iso/catalogue_detail.htm?csnumber=42978)>.

[SRP_LATENCY]

Gunther, C., "Specifying SRP Latency", 2014,
<[http://www.ieee802.org/1/files/public/docs2014/
cc-cgunther-acceptable-latency-0314-v01.pdf](http://www.ieee802.org/1/files/public/docs2014/cc-cgunther-acceptable-latency-0314-v01.pdf)>.

[STUDIO_IP]

Mace, G., "IP Networked Studio Infrastructure for
Synchronized & Real-Time Multimedia Transmissions", 2007,
<[http://www.ieee802.org/1/files/public/docs2047/
avb-mace-ip-networked-studio-infrastructure-0107.pdf](http://www.ieee802.org/1/files/public/docs2047/avb-mace-ip-networked-studio-infrastructure-0107.pdf)>.

Authors' Addresses

Craig Gunther (editor)
Harman International
10653 South River Front Parkway
South Jordan, UT 84095
USA

Phone: +1 801 568-7675
Email: craig.gunther@harman.com
URI: <http://www.harman.com>

Ethan Grossman (editor)
Dolby Laboratories, Inc.
100 Potrero Ave
San Francisco, CA 94103
USA

Phone: +1 415 645 4726
Email: ethan.grossman@dolby.com
URI: <http://www.dolby.com>