

STANDARDS AND INFORMATION DOCUMENTS

AES67-2018
(Rev. AES67-2013)



AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability

Users of this standard are encouraged to determine if they are using the latest printing incorporating all current amendments and editorial corrections. Information on the latest status, edition, and printing of a standard can be found at:
<http://www.aes.org/standards>

AUDIO ENGINEERING SOCIETY, INC.
551 Fifth Avenue, New York, NY 10176. US.



The AES Standards Committee is the organization responsible for the standards program of the Audio Engineering Society. It publishes technical standards, information documents and technical reports. Working groups and task groups with a fully international membership are engaged in writing standards covering fields that include topics of specific relevance to professional audio. Membership of any AES standards working group is open to all individuals who are materially and directly affected by the documents that may be issued under the scope of that working group.

Complete information, including working group scopes and project status is available at <http://www.aes.org/standards>. Enquiries may be addressed to standards@aes.org

The AES Standards Committee is supported in part by those listed below who, as Standards Sustainers, make significant financial contribution to its operation.



audio-technica



This list is current as of 2018/4/18

AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability

Published by

Audio Engineering Society, Inc.

Copyright ©2013, 2015, 2017, 2018 by the Audio Engineering Society

Abstract

High-performance media networks support professional quality audio (16 bit, 44,1 kHz and higher) with low latencies (less than 10 milliseconds) compatible with live sound reinforcement. The level of network performance required to meet these requirements is available on local-area networks and is achievable on enterprise-scale networks. A number of networked audio systems have been developed to support high-performance media networking but until now there were no recommendations for operating these systems in an interoperable manner. This standard provides comprehensive interoperability recommendations in the areas of synchronization, media clock identification, network transport, encoding and streaming, session description and connection management.

An AES standard implies a consensus of those directly and materially affected by its scope and provisions and is intended as a guide to aid the manufacturer, the consumer, and the general public. The existence of an AES standard does not in any respect preclude anyone, whether or not he or she has approved the document, from manufacturing, marketing, purchasing, or using products, processes, or procedures not in agreement with the standard. Prior to approval, all parties were provided opportunities to comment or object to any provision. Attention is drawn to the possibility that some of the elements of this AES standard or information document may be the subject of patent rights. AES shall not be held responsible for identifying any or all such patents. Approval does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties adopting the standards document. This document is subject to periodic review and users are cautioned to obtain the latest edition. Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Audio Engineering Society Inc. 551 Fifth Avenue, New York, NY 10176, US.

www.aes.org/standards standards@aes.org

2018-04-15 printing

 **Documents**
Document Sharing Hub

Contents

0 Introduction	6
0.1 General.....	6
0.2 Patents	6
1 Scope	7
2 Normative references	7
3 Definitions and abbreviations	8
4 Synchronization	13
4.0 General.....	13
4.1 IP network synchronization	13
4.2 IEEE 1588 network synchronization	13
4.3 AVB network synchronization	13
5 Media clocks	13
6 Transport	14
6.0 General.....	14
6.1 Network layer	14
6.2 Quality of service.....	15
6.3 Transport layer	16
7 Encoding and streaming.....	17
7.0 Introduction	17
7.1 Payload format and sampling rate	17
7.2 Packet time	17
7.2.0 General.....	17
7.2.1 Required packet time	18
7.2.2 Recommended packet times.....	18
7.3 Stream channel count	19
7.4 Link offset.....	19
7.5 Sender timing and receiver buffering	20
7.6 Multicasting	21
8 Session description.....	21
8.0 General.....	21
8.1 Packet time	21
8.2 Clock source	22
8.3 RTP and media clock.....	23
8.4 Payload types.....	23
8.5 Example descriptions	23
8.5.0 Errata.....	23
8.5.1 Multicast session description example	23
8.5.2 Unicast session description example	24
9 Discovery	24
10 Connection management.....	24
10.0 General	24
10.1 Unicast connections	24
10.1.1 SIP URI.....	24
10.1.2 Server and serverless modes.....	25
10.1.3 User-Agent	25
10.1.4 Format negotiation	25
10.1.5 Packet time negotiation	25
10.2 Multicast connections.....	25

Annex A (Normative) - Media profile	26
A.0 General	26
A.1 Media profile description	26
A.2 Media profile	26
A.2.1 Identification	26
A.2.2 PTP attribute values	26
A.2.3 PTP options	29
A.2.4 Clock physical requirements	29
Annex B (Informative) - Network QoS configuration recommendations	30
B.0 General	30
B.1 DiffServ network configuration	30
B.1.1 Clock	30
B.1.2 Media	31
B.1.3 Best effort	31
Annex C (Informative) – AVB network transport	33
C.0 General	33
C.1 AVB network transport	33
C.1.1 Interoperable media as AVB time-sensitive streams	33
C.1.2 Interoperable media as other traffic	34
Annex D (Informative) - Interfacing to IEEE 802.1AS clock domains	36
D.0 General	36
D.1 Boundary clock interface	36
D.2 Ordinary clock interface	36
D.3 Traceable reference	36
D.4 AVB network as a boundary clock	37
Annex E (Informative) – Discovery systems	38
E.0 General	38
E.1 Bonjour	38
E.2 SAP	38
E.3 Axia Discovery Protocol	38
E.4 Wheatstone WheatnetIP Discovery Protocol	38
E.5 AMWA NMOS Discovery and Registration Specification (IS-04)	39
Annex F (Informative) - Glossary	40
Annex G (Normative) - Protocol implementation conformance criteria	42
G.1 Introduction	42
G.2 Instructions for completing the PICS proforma	42
G.3 PICS proforma	43
G.3.1 Identification	43
G.3.2 Synchronization	44
G.3.3 Media clocks	45
G.3.4 Transport	46
G.3.5 Encoding and streaming	49
G.3.6 Session description	57
G.3.7 Clock source	58
G.3.8 Discovery	60
G.3.9 Connection management	60
G.3.10 Media profile (Normative)	62
G.3.11 Media profile	62
G.4 – Qualification criteria for encoding and streaming capabilities	69
Annex H Bibliography	70

Foreword

This foreword is not part of the AES67-2013 *AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability*

This document was developed in project AES-X192, in the SC-02-12-H task group on high-performance streaming audio-over-IP interoperability, under the leadership of Kevin Gross.

Members of the writing group that contributed to this document in draft are: R. Abraham, J. Amate, M. Barbour, C. Becker-Foss, J. Berryman, M. Bishop, H. Blum, J. Boqvist, T. Borland, N. Borthwick, L. Bradshaw, P. Briscoe, L. Brito, J. Britton, C. Broad, N. Brunsgaard, D. Brutzman, E. Bukont Jr., A. Calvanese, R. Camprodon, M. Carter, A. Cedronius, A. Clark, H. Clarke, M. Coinchon, P. Cyrta, S. de Jaham, P. Demuytere, J. Dibley, A. Dickens, P. Dietrich, C. Dodds, S. Dove, A. Eales, E. Echols, R. Economaki, T. Edwards, A. Elder, L. Ellison, K. Fitzke, J. Fletcher, S. Flock, R. Foss, P. Foulkes, F. Gierlinger, R. Goforth, J. Grant, M. Graubner, D. Gravereaux, S. Gray, H. Hansen, B. Harshbarger, S. Heinzmann, A. Hildebrand, D. Hoch, M. Holtmann, Os. Igumbor, H. Jahne, T. Johansson, M. J. Teener, S. Johnson, L. Jonsson, A. Karamustafaoglu, K. Kearney, P. Keller, J. Koftinoff, P. Koftinoff, D. Koss, S. Langhans, M. Lave, D. Lazecko, S. Ledergerber, C. Lefebvre, J. Lindsay, G. Linis, S. Loehberg, A. Louko, A. Makivirta, J. A. Martinez, A. Mayo, W. McQuay, C. Merienne, A. Metz, J. Meunier, J. Meyer, R. Michl, S. Millan, D. O'Gwynn, N. O'Neill, H. Okai-Tettey, K. Parker, J. Passaniti, J. Peavey, J. Perez, W. Peters, S. Price, S. Pro, M. Quaix, F. Ragenard, R. Rayburn, D. Rice, T. Rohwedder, G. Rosenboom, M. Saito, M. S. Carreres, J. Sauter, R. Schoonbroodt, V. Schueppel, P. Schwizer, S. Scott, G. Shay, T. Shuttleworth, D. Silver, J. Simpson, M. Sims, A. Smimite, J. Snow, T. Staros, P. Stevens, J. M. Strawn, N. Sturmelm, M. Szlapka, T. Thompson, P. Treleaven, A. Trevena, S. Turner, R. van der Zalm, B. van Kempen, I. Vysick, J. P. Waddell, Y. Wang, P. Warrington, H. Weibel, A. Williams, M. Willsher, A. Witham, J. Wood, J. A. Yeary, J. Yoshio, P. Yurt, U. Zanghieri, D. Zimmermann, R. Zwiebel.

This document was edited by Kevin Gross.

Richard Foss

Chair, working group SC-02-12, 2013-07-18

Foreword to second edition, 2015

This revision includes minor changes identified during 'plugfest' testing in October 2014 and was developed in task group SC-02-12-M. It includes updated references to RFC 7273, and clarifications in 6.3, 8.1, and 8.5.

Members of the task group that contributed to this revision in draft are: R. Abraham, J. Amate, L. Andrieu, R. Barbieri-Carrera, M. Barbour, C. Becker-Foss, F. Bergholtz, J.A. Berryman, J. Boqvist, J. Breitlow, D. Brulhart, B.J. Buchalter, A. Calvanese, C. Cellier, K. Cleary, M. Danielson, I. Dennis, C. Diehl, S. Dove, T. Duffy, J. Dunn, J. Evanson, J. Fletcher, S. Flock, R. Foss, J. Freyberger, G. Gebler, J. Grant, K. Gross, S. Heinzmann, E. Heurtel, A. Hildebrand, F. Hoyer, T. Johansson, L. Jonsson, B. Klinkradt, J. Koftinoff, S. Langhans, S. Ledergerber, C. Lefebvre, S. Leschka, Jo. Lindsay, G. Linis, A. Lucas, A. Makivirta, A. Metz, J. Meyer, R. Michl, G. Passador, J. Peavey, M. Quaix, C.R. Reed, T. Rohwedder, M. Schuchert, V. Schueppel, G. Shay, P. Stevens, N. Sturmelm, M. Szlapka, P. Treleaven, A. van den Broek, B. van Kempen, J.P. Waddell, A. Williams, K. Wu, N. Yamaguchi, R. Zwiebel.

The task group was led by Kevin Gross.

Richard Foss

Chair, working group SC-02-12, 2015-07-27

Foreword to third edition, 2018

This revision contains clarifications and minor corrections and adds a Protocol Implementation Conformance Statement (PICS) as Annex G. A new sender keep-alive recommendation, has been added to clause 6.3. Minor clarifications and corrections include a specification of MTU requirements in the presence of allowed (but not recommended) additional information in the RTP header and correcting SDP examples in clause 8.5 to match an erratum issued by the IETF on RFC 7273. Corrected domainNumber range to conform to IEEE 1588-2008 requirements. These revisions were developed in task group SC-02-12-M.

Members of the task group that contributed to this revision in draft are: R. Abraham, J. Amate, N. Amaya, L. Andrieu, R. Barbieri-Carrera, M. Barbour, C. Becker-Foss, R. Bell, F. Bergholtz, J. A. Berryman, M. Blackburn, A. Boenninghoff, J. Boqvist, L. Bradshaw, D. Breithaupt, J. Breitlow, D. Brulhart, B. J. Buchalter, R. Bugg, R. Cabot, A. Caceres, A. Calvanese, C. Cellier, R. Charlesworth, K. Cleary, B. Cochran, A. Cooper, M. Danielson, T. deBrouwer, I. Dennis, C. Diehl, G. Diehl, S. Dove, T. Duffy, J. Dunn, J. Evanson, J. Fletcher, S. Flock, R. Foss, J. Freyberger, G. Gauthier, G. Gebler, J. Grant, E. Grossman, S. Heinzmann, M. Henke, M. Henry, E. Heurtel, A. Hildebrand, T. Holton, A. Holzinger, F. Hoyer, L. Huapaya, H. Jesuiter, T. Johansson, L. Jonsson, N. Keroe, A. Kitagawa, B. Klinkradt, J. Koftinoff, S. Langhans, J. Laundon, S. Ledergerber, C. Lefebvre, S. Leschka, E. Lestage, J. Lindsay, G. Linis, K. Lyver, A. Makivirta, C. Mannett, S. Mertens, A. Metz, J. Meyer, R. Michl, E. Miehs, T. Neuhaus, C. Nighman, N. Nzoyem, B. Olson, M. Overton, O. Palm, K. Parker, G. Passador, J. Peavey, J. Pruitt, M. Quaix, C. R. Reed, T. Rohwedder, A. Santos, G. Scherling, M. Schettke, M. Schuchert, V. Schueppel, S. Scott, A. Sharifan, G. Shay, M. Smaak, K. Soma, P. Stevens, N. Sturm, M. Szlapka, J. Tikkanen, P. Treleaven, A. van den Broek, B. van Kempen, S. van Tienen, J. P. Waddell, P. Walker, D. Walters, C. Ware, P. Warrington, E. Wetzell, L. Whitcomb, A. Williams, K. Wu, N. Yamaguchi, M. Yonge, R. Zwiebel.

The task group was led by Kevin Gross.

The PICS (annex H) was edited by Gints Linis

Morten Lave

Chair, working group SC-02-12, 2017-12-11

Note on normative language

In AES standards documents, sentences containing the word “shall” are requirements for compliance with the document. Sentences containing the verb “should” are strong suggestions (recommendations). Sentences giving permission use the verb “may”. Sentences expressing a possibility use the verb “can”.

2018-04-15 printing

AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability

0 Introduction

0.1 General

High-performance media networks support professional quality audio (16 bit, 44,1 kHz and higher) with low latencies (less than 10 ms) compatible with live sound reinforcement. The level of network performance required to meet these requirements is available on local-area networks and is achievable on enterprise-scale networks but is generally not available on wide-area networks or the public internet.

The most recent generation of these media networks use a diversity of proprietary and standard protocols. Despite a common basis in Internet Protocol, the systems do not interoperate.

This standard provides specific recommendations for interoperability. The standard focuses on defining how existing protocols are used to create an interoperable system. No new protocols have been developed to achieve this.

The standard is expected to be useful for commercial audio applications including fixed and touring live sound reinforcement. It is also expected to be useful for distribution within broadcast, music production and post-production facilities.

This standard depends on established network protocols (see clause 2). These protocols may include additional options that are not required by this standard. Implementations of AES67 should tolerate these additional options.

0.2 Patents

The Audio Engineering Society draws attention to the fact that it is claimed that compliance with this AES standard or information document may involve the use of a patent.

The AES holds no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured the AES that it is willing to negotiate licenses under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is archived with the AES and listed on the public AES web site.

Information may be obtained from.

Audinate Pty Ltd.
PO Box 855
Ultimo, NSW 2007
Australia

Attention is drawn to the possibility that some of the elements of this AES standard or information document may be the subject of patent rights other than those identified above. AES shall not be held responsible for identifying any or all such patent rights.

1 Scope

This standard defines an interoperability mode for transport of high-performance audio over networks based on the Internet Protocol. For the purposes of the standard, high-performance audio refers to audio with full bandwidth and low noise. These requirements imply linear PCM coding with a sampling frequency of 44,1 kHz and higher and resolution of 16 bits and higher. High performance also implies a low-latency capability compatible with live sound applications. The standard considers latency performance of 10 milliseconds or less.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

AES11 - *AES recommended practice for digital audio engineering - Synchronization of digital audio equipment in studio operations*; Audio Engineering Society, New York, NY., US.

IEEE 1588-2008 - *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, July 2008, Institute of Electrical and Electronics Engineers (IEEE), US.

RFC 768 – *User Datagram Protocol*”, Internet Engineering Task Force

RFC 791 – *Internet Protocol*, Internet Engineering Task Force

RFC 1112 – *Host Extensions for IP Multicasting*, Internet Engineering Task Force

RFC 2236 - *Internet Group Management Protocol, Version 2*, Internet Engineering Task Force

RFC 2474 – *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, Internet Engineering Task Force

RFC 2616 - *Hypertext Transfer Protocol - HTTP/1.1*, Internet Engineering Task Force

RFC 2974 – *Session Announcement Protocol*, Internet Engineering Task Force

RFC 3190 – *RTP Payload Format for 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio*, Internet Engineering Task Force

RFC 3261 - *SIP: Session Initiation Protocol*, Internet Engineering Task Force

RFC 3264 - *An Offer/Answer Model with the Session Description Protocol (SDP)*, Internet Engineering Task Force

RFC 3376 - *Internet Group Management Protocol, Version 3*, Internet Engineering Task Force

RFC 3550 – *RTP: A Transport Protocol for Real-Time Applications*, Internet Engineering Task Force

RFC 3551 - *RTP Profile for Audio and Video Conferences with Minimal Control*, Internet Engineering Task Force

RFC 4566 – *Session Description Protocol*, Internet Engineering Task Force

RFC 5939 – *Session Description Protocol (SDP) Capability Negotiation*, Internet Engineering Task Force

RFC 7273 – *RTP Clock Source Signalling*, Internet Engineering Task Force

3 Definitions and abbreviations

For the purposes of this document, the following terms, definitions, and abbreviations apply.

3.1

Audio stream

See RTP stream.

3.2

Audio Video Bridging**AVB**

describes enhanced Ethernet networks specified in IEEE 802.1BA, IEEE 802.1Q-2011 and IEEE 802.1AS.

3.3

Boundary Clock

A clock that has multiple Precision Time Protocol (PTP) ports in a domain and maintains the timescale used in the domain. It may serve as the source of time, that is, be a master clock; and may synchronize to another clock, that is, be a slave clock. See IEEE 1588-2008.

3.4

Byte

A unit comprising 8 bits of data. Over IP networks, data is transported in units of bytes.

3.5

Digital Audio Reference Signal**DARS**

an audio clock signal defined in AES11.

3.6

CSRC

The contributing source (CSRC) is the source of a stream of RTP packets that has contributed to the combined stream produced by an RTP mixer

3.7

DiffServ

Differentiated services (DiffServ) is a system for classifying traffic and providing quality of service (QoS) on an IP network.

3.8

DSCP

The differentiated services code point (DSCP) is a 6-bit field in the IP packet header that is used for classification purposes. DSCP is part of the differentiated services architecture.

3.9

End-to-end Transparent Clock

A transparent clock that supports the use of the end-to-end delay measurement mechanism between slave clocks and the master clock. See IEEE 1588-2008.

3.10

Ethernet

Ethernet is a physical and data link layer family of computer networking technologies for local area networks (LANs). Ethernet uses a bus or star topology and supports data transfer rates from 10 Mbps, through 100 Mbps (Fast Ethernet) and onto Gigabit Ethernet, supporting data rates of 1 gigabit (1,000 megabits) per second.

3.11

EUI-64

A 64-bit globally unique identifier formed by combining a registered 24 or 36-bit company identifier and a company unique device identifier. The EUI-64 is similar to the EUI-48 which is used to assign Ethernet media access control (MAC) addresses.

3.12**Grandmaster identifier****GMID**

an EUI-64 used in IEEE 1588 and IEEE 802.1AS synchronization standards to uniquely identify the grandmaster serving a synchronization domain.

3.13**Grandmaster**

The master source of synchronization for clock distribution via PTP. The grandmaster is a network device and is identified by an EUI-64.

3.14**IEEE**

Institute of Electrical and Electronics Engineers is a professional association dedicated to advancing technological innovation and excellence. The IEEE publishes a wide range of communications standards.

3.15**IETF**

Internet Engineering Task Force is the volunteer standards-developing organization responsible for the Internet Protocol suite.

3.16**IGMP**

Internet Group Management Protocol (IGMP) is a communications protocol used by hosts to report their multicast group memberships to IPv4 routers.

3.17**Internet Protocol****IP**

the network layer protocol commonly used to transport data on networks built through interconnection of one or more local-area networks.

3.18**IPv4**

Internet Protocol version 4 is the most widely deployed version of the protocol and is widely used on the Internet and on local area networks (LANs).

3.19**IPv6**

Internet Protocol version 6 is the most recent revision of the Internet Protocol and is intended to replace IPv4 eventually.

3.20**Link offset**

Link offset specifies the amount of time media spends on the network and in buffers at the sender and receiver as illustrated in figure 1. Link offset is also known as *network latency* or *playout delay*.

3.21**Media clock**

The clock used by senders to sample and receivers to play digital media streams. The media clock for audio streams reads in units of samples. The relationship between media clock and network clock is defined in 5.

3.22**Media packet**

One of the data packets carrying media data as part of a media stream. A media packet contains one or more samples for one or more audio channels.

3.23**Media stream**

See RTP stream.

3.24**Maximum transmission unit****MTU**

The size of the IP packet, measured in bytes, that can be transferred using a specific data link connection. The MTU for an Ethernet data link is 1500 bytes.

3.25**Network clock**

The time delivered by the network synchronization mechanism defined in 4. The network clock reads in units of seconds.

3.26**Network layer**

The network layer is layer 3 of the OSI model and is responsible for packet forwarding and routing of variable-length data sequences from a source to a destination.

3.27**OSI model**

The Open Systems Interconnect Model characterizes and standardizes the functions of a communications system in terms of abstraction layers.

3.28**Packet time**

The real-time duration of the media data contained in a media packet. For example, a packet containing 12 samples of 48 kHz audio has a packet time of $12 \div 48 \text{ kHz} = 250$ microseconds.

3.29**Peer-to-peer Transparent Clock**

A transparent clock that, in addition to providing Precision Time Protocol (PTP) event transit time information, also provides corrections for the propagation delay of the link connected to the port receiving the PTP event message. In the presence of peer-to-peer transparent clocks, delay measurements between slave clocks and the master clock are performed using the peer-to-peer delay measurement mechanism. Source: IEEE 1588-2008.

3.30**Precision time protocol****PTP**

The general class clock distribution protocol standardized in IEEE 1588-2002, IEEE 1588-2008 and IEEE 802.1AS-2011.

3.31**Quality of service****QoS**

describes a system for classifying, marking and delivering traffic across a network in accordance with its performance requirements.

3.32**Receiver**

A network device with ability to receive at least one media stream from the network.

3.33**Request for Comment****RFC**

Request for Comments are memorandums published by the IETF relevant for the working of the Internet and Internet-connected systems. RFCs are referenced by number. RFC 791, for example, defines the Internet Protocol version 4 (IPv4).

3.34**RTCP**

A companion protocol of the Real-time Transport Protocol (RTP), providing statistics and control information for RTP media packets.

3.35**Real-time Transport Protocol****RTP**

is defined in RFC 3550 and provides a means for applications to organize, mark and transport their media packets using UDP/IP networking.

3.36**RTP clock**

Timestamps are carried in RTP packets containing stream data. Each stream has its own RTP clock. There is a constant offset between the media clock and the RTP clock (see 8.2).

3.37**RTP session**

An RTP session is a media connection between sender and receiver. RTP sessions may be unicast or multicast. In teleconferencing RTP applications, multicast sessions may have multiple senders and receivers. However, under this standard, a session is allowed only one sender (see 7.6).

3.38**RTP stream**

An RTP stream is a sequence of RTP packets with media data sent at regular interval. A stream may contain multiple channels. There may be multiple media streams per RTP session.

3.39**Session Description Protocol****SDP**

a format for describing RTP sessions and their operating parameters including network addressing, encoding format and other metadata. SDP is defined in RFC 4566.

3.40**Sender**

A network device with ability to source at least one media stream onto the network.

3.41**Session**

See RTP session.

3.42**Session Initiation Protocol****SIP**

a telecommunications connection management protocol defined in RFC 3261.

3.43**SIP URI**

A SIP URI is a URI used by SIP to identify user agents. SIP URI take the form sip:<user>@<domain> or sips:<user>@<domain>. See 10.1.1.

3.44**Slave Clock**

A clock that is synchronized to a master clock (the provider of time) within an environment that uses the Precision Time Protocol (PTP). A slave may, in turn, be a master to another clock and may simultaneously be a boundary clock.

3.45**Stream**

See RTP stream.

3.46**Transport Layer Security****TLS**

a cryptographic protocol for secure communication over IP networks.

3.47**Transparent clock**

A device that measures the time taken for a Precision Time Protocol (PTP) event message to transit the device and provides this information to clocks receiving this PTP event message. See IEEE 1588-2008. See also: end-to-end transparent clock; peer-to-peer transparent clock.

3.48**Transport layer**

The network layer is layer 4 of the OSI model and provides end-to-end communication services for network applications.

3.49**User datagram protocol****UDP**

constitutes a simple transport layer for the IP network layer. Defined in RFC 768.

3.50**Uniform resource identifier****URI**

an identifier for a network resource. An identification URI enables interaction with the resource over a network.

3.51**User agent**

A SIP endpoint device such as a VoIP telephone.

3.52**Virtual LAN****VLAN**

A single layer-2 network may be partitioned to create multiple distinct broadcast domains, which are mutually isolated so that packets can only pass between them via one or more routers; such a domain is referred to as a Virtual Local Area Network.

4 Synchronization

4.0 General

The ability for network participants to share an accurate common clock distinguishes high-performance media streaming from its lower-performance brethren such as Internet radio and IP telephony. Using a common clock, receivers anywhere on the network can synchronize their playback with one another. A common clock allows for a fixed and determinable latency between sender and receiver. A common clock assures that all streams are sampled and presented at exactly the same rate. Streams running at the same rate may be readily combined in receivers. This property is critical for efficient implementation of networked audio devices such as digital mixing consoles.

Synchronization of a common clock shall be achieved using IEEE 1588-2008 Precision Time Protocol (PTP).

IEEE 1588-2008 is profiled for use in different synchronization applications. A profile describes protocol attributes, available options and required device performance. IEEE 1588-2008 specifies default profiles for delay request-response (IEEE 1588-2008 annex J.3) and peer-to-peer (IEEE 1588-2008 annex J.4) mechanisms.

Devices, with the exception of certain AVB devices (see below), shall support the IEEE 1588-2008 default profiles. Devices supporting the default profiles shall use IPv4 encapsulation as described in IEEE 1588-2008 annex D.

As a single exception, devices that use the AVB synchronization mechanism described in 4.3, and that need to be connected to an AVB network in order to accomplish media streaming, are not required to implement the IEEE 1588-2008 default profiles.

4.1 IP network synchronization

Devices on standard IP networks should use the media profile defined in annex A to assure adequate performance for all applications. Devices may use the default profiles on IP networks but should recognize that lock time and accuracy will be degraded.

4.2 IEEE 1588 network synchronization

On networks built using switches with IEEE 1588-2008 capabilities (boundary clocks or transparent clocks), adequate performance for audio transport is achieved using the appropriate default profile.

NOTE Due to performance constraints, some IEEE 1588 network equipment may be unable to support the media profile.

4.3 AVB network synchronization

Enhanced Ethernet networks, as specified in IEEE 802.1Q-2011, commonly known as Audio Video Bridging (AVB), deliver synchronization using IEEE 802.1AS. IEEE 802.1AS defines an IEEE 1588-2008 profile. AVB networks may use their native IEEE 802.1AS synchronization profile in preference to the default profiles or media profile. Methods for building heterogeneous synchronization networks using IEEE 1588-2008 and IEEE 802.1AS-2011 are described in annex D.

5 Media clocks

The media clock is used by senders to sample and by receivers to play digital media streams. The media clock has a fixed relationship to the network clock. The media clock and the network clock shall share the IEEE 1588 epoch of 1 January 1970 00:00:00 TAI, as defined in IEEE 1588-2008 clause 7.2.2. Digital audio to be carried on the network shall be sampled according to the media clock, or sampling-frequency converted to conform to the media clock.

NOTE With the introduction of leap seconds at the beginning of 1972, the offset between TAI and UTC became an integer number of seconds. This integer relationship exists for timestamps in 1972 and thereafter. The relationship is non-integral for timestamps in 1971 and

prior. This leads to the alternate IEEE 1588 epoch definition given in IEEE 1588-2008 clause 7.2.2 of 31 December 1969 23:59:51.999918 UTC. Again, this non-integral UTC offset exists only for network clock time prior to 1972.

The media clock shall advance at an exact rate with respect to the network clock. The rate of the media clock shall be the same as the audio sampling frequency.

This standard supports three sampling frequencies: 44,1 kHz, 48 kHz and 96 kHz (see 7.1). The media clock for an audio stream sampled at 48 kHz advances exactly 48 000 samples for each elapsed second on the network clock, for example. The value of the media clock shall be 0 at the IEEE 1588 epoch and change to 1 exactly one sample period after the epoch.

RTP clocks operate with a constant offset with respect to the media clock. The offset shall be conveyed through session description (see 8.3) on a per-stream basis.

In network protocols and management interfaces, RTP and media clocks are typically represented as 32-bit integers. The media clock for a 48 kHz stream will overflow its 32-bit representation approximately every 24,86 hours. To assure proper phasing with respect to the network clock, a media clock using 32-bit representation shall accurately take into account all such overflows (rollovers) between the epoch and the current time.

6 Transport

6.0 General

Transport aspects describe how media data, once encoded and packetized, is transported across the network. In terms of the OSI model, this clause defines operation on layer 3 (network layer) and layer 4 (transport layer). The standard does not specify how interoperability is achieved at lower layers in the model. It is assumed that best practices in transport of IP over the network technologies in question are employed.

NOTE Carriage of IP over Ethernet is described in RFC 894 - A Standard for the Transmission of IP Datagrams over Ethernet Networks.

6.1 Network layer

Media packets shall be transported using IP version 4 as defined in RFC 791.

NOTE 1 Although care has been taken in design of this standard to facilitate future support for IPv6, support for IPv6 is outside the scope of this revision of the standard.

Despite a requirement in RFC 791, receivers are not required under this standard to support reassembly of fragmented packets. A receiver that does not support reassembly shall ignore IP packet fragments.

Senders may set the Don't Fragment flag (DF) bit in the IP header of outgoing media packets. In the event that a packet marked DF needs to be fragmented by the network, it will instead be dropped and an ICMP "Too Big" message will be sent back to the sender. Senders should terminate transmission of the offending stream in response to receipt of an ICMP "Too Big" message.

Multicast messaging, such as that used for synchronization, shall be accomplished using IP multicasting as described in RFC 1112.

NOTE 2 Additional tutorial information on IP multicasting is available in RFC 3170.

To ensure that desired multicasts are received and to allow the network to filter undesired multicasts, all devices shall support IGMPv2 as defined in RFC 2236 and may support IGMPv3 as defined in RFC 3376.

NOTE 3 IGMPv2 support is required because IGMPv3 devices operating on an IGMPv2 network experience a two minute startup delay looking for IGMPv3 services on the network.

NOTE 4 RFC 2236 and RFC 3376 include backwards-compatibility requirements. A device supporting IGMPv2 is able to correctly operate on a network supporting IGMPv1 or IGMPv2. A

device supporting IGMPv3 is able to correctly operate on a network supporting IGMPv1, IGMPv2 or IGMPv3.

Devices shall use IGMP to request reception of any multicasts required. These include receipt of IEEE 1588 synchronization messages (see IEEE 1588-2008 clause D.3), media streams using multicast addressing (see 7.6) and also messages of other application protocols that may be used on the device, such as discovery (clause 9), that use multicast messaging.

NOTE 5 IGMP registration data can be purged by the network in some routing reconfiguration scenarios. Such a purge may result in an interruption of streamed data. An effective method of expediting restoration of service is to retransmit IGMP membership reports. This can be achieved by closing and immediately reopening any affected multicast network sockets.

Senders shall use IGMP to request receipt of any multicast media packets they are going to send before starting to send such packets.

NOTE 6 By making this request, senders will not actually receive the packets they are sending. Rather, the purpose of this requirement is to prevent unnecessary flooding of multicast media data. Some IGMP snooping implementations flood any multicast which has no registered group members.

NOTE 7 This requirement is normally covered by a sender's desire to receive RTCP messages. However, under RFC 3550, devices are strongly encouraged but not absolutely required to send or receive RTCP packets.

6.2 Quality of service

On a network shared with unregulated non-real-time traffic, time-critical media traffic generally requires prioritized handling known as QoS. In order to facilitate the implementation of suitable QoS in the network, devices shall implement the DiffServ method as described in RFC 2474. DiffServ uses the DSCP field in each IP packet header to mark packets according to their traffic class so that the network can easily recognize packets that need to be treated preferentially.

Minimally the three traffic classes described in table 1 shall be supported. Devices shall tag outgoing traffic with an appropriate DSCP value. Devices should use the default values for the DSCP field as given in table 1, but traffic may be marked with alternate DSCP values as provided by a network administrator or user through a management interface. Senders may be configured to use the same DSCP values for multiple classes - this has the effect of combining classes. Devices are not required to implement a management interface and may use default values exclusively.

Media streams that require very low delivery latency may not traverse a network reliably when transported in the same QoS class with media streams using longer packet times. In order to differentiate media streams with different requirements, senders may be configured to use classes in addition to those shown in table 1.

Table 1 - QoS classes and DiffServ associations

Class name	Traffic type	Default DiffServ class (DSCP decimal value)
Clock	IEEE 1588-2008 <i>Announce, Sync, Follow_Up, Delay_Req, Delay_Resp, Pdelay_Req, Pdelay_Resp</i> and <i>Pdelay_Resp_Follow_Up</i> packets	EF (46)
Media	RTP and RTCP media stream data	AF41 (34)
Best effort	IEEE 1588-2008 signaling and management messages. Discovery and connection management messages.	DF (0)

NOTE 1 The DSCP markings on packets do not define any particular behavior of network devices or imply particular policies the network must implement. As a security measure, a network may even ignore incoming DSCP markings in which case it may distinguish and

prioritize the traffic through other means (for example, UDP port number, IP addressing). Such network issues are outside the scope of this standard, although annex B provides some informative guidelines for network administrators.

NOTE 2 This standard makes no specific recommendation as to DSCP marking of traffic outside the scope of this standard. Traffic generated by other applications is marked as DF (0) by most systems, a situation compatible with this standard.

Senders should mark outgoing RTCP packets with the same DSCP value as the respective RTP stream packets. Receivers should mark outgoing RTCP packets with the same DSCP value they would use on RTP packets if transmitting a similar stream.

Receivers shall make no assumptions about class associations from DSCP markings on received packets.

NOTE 3 DSCP markings may be changed in route by the network or assignments may be reconfigured at the sender without the receiver's knowledge.

6.3 Transport layer

The transport layer provides end-to-end communications between devices on a network. The layer handles issues of packet loss and reordering and implements multiplexing so that a single network connection can serve multiple applications on the end station.

Devices shall use Real-time Transport Protocol as defined in RFC 3550. Devices shall operate in accordance with RTP Profile for Audio and Video Conferences with Minimal Control as defined in RFC 3551. Devices should use the default ports allocated for RTP: 5004 for RTP and 5005 for RTCP (see RFC 3551, section 8). Senders may use other or additional ports. Receivers shall support use of other or additional ports by corresponding senders.

Devices shall use UDP as defined in RFC 768 for transport of RTP.

Fragmentation is undesirable and, under this standard, receivers are not required to perform reassembly (6.1). The standard 1500-byte Ethernet MTU is assumed. To prevent fragmentation through a standard Ethernet infrastructure when using IPv4, and to assure future compatibility with IPv6, the maximum allowed RTP payload size shall be 1440 bytes.

NOTE 1 On connections offering lower MTU than Ethernet's 1500 bytes, senders may wish to use a smaller maximum payload than specified here.

Encrypted streaming using TLS, while supported in the RTP protocol suite, is not supported as part of this interoperability standard.

Senders should not include contributing source (CSRC) identifiers in the RTP header. Senders should not add RTP header extensions (RFC 3550 clause 5.1). However, as per RFC 3551, receivers shall tolerate the presence of CSRC identifiers and header extensions.

If senders do include header extensions or CSRCs, the 1440 byte maximum allowed payload shall be adjusted downwards by the size of the added header material so as to avoid fragmentation through a standard Ethernet infrastructure.

Both senders and receivers should transmit RTCP messages as specified in RFC 3550 clause 6. Senders and receivers should allocate RTCP bandwidth as recommended in RFC 3551 clause 2 (RTCP report interval).

Unicast senders should monitor connectivity to their respective receivers in such a way as to detect failure of the receiver, and stop transmission, within 60 seconds.

NOTE 2 Senders that continue their unicast transmissions to a missing receiver unnecessarily consume network resources, and can generate excessive ARP traffic and potentially have their transmissions flooded to all devices on the network segment.

If a receiver implements RTCP as recommended by RFC 3550, RTCP receiver reports are generally sufficient to achieve the recommended monitoring. For monitoring of receivers that do not implement RTCP, senders may use any other monitoring means available to them including any of the following techniques.

- SIP session timers as described in RFC 4028
- SIP OPTION ping (see IETF draft-jones-sip-options-ping)
- ICMP Echo request (ping, see RFC 792)

7 Encoding and streaming

7.0 Introduction

Encoding describes the means in which audio is digitized and formatted into the sequence of packets that constitutes a stream.

7.1 Payload format and sampling rate

Payload format defines audio sample encodings. The following payload formats are supported:

- L16** 16-bit linear format defined in RFC 3551 clause 4.5.11
- L24** 24-bit linear format defined in RFC 3190 clause 4

All devices shall support 48 kHz sampling rate. Devices should support 96 and 44,1 kHz sampling rates. See also AES5.

When operating at 48 kHz sampling rate:

1. Receivers shall support both L16 and L24 encodings
2. Senders shall support either L16, or L24, or both encodings

When operating at 96 kHz sampling rate:

1. Both senders and receivers shall support L24 encoding

When operating at 44,1 kHz sampling rate:

2. Both senders and receivers shall support L16 encoding

While all devices shall be able to support 48 kHz sampling frequency, they are not required to accept 48 kHz connections at all times. They may, for instance, have a global sample frequency configuration and only accept 48 kHz connections when the user selects global 48 kHz mode. Although not required, devices may support multiple sampling frequencies simultaneously.

Devices may support L16 at 96 kHz and other audio format combinations defined in RFC 3551 and RFC 3190. However, sampling rate and payload combinations beyond those defined above are outside the scope of this standard.

NOTE 1 The indication of the format and sampling rate in use for a given stream is given by a combination of the payload type field in the RTP header (RFC 3550 clause 5.1) and description information associated with the stream (see 8).

NOTE 2 While RFC 3190 mentions an “emphasis” signaling parameter, emphasis is generally considered to be a legacy issue dating back to the early 1980s. Emphasis is no longer generally used in digital audio signals and it is not expected to be used in these inter-operable networked streams.

7.2 Packet time

7.2.0 General

Packet time is the real-time duration of the media data contained in a media packet. Given the sampling rate and packet time, the number of samples per packet can be calculated.

Short packet times allow for lower latency but introduce overhead and high packet rates that may overtax some devices or networks. Long packet times imply higher latency and require additional buffering which may not be available on memory-constrained devices.

Packet time is determined by the sender, is specified in the session description (see 8.1) and negotiated through connection management (see 10). Senders shall not change packet time for the duration of a session. Although not required, receivers may adapt to packet time changes during a session. To enable interoperation with standard RTP implementations, receivers should not rely on the presence or accuracy of any packet time description. Receivers should be able to determine packet time based on the timestamps in received packets.

Interoperability is addressed by the requirement that devices implement the 1-millisecond packet time defined in 7.2.1. Further interoperability is encouraged through additional packet time recommendations in 7.2.2.

Product documentation for a device shall indicate which packet times are supported in send and receive directions.

7.2.1 Required packet time

A packet time of 1 millisecond offers the widest possible interoperability and compatibility with audio and network equipment.

Senders shall be able to load each audio packet with 48 samples of audio data when operating at a sampling frequency of 48 kHz or 44,1 kHz and 96 samples when operating at a sampling frequency of 96 kHz. Receivers shall be able to receive and decode 48-sample packets when operating at a sampling frequency of 48 kHz or 44,1 kHz, and 96-sample packets when operating at a sampling frequency of 96 kHz.

While all devices shall support the packet time requirements specified above, they are not required to accept these connections at all times. They may, for instance, have a global packet time configuration and only accept these connections when so configured. Although not required, devices may support multiple packet times simultaneously.

7.2.2 Recommended packet times

For enhanced interoperability over a range of applications, senders and receivers should support one or more of the other packet times listed in table 2.

Senders and receivers may support additional packet times. Maximum packet time is limited by network MTU as described in 6.3.

Table 2 - Required and recommended packet times

Packet time	Packet samples (48 kHz)	Packet samples (96 kHz)	Packet samples (44,1 kHz)	Notes
“125 microseconds”	6	12	6	Compatible with class A AVB transport
“250 microseconds”	12	24	12	High-performance, low-latency operation. Interoperable with class A and compatible with class B AVB transport.
“333 microseconds”	16	32	16	Efficient low-latency operation
“1 millisecond”	48	96	48	Required common packet time for all devices adhering to this standard (see 7.2.1)

“4 milliseconds”	192	n.a.	192	For applications desiring interoperability with EBU Tech 3326 or transport over wider areas or on networks with limited QoS capability
------------------	-----	------	-----	--

NOTE 96 kHz is not discussed in EBU Tech 3326. MTU restrictions of clause 6.3 limit a 96 kHz audio stream using 4-ms packet time to a single channel.

7.3 Stream channel count

The maximum number of channels per stream is limited by the packet time, encoding format and network MTU as described in 6.3.

Receivers shall support reception of streams with 1 to 8 audio channels. Receivers may support streams with more than 8 channels. Senders shall be able to offer at least one stream with 8 channels or fewer. Senders may support streams with more than 8 channels.

Table 3 - Examples: maximum channel capacities per stream

Format, sampling rate	Packet time	Maximum channels per stream
L24, 48 kHz	125 microseconds	80
L16, 48 kHz	250 microseconds	60
L24, 48 kHz	250 microseconds	40
L24, 48 kHz	333-1/3 microseconds	30
L24, 96 kHz	250 microseconds	20
L24, 48 kHz	1 millisecond	10
L24, 48 kHz	4 milliseconds	2

NOTE- Although bundling multiple channels in a stream can improve network and processing efficiency, it is recommended that bundling be used primarily in service of the application. Channels of related material (for example, stereo or surround sound) are good candidates for bundling. Bundling of unrelated channels destined for different receivers in an effort to reduce network overhead is discouraged as this complicates media routing configuration.

7.4 Link offset

Link offset describes the latency through a media network. It is defined as the difference in time between when audio enters the sender (*ingress time*) and when it leaves the receiver (*egress time*).

Ingress time is referenced at ingress to the *sender network system*. RTP packets are marked with origination timestamps in the *timestamp* field (RFC 3550 clause 5.1) based on this reference point. *Egress time* is referenced at egress from the *receiver network system*. *Link offset* is therefore the time difference between ingress at the sender and egress at the receiver. Link offset and ingress and egress reference points are illustrated in figure 1.

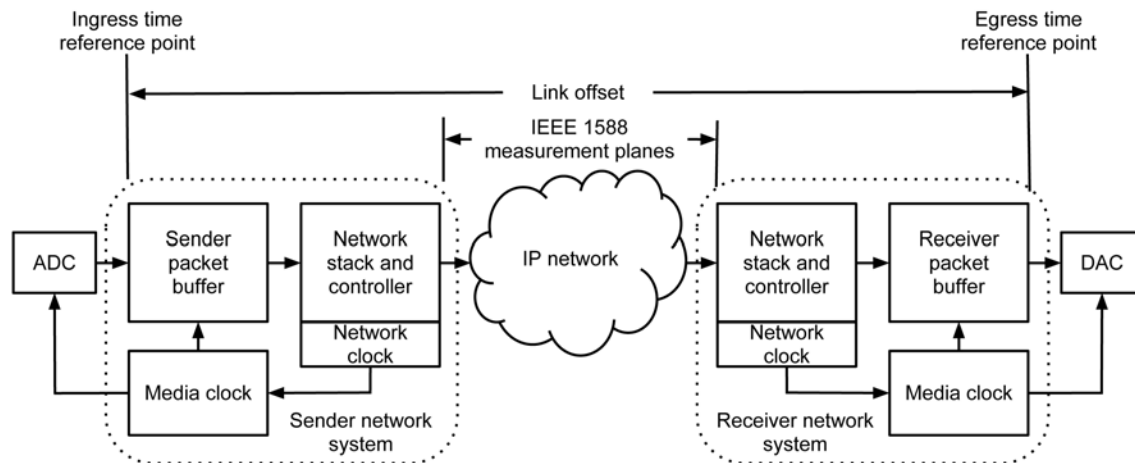


Figure 1 - Example network illustrating link offset and ingress and egress reference points

Link offset is determined at the receiver and is dependent on multiple factors, including packet time, propagation and queuing delays through the network, packet handling in the devices and buffering at the receiver. A receiver should attempt to maintain a constant link offset. At the same time it is recognized that unexpected changes to network conditions may require changing the buffering at the receiver resulting in a change of the link offset. The link offset and any changes in link offset should be communicated to the management entity within the receiver, if present.

NOTE 1 Minimum possible link offset is packet time (see 7.2) plus network forwarding time. Forwarding time for a minimum-sized packet on a point-to-point gigabit Ethernet connection can be as low as 0,5 microseconds. Minimum link offsets in an actual implementation under realistic network conditions will approach twice the packet time and beyond.

NOTE 2 Future work may specify a link offset management mechanism which is expected to require additional buffering and reporting of link offset to a central latency management server on the network and means for receivers to adjust link offset based on commands from the latency management server. See RFC 7272 for details.

7.5 Sender timing and receiver buffering

Buffering at the receiver is required to absorb jitter generated by packetization, network delivery and in the sender or receiver's network stacks and controllers. The receiver's buffer must accommodate media for the duration of the link offset minus the minimum delivery time between sender and receiver. The sender's packet buffer must accommodate packet time plus any variation introduced by the sender's network stack and controller. If buffering is too short, data may not arrive in time to be played, resulting in audio dropouts. Longer buffering improves robustness but introduces additional latency.

Receivers shall have a buffer capacity at least 3 times the packet time. Receivers should have a buffer capacity at least 20 times the packet time or 20 ms whichever is smaller.

Senders nominally send packets associated with a stream at packet time intervals. Senders should transmit at the nominal transmission time with a variation of 1 packet time or less. Senders shall transmit data at the nominal transmission time with a variation of no more than 17 packet times or 17 ms whichever is smaller.

The above requirements are designed to allow a range of implementations from hardware to applications running on desktop operating systems. Additional buffering and more accurate transmission timing are encouraged and will produce improved robustness and interoperability.