



Presidio of San Francisco
P.O. Box 29920
572 B Ruger Street
San Francisco, CA 94129-0920

T 415.561.6276
F 415.561.6120

www.isma.tv

INTERNET STREAMING MEDIA ALLIANCE

Encryption and Authentication

Version 2.0

AREA / Task Force: DRM

Status: External Proposed Specification

November 2007



ISMA SPECIFICATION LIMITATIONS AND CONDITIONS OF USE

LEGAL LIMITATIONS AND CONDITIONS OF USE

USERS OF THE ISMA SPECIFICATION ARE NOT PERMITTED OR AUTHORIZED TO STATE OR CLAIM THAT THEIR PRODUCTS OR APPLICATIONS COMPLY WITH THE SPECIFICATION, PENDING ISMA'S DEVELOPMENT AND IMPLEMENTATION OF A COMPLIANCE OR CERTIFICATION PROGRAM AND USER'S EXPRESS AGREEMENT WITH THE TERMS AND CONDITIONS THEREOF. BY REQUESTING OR USING THE SPECIFICATION, USER AGREES TO THIS LIMITATION AND CONDITION.

Table of Contents

1.0 Introduction	5
2.0 References	5
2.1 Normative	5
2.2 Informative	7
3.0 Terminology and Conventions.....	9
3.1 Conventions.....	9
3.2 Definitions / Glossary of Terms.....	9
3.3 Abbreviations	10
4.0 Goals and Requirements.....	11
5.0 ISMA Encryption and Authentication	12
5.1 Receiver Architecture and End-to-End Flows	12
5.2 Transforms.....	13
5.2.1 Default message authentication (integrity) transform	14
5.2.2 Default cipher and mode	14
6.0 ISMACryp File Structure.....	15
6.1 General principles	15
6.1.1 Sample transformation.....	15
6.1.2 Sample description transformation	15
6.2 ISMA Encryption.....	16
6.2.1 Encryption scheme.....	16
6.2.2 Encryption information	16
6.2.3 Sample transformation.....	17
6.3 Sub-samples.....	17
6.4 Bytestream format for encrypted AVC video stream	17
6.5 Usage of Brands	18
7.0 ISMACryp RTP Transport Packet Structure.....	19
7.1 General principles	19
7.2 Decryption process.....	20
7.3 Payload header	20
7.3.1 enc-isoff-generic header	20
7.3.2 enc-mpeg4-generic header	24
7.4 Encrypted payload modes.....	25
7.4.1 enc-isoff-generic mode.....	25
7.4.2 enc-mpeg4-generic mode	26
8.0 ISMACryp SDP Signaling	29
8.1 Overview	29
8.2 ISMACryp version identification	29
8.3 ISMACryp parameters.....	30
8.4 'enc-isoff-generic' SDP signaling.....	31
8.4.1 Codec identification.....	32
8.4.2 Codec initialization.....	32
8.4.3 Optional parameters.....	33
8.5 'enc-mpeg4-generic' SDP signaling	33
9.0 ISMACryp Encryption (Default) Cryptography Specification	35
9.1 ISMACryp AES-CTR Encryption Transform	35
9.1.1 Default cipher, mode, and configuration.....	35
9.1.2 Fixed parameters and signaling values.....	37
9.1.3 Transport packetization values.....	38
9.2 ISMACryp Message Authentication (Integrity) Transform.....	39
9.3 The Security of ISMACryp Cryptography.....	39
10.0 Name Assignment and Registration.....	41

Annex A: Local Playback (Informative)	42
Annex B: Encryption Process Example (Informative)	43
Annex C: 'enc-mpeg4-generic' SDP Examples (Informative)	47
Annex D: 'enc-isoff-generic' RTP packetization and SPD examples (Informative)	49
Annex E: 'enc-isoff-generic' RTP receiver behaviour in case of packet loss (Informative)	52
Annex F: Interoperability with OMA DRM Version 2.0 (Informative)	54
F.1 Overview	54
F.2 MPEG-4 File Structure.....	54
F.2.1 Sample description transformation.....	54
F.3 Transport Signaling	54
Annex G: Use of ISMACryp prior to OMA DRM 2.0 super-distribution (informative)	56
G.1 Introduction	56
G.2 ISMACryp streaming followed by OMA DRM 2.0 super-distribution	56
G.3 Requirements for ISMACryp streaming.....	57
G.4 Conclusion.....	58
Annex H: 'enc-isoff-generic' SVC protection (Informative)	59

1.0 Introduction

This document specifies content encryption, message authentication (integrity) services, an RTP payload format and a file format for pre-encrypted content for ISMA 1.0 [ISMASPEC], ISMA 2.0 [ISMASPEC2] and more generally any media that can be stored as elementary stream in an ISO media file [14496-12]. The official name for this service specification is "ISMA Encryption and Authentication" but it is unofficially called "ISMACryp" throughout much of this document.

The ISMACryp framework is extensible to new media encodings, can be upgraded to new cryptographic transforms, and is applicable to a variety of key management, security, or digital rights management (DRM) systems. ISMACryp defines a default encryption of media streams and authentication of media messages. The ISMA Encryption and Authentication transforms conform to the ISMA DRM Recommendations [ISMADRM].

Section 4 briefly describes requirements and architecture. Sections 5, 6, 7, 8 and 9 define ISMACryp cryptography, ISO file format, RTP payload format and SDP signaling.

2.0 References

2.1 Normative

[14496-1] ISO/IEC 14496-1

Information technology -- Coding of audio-visual objects --
Part 1: Systems,
Third Edition, November 2004.

[14496-8] ISO/IEC 14496-8

Information technology -- Coding of audio-visual objects --
Part 8: Carriage of ISO/IEC 14496 contents over IP networks,
First Edition, May 2004.

[14496-10] ISO/IEC 14496-10

Information technology -- Coding of audio-visual objects --
Part 10: Advanced Video Coding
2003
ITU-T Recommendation H.264 (2003): "Advanced video coding for generic audiovisual services"

[14496-12] ISO/IEC 14496-12|15444-12

Information technology -- Coding of audio-visual objects --
Part 12: ISO base media file format,
Information technology -- JPEG 2000 image coding system --
Part 12: ISO base media file format,
Second Edition, April 2005.

[14496-14] ISO/IEC 14496-14

Information technology -- Coding of audio-visual objects --
Part 14: MP4 file format,
First Edition, November 2003.

[AES] NIST FIPS 197

Advanced Encryption Standard (AES),
(<http://csrc.nist.gov/encryption/aes/index.html>),
(<http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf>).

[AES-C] NIST SP 800-38

Recommendation for Block Cipher Modes of Operation Methods and Techniques,
M. Dworkin, December 2001,
(<http://csrc.nist.gov/publications/nistpubs/800-38a/sp800-38a.pdf>).

[AES-CTR] NIST Workshop1

CTR-Mode Encryption,
H.Lipmaa, P.Rogaway, D. Wagner,

[ISMADRM] ISMA DRM

Recommendations Version 1.0, Internet Streaming Media Alliance, January 2002.

[ISMASPEC] ISMA Implementation Specification

Version 1.0, August 2001.

[ISMASPEC2] ISMA Implementation Specification

Version 2.0, April 2005.

[ISMACRYP11] ISMA Encryption and Authentication Specification

Version 1.1, July 2006.

[RFC1630] IETF RFC 1630

Uniform Resource Identifiers in WWW,
T. Berners-Lee, June 1994,
(<ftp://ftp.rfc-editor.org/in-notes/rfc1630.txt>).

[RFC2119] IETF RFC 2119

Key words for use in RFCs to Indicate Requirements Levels,
S. Bradner, March 1997,
(<ftp://ftp.rfc-editor.org/in-notes/rfc2119.txt>).

[RFC2327] IETF RFC 2327

SDP: Session Description Protocol,
M. Handley, V. Jacobson, April 1998,
(<ftp://ftp.rfc-editor.org/in-notes/rfc2327.txt>).

[RFC3016] IETF RFC 3016

RTP Payload Format for MPEG-4 Audio/Visual Streams,
Y. Kikuchi, et. Al, November 2000,
(<ftp://ftp.rfc-editor.org/in-notes/rfc3016.txt>).

[RFC3640] IETF RFC 3640

RTP Payload Format for Transport of MPEG-4 Elementary Streams,
J. van der Meer, D. Mackie, V. Swaminathan, D. Singer, P. Gentric, November 2003,
(<ftp://ftp.rfc-editor.org/in-notes/rfc3640.txt>).

[RFC3711] IETF RFC 3711
SRTP: The Secure Real Time Transport Protocol,
M. Baugher, D. McGrew, D. Oran, R. Blom, E. Carrera, M. Näslund, K. Normann, March 2004,
(<ftp://ftp.rfc-editor.org/in-notes/rfc3711.txt>).

[RFC4568] IETF RFC 4568
SDP Security Descriptions for Media Streams,
F.Andreasen, M.Baugher, D.Wing, July 2006
(<ftp://ftp.rfc-editor.org/in-notes/rfc4568.txt>).

2.2 Informative

[21000-5] ISO/IEC 21000-5
Part 5: Rights Expression Language,
First Edition, April 2004.

[21000-6] ISO/IEC 21000-6
Information Technology — Multimedia Framework —
Part 6: Rights Data Dictionary,
First Edition, May 2004.

[AMPBH] Proc. Security Protocols Workshop '97
Secure Books: Protecting the Distribution of Knowledge,
R.J.Anderson, V. Matyás Jr., F. Petitcolas, I.E. Buchan, R. Hanka, 1997,
(<http://www.cl.cam.ac.uk/users/rja14/>).

[CPRM] 4C Entity Content Protection for Recordable Media
(<http://www.4centity.com/docs/versions.html>).

[EPIC] Pretty Poor Privacy: An Assessment of P3P and Internet Privacy
EPIC, June 2000,
(<http://www.epic.org/Reports/pretypoorprivacy.html>).

[FFSS] Privacy Engineering in Digital Rights Management Systems
Revised papers from the ACM CCS-8 Workshop on Security and Privacy in DRM,
J. Fegenbaum, M.Freedman, T. Sander, A. Shostack, 2001,
(<http://www.homeport.org/~adam/privacyeng-wspdrm01.pdf>).

[IPSEC] IETF RFC 2411
IP Security Document Roadmap,
R. Thayer, N. Doraswamy, R. Glenn, November 1998,
(<ftp://ftp.rfc-editor.org/in-notes/rfc2411.txt>).

[Krawczyk] SKEME: A Versatile Secure Key Exchange Mechanism for Internet,
1996 Symposium on Network and Distributed System Security,
H. Krawczyk, February 1996.
(<http://www.csl.mtu.edu/cs6461/www/Reading/Krawczyk96.pdf>)

[MF00] Attacks on Encryption of Redundant Plaintext and Implications on Internet Security
Proceedings of the Seventh Annual Workshop on Selected Areas in Cryptography (SAC),
D. McGrew, S. Fluhrer, 2000,
(<http://www.mindspring.com/~dmcgrew/dam-srf-sac00.pdf>).

- [NNL] Revocation and Tracing Schemes for Stateless Receivers,
Lecture Notes in Computer Science,
D. Naor M. Naor, J. Lotspiech, March 2002
(<http://www.wisdom.weizmann.ac.il/%7Enaor/PAPERS/2nl.pdf>),
(http://www.securemulticast.org/GSEC/gsec1_Naor.pdf).
- [OFT] Key Establishment in Large Dynamic Groups Using One-Way Function Trees
D.McGrew, A.Sherman, 1998
- [OMADCFV2] OMA-TS-DRM-DCF-V2_0-20060303-A
DRM-DCF Specification V2.0 (Approved version), March 2006
(http://www.openmobilealliance.org/release_program/docs/CopyrightClick.asp?pck=DRM&file=V2_0-20060303-A/OMA-TS-DRM-DCF-V2_0-20060303-A.pdf)
- [OMADRMV2] OMA-TS-DRM-DRM-V2_0-20060303-A
DRM-DRM Specification V2.0 (Approved version), March 2006
(http://www.openmobilealliance.org/release_program/docs/CopyrightClick.asp?pck=DRM&file=V2_0-20060303-A/OMA-TS-DRM-DRM-V2_0-20060303-A.pdf)
- [OMARELV2] OMA-TS-DRM-REL-V2_0-20060303-A
DRM-REL Specification V2.0 (Approved version), March 2006
(http://www.openmobilealliance.org/release_program/docs/CopyrightClick.asp?pck=DRM&file=V2_0-20060303-A/OMA-TS-DRM-REL-V2_0-20060303-A.pdf)
- [RFC2104] IETF RFC 2104
HMAC: Keyed-Hashing for Message Authentication,
H.Krawczyk, M.Bellare, R.Canetti, February 1997,
(<ftp://ftp.rfc-editor.org/in-notes/rfc2104.txt>).
- [RFC2627] IETF RFC 2627
Key Management for Multicast: Issues and Architectures,
D.Wallner, E.Harder, R.Agee, June 1999,
(<ftp://ftp.rfc-editor.org/in-notes/rfc2627.txt>).
- [RFC3547] IETF RFC 3547
The Group Domain of Interpretation, IETF,
M.Baughner, T.Hardjono, H.Harney, B.Weis, July 2003,
(<ftp://ftp.rfc-editor.org/in-notes/rfc3547.txt>).
- [RFC3984] IETF RFC 3984
RTP Payload Format for H.264 Video,
S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund, D. Singer, February 2005,
(<ftp://ftp.rfc-editor.org/in-notes/rfc3984.txt>).
- [SECURITY] A concrete security treatment of symmetric encryption
Proceedings of the 38th Symposium on Foundations of Computer Science, IEEE,
M.Bellare, A.Desai, E.Jokipii, P.Rogaway, 1997,
(<http://www-cse.ucsd.edu/users/mihir/papers/sym-enc.html>).
- [RFCSDPLAYER] IETF draft-schierl-mmusic-layered-codec-04
Signaling of layered and multi description media in Session Description Protocol (SDP)
T. Schierl, December 2007, Work in Progress.
(<http://tools.ietf.org/id/draft-schierl-mmusic-layered-codec-04.txt>)

[SVCFF] ISO/IEC 14496-15/Amd.2

Information technology -- Coding of audio-visual objects -- Part 15: Advanced Video Coding (AVC) file format, AMENDMENT 2: File format support for Scalable Video Coding FPDAM, July 2007.

3.0 Terminology and Conventions

3.1 Conventions

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 [RFC2119].

All sections and appendixes, except “Introduction”, are normative, unless they are explicitly indicated to be informative.

3.2 Definitions / Glossary of Terms

Access Unit (AU)	An access unit is the smallest data entity to which timing information can be attributed
Authentication	See "Entity authentication" and "Message authentication"
Authorization	The process of allocating access to resources, such as key-management keying material, to an authenticated entity. Authorization is outside of the the scope of ISMA Encryption and Authentication.
Byte Stream Offset (BSO)	The position of a stream byte relative to the first stream byte, which has a BSO of zero. The default encryption transform defines and uses the BSO for the IV (see "Initialization Vector").
Confidentiality	Access control that is applied to media using ISMA Encryption and Authentication.
Elementary stream	An elementary stream contains only one kind of data, e.g. audio, video or closed caption. An elementary stream is often referred to as "elementary", "data", "audio" or "video" bitstreams or streams. The format of the elementary stream depends upon the codec or data carried in the stream, but will often carry a common header when turned into a packetized elementary stream .
Entity authentication	Entity authentication confirms that an entity (e.g. a rights holder) has possession of a secret associated with its identity. Key management procedures often perform entity authentication during key establishment such as in an authenticated key exchange [Krawczyk].
Initialization Vector (IV)	The cryptographic metadata needed by a payload-encryption and/or message-authentication transform in the ISMA Encryption and Authentication framework. The contents of the IV depend on the particular transform. The default ISMACryp encryption transform uses the BSO as the IV (See Section 10.0). ISMACryp default message-authentication does not use the IV.
Integrity	See "Message integrity" and "Message Authentication"
Media	Continuous-time data that share a common timebase.
Media authentication	Media authentication validates the integrity of the media data, independently of the message, and authenticates the rights holder who mastered and authenticated the data.
Media frame	The smallest clocked unit of media.

Message authentication	Message authentication validates that the received message is identical to what was sent by the sender. A transport security protocol performs message authentication using an integrity check of a message authentication code or through the verification of a digital signature.
Message integrity	See "Message authentication."
NAL Unit	MPEG-4 AVC Network Abstraction Layer Unit; see also "Slice"
Packetization	Process of assigning media frames or fragments of media frames to MPEG transport packets or to RTP packets before, during, or after encryption of the media data. Packetization is performed by a "packetizer."
Privacy	Access controls applied to the user's identity and activity on a network.
Sample	See "media frame"
Slice	Each slice is parsable (i.e. syntax decodable) independently of all other bytes in the access unit. In other words, if a decoder receives only a slice and not the whole access unit, it will be able to decode it. Note: The definition of slices matches with the definition of video-packets for MPEG-4 Part 2, slices for H.263, NAL units for MPEG-4 AVC and SVC.
Sub-Sample	See "Slice"

3.3 Abbreviations

4CC	Four character code
AU	Access Unit
BSO	Byte Stream Offset
CTS	Composition Time Stamp
DTS	Decoding Time Stamp
IV	Initialization Vector
LSB	Least Significant Bit
KMS	Key Management System
MAC	Message Authentication Code
MSB	Most Significant Bit
MTU	Maximum Transmission Unit
NALU	NAL Unit
PKC	Public Key Cryptography
RAP	Random Access Point
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

4.0 Goals and Requirements

The ISMA DRM Architecture, Goals and Requirements are unchanged from ISMACryp version 1.1 [ISMALCryp11] and not repeated in this version of the specification. The following figure briefly reviews the architecture. For more background information refer to the ISMACryp version 1.1 specification.

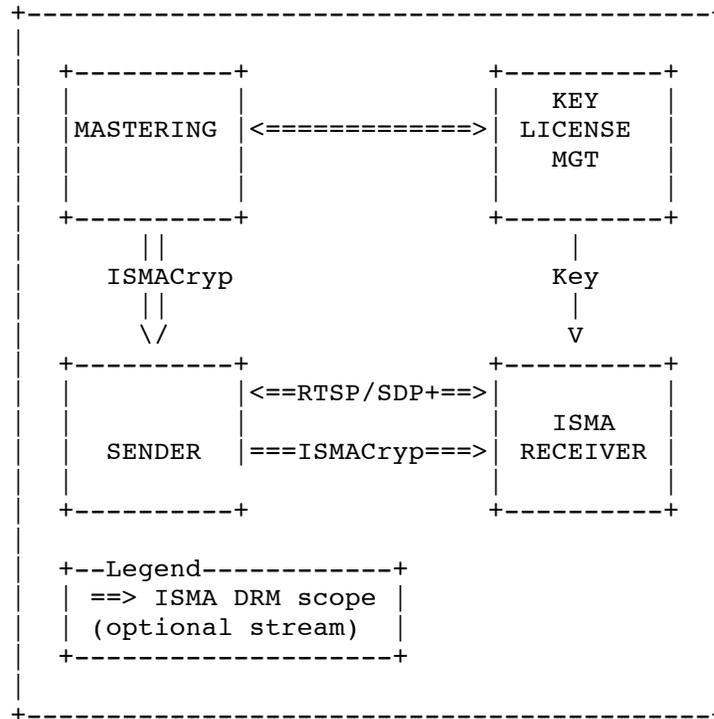


Figure 6.0-1: ISMA DRM Architecture

This version 2.0 of the ISMACryp specification adds the requirement to offer a generic solution for any elementary stream that can be stored in an ISO media file [14496-12]:

- It MUST be possible to encrypt any elementary stream that can be stored in an ISO media file; media encryption may take place either during mastering or sending. Transport MUST be designed to handle any of these encrypted medias.

5.0 ISMA Encryption and Authentication

This section specifies the ISMA Encryption and Authentication cryptographic framework. Nicknamed "ISMACryp" throughout this document, ISMA Encryption and Authentication is a family of cryptographic media encodings and protocols. Section 9 specifies the default encryption and authentication transforms for ISMACryp. Sections 5, 6, 7, 8, 9 form a complete specification for ISMACryp. "ISMACryp 2.0" targets all codecs that can be stored in ISO media files, and, in particular, codecs that are used by the ISMA 1.0 and 2.0 specifications.

5.1 Receiver Architecture and End-to-End Flows

Figure 5.1-1 shows the Receiver architecture in more detail with interfaces to Key/License Management (KEY MGT), an RTSP control interface, and ISMACryp, the cryptographic services for media data. The ISMACryp Receiver can decrypt, authenticate, and check the integrity of encrypted media data.

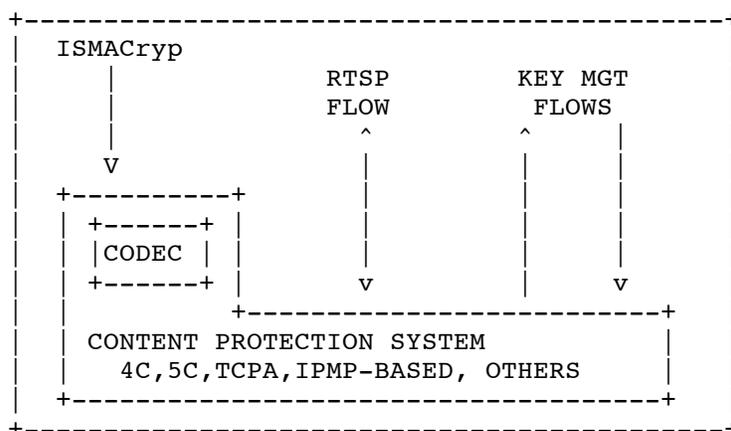


Figure 5.1-1: ISMACryp Receiver Architecture

Figure 5.1-1 also indicates the scope of the ISMACryp specification, which ends upon the decryption of the encrypted media stream. The decoding and presentation of the decrypted media is out of scope for this specification. Receiver content protection system and key management are also out of scope.

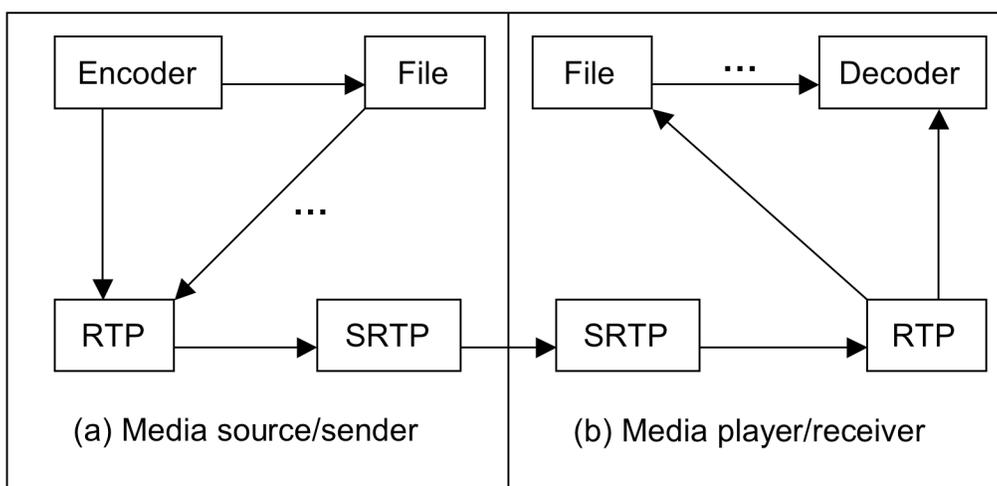


Figure 5.1-2: ISMACryp End-to-end Flows

Figure 5.1-2a shows the ISMACryp environment where a stream MAY be mastered into a file or streamed directly from an encoding application to a network. In both cases, encryption occurs prior to transport although message authentication SHOULD be performed at the transport layer. At the receiver (media player/decoder) of Figure 5.1-2b, a stream may be received into a file, such as a personal recorder at a player or cache server, or directly to a decoder. ISMACryp's transforms are applied at the arcs that emanate from the Figure 5.1-2 ENCODER; decryption occurs at the arcs that terminate at the DECODER; message authentication transforms are applied at the SRTP sender and receiver shown in Figure 5.1-2.

The following sections describe the cipher and bit-stream specifications for the ISMACryp specifications.

5.2 Transforms

ISMACryp supports the replacement of its encryption and message-authentication (integrity) transforms. This section identifies the default transforms and references their specification, which is in Section 9. It is possible to replace the ISMACryp default encryption or message authentication transform with different ones: the present (framework) specification will not need revision but Section 9 would be replaced with a new transforms specification (i.e. in an ISMACryp version later than version 2.0, which is revision-controlled by ISMA). In some cases, existing transforms MAY be augmented. For example, the default ISMACryp transforms include only message authentication, but media authentication can be added to ISMACryp without necessarily replacing message authentication.

The ISMACryp modular design relies on external standards where appropriate and is therefore suitable for the greatest variety of network environments. Thus ISMACryp uses SRTP for message authentication of real-time media packets [RFC3711]. A modular design also permits use of a variety of key management and key establishment systems.

In addition to message authentication, there is also media authentication (see Section 3.2). Authentication of the rights holder that creates or publishes a content work - and validation of the integrity of that work - can be an important function for many applications. Typically, digests from hash functions will suffice for the integrity check, and digital signatures serve to authenticate the creator of the digest or catalog of published works [AMPBH]. This solution is efficient for files that are reliably delivered and for which a single hash of the contents is feasible. Such a hash cannot be computed when parts of the file are streamed over a lossy channel.

Whereas public-key cryptography (PKC) is suitable to authenticate media data in files, this is inefficient for packet data: Practical security systems generally do not use asymmetric cryptography for packets owing to the excessive per-packet overhead of digital signatures or public-key encryption. The packet-size and computation overhead are worse for media frames since there are often multiple frames to a packet. There are, however, two methods more efficient than PKC to authenticate stored and streamed media data independently of the message.

The first method is to authenticate each media frame using symmetric keys. Use of a message authentication code (MAC) is arguably feasible for large media frames, but these frames may be fragmented across transport packets thus causing additional complications. In fact, even symmetric authentication techniques are generally infeasible for small media frames since a message authentication code (MAC) can add 10 or more bytes to the length of each media frame [RFC2104]. This technique has orders of magnitude less overhead than PKC and is not excessive for large (~500 byte) video frames. But it is excessive for small frames and low bit-rate audio and video data.

The second method avoids the problem of authenticating fragments of media frames by pre-assigning media data into packet payloads, which are authenticated independently of the transport packet. This pre-assignment assumes that the size of the transport-packet payload is fixed prior to the time of transport.

Even the second choice, which has much less overhead than the first, usually doubles the amount of MAC data in each transport packet. This doubling is unavoidable if both message authentication and media authentication are desired. Of the two, message authentication is chosen as more important for several reasons. First, the receiver must trust the sender to have rights to disseminate the media and this trust relationship is realized in message authentication. Second, the authentication of the file creator can be no better than that of the sender whom the creator has authorized to disseminate the work. Third, it is not necessarily the receiver's responsibility to ensure that what the sender sends is exactly what the file creator authorized it to send. Finally, the sender may have rights to alter the media in various ways. Thus, media authentication is redundant to message authentication for many practical applications. The preferred packet design of section 7, therefore, uses SRTP message authentication only and does not support media authentication independent of the SRTP message.

5.2.1 Default message authentication (integrity) transform

SRTP message authentication [RFC3711] SHOULD be used for ISMACryp messages. This transform is described in Section 9.

5.2.2 Default cipher and mode

The AES [AES] in Counter mode [AES-C, AES-CTR] SHALL be used as the encryption cipher. This cipher is described in Section 9.

6.0 ISMACryp File Structure

The ISMACryp file format transformation supports the encryption of files either for local playback (including file download) or prior to hinting for streaming. The transformation is self-contained; all the information needed to either play the file, or hint it for streaming (including generating SDP information) is in the file. This does not mean, of course, that the file contains, for example, all the keys; but it does mean that enough information MAY be included to identify the KMS used and to enable a client to contact it and acquire the correct set of keys.

The file transformation involves:

- a) transforming the media samples themselves (encrypting them);
- b) transforming the description of the media samples, both to document the transformation of the media samples, and to avoid the encrypted samples being read as if they were in the clear.

The file may be optionally hinted, and the extra signaling MUST be generated in the hint tracks.

6.1 General principles

This section documents the general format of encrypted media and the principles applied.

6.1.1 Sample transformation

In encryption, the samples are transformed – encrypted – so that the underlying media cannot be accessed by readers without the appropriate information (e.g. keys). The format of the encrypted samples is "owned" and documented by the encryption system.

6.1.2 Sample description transformation

The purpose of the sample description transformation is twofold: The sample description prevents accidental treatment of encrypted data as if it were un-encrypted and documents the transforms applied.

The transformation of the sample description is described entirely by the following procedure as defined in chapter 8.45 of [14496-12]:

1. The 4CC of the sample description is replaced with a 4CC indicating the encryption: 'encv' for "encrypted video" stream (instead of e.g. 'mp4v', 'avc1'), 'enca' for "encrypted audio" stream (instead of e.g. 'mp4a', 'samr'), 'enct' for "encrypted text" stream and 'encs' for encrypted "systems" stream.
2. A ProtectionSchemeInfoBox 'sinf' is appended to the sample description, leaving all other boxes unmodified. The ProtectionSchemeInfoBox contains all the information required both to understand the encryption transform applied and its parameters, and also to find other information such as the kind and location of the key management system. It also documents the original (unencrypted) format of the media.
3. The original format 4CC of the track is stored in the OriginalFormatBox 'frma' that is a sub-box of the ProtectionSchemeInfoBox 'sinf'.
4. The SchemeTypeBox 'schm' is also a sub-box of the ProtectionSchemeInfoBox and specifies the encryption scheme as 4CC and its version. Additionally, this box may contain an optional URI that directs the user to a web-page if they do not have the scheme installed on their system.
5. Finally, the SchemeInformationBox 'schi' that is also a sub-box of the ProtectionSchemeInfoBox contains any information the protection system needs to store. The SchemeInformationBox is a container box that is only interpreted by the scheme being used. The content of this box is a series of boxes whose type and format are defined by the scheme declared in the SchemeTypeBox.

Note: The sub-boxes of the ProtectionSchemeInfoBox may occur in any order.

6.2 ISMA Encryption

In the definitions which follow, the value n in $\text{bit}(n)$, $\text{unsigned int}(n)$ and $\text{int}(n)$ is always a bit count.

6.2.1 Encryption scheme

The AES-CTR-128 mode (Section 9) used by ISMA uses the `scheme_type` "iAEC" in the `SchemeTypeBox` with a `scheme_version` of "1" for ISMACryp, which is revision-controlled by ISMA.

6.2.2 Encryption information

This section describes, for the ISMACryp scheme, how to convey similar parameters to some of those in Table 8.3.1. All boxes defined in this section are stored as sub-boxes of the `SchemeInformationBox` and shall only be used if the ISMACryp `scheme_type` "iAEC" is used in the `SchemeTypeBox`.

For example, following information is required in the `SchemeInformationBox`:

- a) the identification of the Key Management System (KMS) used, its URI and KMS version
- b) the description of the format of the samples when in the file format; this includes the presence of a selective-encryption indicator, the size of the `key_indicator`, and the size of the initial-offset.

The KMS supplies the keying material. The "string" used for `KMS_URI` below is a null-terminated string.

```
aligned(8) class ISMAKMSBox extends FullBox('iKMS', version, 0) {
    if (version==0) {
        string KMS_URI;           // the KMS URI which the hinter or server
                                // MAY add to the ISMACryp SDP information
    } else { // version ==1
        unsigned int(32) KMS_ID;   // 4CC identifying the KMS
        unsigned int(32) KMS_version; // KMS version
        string kms_URI;           // the KMS URI which the hinter or server
                                // MAY add to the ISMACryp SDP information
    }
}
```

Note: Writers should use version 0 to be compatible with ISMACryp 1.0 readers and should use version 1 if extended KMS information is needed.

```
aligned(8) class ISMASampleFormatBox extends FullBox('iSFM', 0, 0) {
    bit(1) selective_encryption; // see Section 8.1
    bit(7) reserved;           // MUST be zero
    unsigned int(8) key_indicator_length; // see Section 8.1
    unsigned int(8) IV_length; // see Section 8.1
}
```

Other additional information about the ISMA scheme may be required. It is possible to add an `ISMACrypSaltBox` to convey the salt key used in the media encryption. This parameter is similar to the `fntp` parameter defined in Table 8.3.2. This is an OPTIONAL sub-box of the `SchemeInformationBox`.

```
aligned(8) class ISMACrypSaltBox extends FullBox('iSLT', 0, 0) {
    unsigned int(64) salt; // see Section 8.1, MUST be non null
}
```

6.2.3 Sample transformation

ISMACryp adopts the approach of embedding ISMACryp signaling information (Section 8.1) in the media data. While this scheme has redundant data, which is bad, it does not require redefinition of the ISO file format. In this scheme, the samples are encrypted using a key from the key-set and using the ISMACryp default encryption in ISMACryp or some other encryption transform (in a future ISMACryp version). In order to permit random access, editing, and hinting, without scanning the file, we add an ISMACryp sample header to each sample that contains the following parameters as defined in Section 8.1:

- a) selective encryption indicator;
- b) key indicator;
- c) the initialization vector (for the ISMACryp default encryption, the initial counter value).

```
aligned(8) class ISMACrypSample {
    if (selective_encryption == 1) { // from the sample description
        bit(1) sample_is_encrypted;
        bit(7) reserved;           // must be zero
    }
    else sample_is_encrypted = 1;

    if (sample_is_encrypted==1) {
        unsigned int(8 * IV_length)      IV;
        unsigned int(8 * key_indicator_length) key_indicator;
    }
    unsigned int(8) data[]; // encrypted media data, to end of sample
}
```

When selective-encryption (from the ISMASampleFormatBox) is zero, there is no storage allocated for the fields sample-is-encrypted and reserved. When no storage is allocated for the sample-is-encrypted field, the "else sample-is-encrypted = 1" refers to a local variable named "sample-is-encrypted" and not the field named "sample-is-encrypted". If key_indicator_length is zero (0), then the key_indicator is also always zero (0). This means a single key is being used for the stream.

6.3 Sub-samples

Some video codecs (like AVC [14496-10]) allow encoding an Access Unit (a video frame) as slices that can be decoded independently. If such video frames are encrypted, it is not possible to identify the slice boundaries by parsing for e.g. slice start codes or length fields that are part of the Access Unit.

In Section 7.4.1.1 of this specification a fragmentation mode is defined that the packetizer (hinter or server) MAY use to align slice boundaries with AU fragment boundaries. To allow such packetization, the use of the SubSampleInformationBox 'subs' (as defined in chapter 8.42 of [14496-12]) is RECOMMENDED in the encrypted file to mark the slice boundaries in the encrypted Access Unit.

The ISMACryp sample header as defined in chapter 6.2.3 MUST be considered as the first sub-sample.

6.4 Bytestream format for encrypted AVC video stream

Alternatively to the "length-field mode", the byte-stream format may optionally be used for encrypted AVC video (see Section 7.4.2.3).

Note:

- This mode MUST be used for 'enc-mpeg4-generic' stream delivery.
- This mode MAY be used for 'enc-isoff-generic' stream delivery (backward compatibility)

In this case, the 'original format' indicator for the stream MUST be changed from 'avc1' to '264b'.

Unencrypted streams **MUST NOT** be stored in files using the byte-stream format as this is not the standard format for AVC. The 'avc1' sample-entry name **MUST NOT** be used for byte-stream structured AVC, either as a sample-entry name or as an original-format name.

Note: The byte-stream format [14496-10 Annex B] allows two types of start-codes: "four bytes 00 00 00 01" and "three bytes 00 00 01". Since the AVC File format [14496-15] does not permit 3-byte length fields, the byte-stream **MUST** use 4-byte start-codes (to enable easy transformation back into length fields after decryption). In other words, the fields "leading_zero_8bits" and "trailing_zero_8bits" **MUST NOT** be present and the field "zero_byte" **MUST** be present in each NAL unit. After decrypting the stream, two consecutive start-codes are needed to discover the length of each NAL unit.

In the `AVCDecoderConfigurationRecord`, the `LengthSizeMinusOne` field indicates the length of start-codes and **MUST** be set to 3 (corresponding to 4-byte start-codes).

```
aligned(8) class AVCDecoderConfigurationRecord {
    unsigned int(8) configurationVersion = 1;
    unsigned int(8) AVCProfileIndication;
    unsigned int(8) profile_compatibility;
    unsigned int(8) AVCLevelIndication;
    bit(6) reserved = '111111'b;
    unsigned int(2) lengthSizeMinusOne = 3;
    bit(3) reserved = '111'b;
    unsigned int(5) numOfSequenceParameterSets;
    for (i=0; i< numOfSequenceParameterSets; i++) {
        unsigned int(16) sequenceParameterSetLength;
        bit(8*sequenceParameterSetLength) sequenceParameterSetNALUnit;
    }
    unsigned int(8) numOfPictureParameterSets;
    for (i=0; i< numOfPictureParameterSets; i++) {
        unsigned int(16) pictureParameterSetLength;
        bit(8*pictureParameterSetLength) pictureParameterSetNALUnit;
    }
}
```

6.5 Usage of Brands

The brand 'isc2' indicates that support for ISMACryp2.0 is required. However, this brand is silent about what support is required after decryption. It is recommended that any integration specification using ISMACryp2.0 as the protection layer, also have a new brand which indicates that support both for ISMACryp2.0 and the appropriate codec(s) etc. is required. In this case, the 'isc2' is no longer needed as the new brand is more specific than the 'isc2' brand.

Note that once a file is encrypted, the brands documenting the support for the codecs by themselves would normally be removed. For example, a file readable under the 'mp42' brand is no longer readable under that brand once encrypted, and the 'mp42' brand would be removed from the compatible brands list. Therefore a common brand list would be a brand indicating the required encryption and probably codec support (an integration brand, as above, or 'isc2'), and the appropriate structural brand (e.g. 'iso2').

7.0 ISMACryp RTP Transport Packet Structure

This section defines the RTP payload formats for content encrypted according to this specification. The reader needs to be familiar with the 'mpeg4-generic' packet format [RFC3640] in order to understand the ISMACryp packet design.

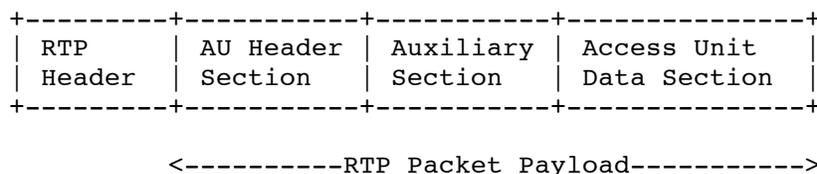


Figure 7.0-1: Data sections within an mpeg4-generic RTP packet

7.1 General principles

In RTP, a payload format is generally designed for the specific media being carried. ISMACryp uses a generic approach for encrypted content which is as much as possible codec agnostic.

Each ISMACryp RTP packet SHALL contain either:

1. Exactly one access unit,
2. Two or more complete access units, or
3. One fragment of an access unit.

Two payload formats are defined:

1. 'enc-isoff-generic' format :

This supports any media type that can be stored as elementary stream in ISO Media File format based files. It is derived from the 'mpeg4-generic' specification for MPEG-4 content [RFC3640]. It is also known as "ISMACryp 2.0".

2. 'enc-mpeg4-generic' format :

This is a subset of 'enc-isoff-generic' and targets ISMA 1.0 payloads (MPEG-4 Video, AAC and CELP MPEG-4 Audio) and ISMA 2.0 payloads (MPEG-4 AVC, AAC and HE-AAC MPEG-4 Audio). It is also known as "ISMACryp 1.1" and fully specified in the previous ISMACryp version 1.1 [ISMALCryp11].

The payload format 'enc-isoff-generic' is RECOMMENDED for any new application. The payload format 'enc-mpeg4-generic' should only be used if backward compatibility with existing ISMACryp 1.1 applications is desired.

Receivers that conform to the ISMACryp2.0 specification MUST support both payload formats 'enc-isoff-generic' and 'enc-mpeg4-generic'.

The ISMACryp payload formats specified here try to preserve optimal packetization and loss recovery while acknowledging that the packetization or the de-packetization processes might not have access to media data, which is encrypted at these stages. For the 'enc-mpeg4-generic' payload format, the media data is invisible to the packetization process except for access unit boundaries, which are available to the packetization process. Therefore, the 'enc-mpeg4-generic' payload format imposes specific restrictions at the access-unit level, such as allowing or disallowing fragmentation, or interleaving of access units. The 'enc-isoff-generic' payload format adds support for identifying slice boundaries even in encrypted form. No restrictions that require access to unencrypted media are imposed here, e.g., requiring that video frames be split at video packet boundaries.

7.2 Decryption process

Each packet is sent according to the ISMACryp payload format to the receiver, who must have all the needed cryptographic data to decrypt the packet. Upon decryption, the receiver must have all the needed information to process the packet. ISMACryp packetization accomplishes this by inserting an initialization vector (IV) in each packet, and the IV contains all the needed information to decrypt each access unit (AU) contained in the packet. How this is done is specific to the cipher. See Section 9 for operational details for the default cipher (where the IV is instantiated as a BSO). It is REQUIRED that ISMACryp ciphers accommodate the loss, delay, and reordering of a packet stream. The IV contains all cryptographic data that are needed to make the decryption of a packet independent of previous or successive packets in a stream.

7.3 Payload header

The RTP payload structures defined in this specification are based upon the RTP payload format defined in mpeg4-generic [RFC3640]. Support for encrypted media is enabled by adding new fields to the access unit (AU) header section. These new fields are defined in the following section. For a description of all other fields refer to [RFC3640].

7.3.1 *enc-isoff-generic* header

ISMACryp inserts cryptographic metadata at the beginning of each AU header. The format of the first AU header is different from the second and subsequent AU headers (similar to the treatment of AU-Index and AU-Index-Delta in mpeg4-generic [RFC3640]). This block supplies the Cryptographic context for each AU or AU fragment in an RTP packet and is defined as follows:

```
class ISMACrypContextAU(int auNum)
{
    if (ISMACrypSelectiveEncryption || ISMACrypSliceIndication ||
        ISMACrypPaddingIndication )
        //ISMACryp Header Byte exists if at least one of these fields is used
        {
            bit(1) AU_is_encrypted;
            bit(1) Slice_start;
            bit(1) Slice_end;
            bit(3) Padding_bitcount;
            bit(2) Reserved;
        }

    if (ISMACrypExtensionHeaderSize !=0) // default value is "0" bytes
    {
        // no ISMACrypExtensionHeader in this version of ISMACryp
        bit(ISMACrypExtensionHeaderSize*8) extension_header;
    }

    if (auNum==0) // First AU in packet?
    {
        unsigned int(ISMACrypIVLength*8)          initial_IV;
        unsigned int(ISMACrypKeyIndicatorLength*8) key_indicator;
    }
    else
    {
        int(ISMACrypDeltaIVLength*8)              delta_IV;
        if (ISMACrypKeyIndicatorPerAU)
            unsigned int(ISMACrypKeyIndicatorLength*8) key_indicator;
    }
}
```

Note: in bit(n), unsigned int(n) and int(n), n is always a bit count. Note also that delta_IV is the only signed int in the above definition.

See Section 8 for the signaling of the parameter constants:

- ISMACrypSelectiveEncryption
- ISMACrypSliceIndication
- ISMACrypPaddingIndication
- ISMACrypExtensionHeaderSize
- ISMACrypIVLength
- ISMACrypKeyIndicatorLength
- ISMACrypDeltaIVLength
- ISMACrypKeyIndicatorPerAU

All "Reserved" fields MUST be zero and SHOULD be ignored by the receiver. The first byte of ISMACrypContextAU MUST be present if one of the fields "AU_is_encrypted", "Slice_start", "Slice_end" and "Padding_bitcount" is used. Any of these fields SHOULD be ignored by the receiver if their usage is switched off by the corresponding parameters (ISMACrypSelectiveEncryption, ISMACrypSliceIndication, ISMACrypPaddingIndication) in the SDP description. The first byte is omitted if none of these fields is switched on through their corresponding SDP parameters.

Note: For backward compatibility of 'enc-isoff-generic' packets with 'enc-mpeg4-generic' receivers, ISMACrypSelectiveEncryption should also be switched on if ISMACrypSliceIndication and/or ISMACrypPaddingIndication is switched on, even if it is not intended to use the selective encryption feature. In this case, the AU_is_encrypted (SE) field should always be set to 'one'.

In the following table, the columns indicate the following fields:

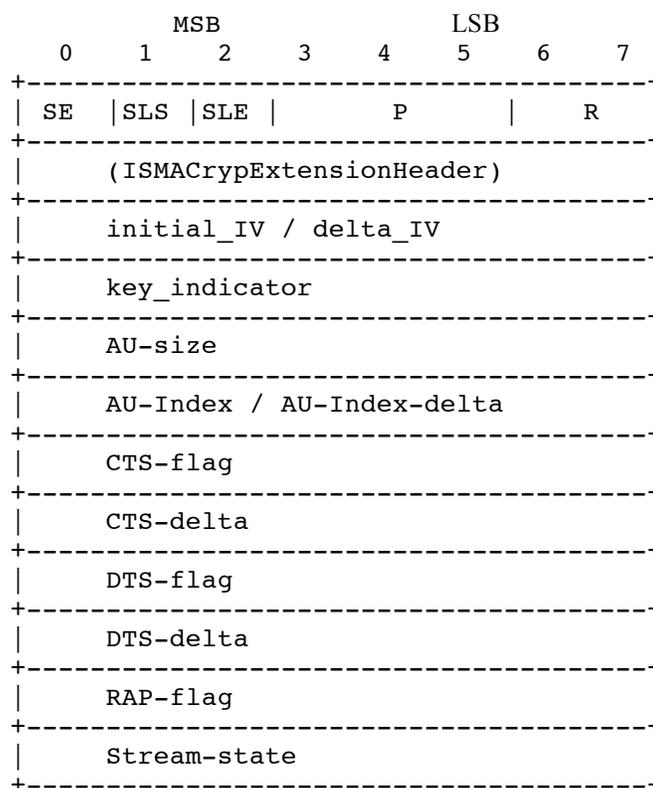
- SE AU_is_encrypted;
- SLS: Slice_start;
- SLE: Slice_end;
- P: Padding_bitcount;
- R: Reserved

Case	Functionality		SE	SLS	SLE	P	R
1	SelectiveEncryption Slices boundaries Padding	not used not used not used	∅ ⁽¹⁾	∅ ⁽¹⁾	∅ ⁽¹⁾	∅ ⁽¹⁾	∅ ⁽¹⁾
2	SelectiveEncryption Slices boundaries Padding	used not used not used	0/1 ⁽²⁾	0	0	0	0
3	SelectiveEncryption Slices boundaries Padding	not used used not used	1	0/1 ⁽³⁾	0/1 ⁽⁴⁾	0	0
4	SelectiveEncryption Slices boundaries Padding	not used not used used	1	0	0	[0..7] ⁽⁵⁾	0
5	SelectiveEncryption Slices boundaries Padding	used used not used	0/1 ⁽²⁾	0/1 ⁽³⁾	0/1 ⁽⁴⁾	0	0
6	SelectiveEncryption Slices boundaries Padding	used not used used	0/1 ⁽²⁾	0	0	[0..7] ⁽⁵⁾	0

7	SelectiveEncryption Slices boundaries Padding	not used used used	1	0/1 ⁽³⁾	0/1 ⁽⁴⁾	[0..7] ⁽⁵⁾	0
8	SelectiveEncryption Slices boundaries Padding	used used used	0/1 ⁽²⁾	0/1 ⁽³⁾	0/1 ⁽⁴⁾	[0..7] ⁽⁵⁾	0

- (1) the field does not exist in this case
- (2) "1" if data is encrypted, "0" if not (see below)
- (3) "1" if it is the first fragment of the slice or if is a complete slice, "0" if not (see below)
- (4) "1" if it is the last fragment of the slice or if it is a complete slice, "0" if not (see below)
- (5) number of bits for the padding (can be from 0 to 7 bits) (see below)

Diagrammatically, this means that the above specified fields are inserted just before the AU-size field. The complete AU Header for each AU is defined in 'enc-isoff-generic' as follows:



The ISMACrypExtensionHeader MUST not be present in this version of the specification. The corresponding SDP parameter "ISMACrypExtensionHeaderSize" is defined for future extensions and MUST be set to zero in this version of the specification. Future versions may define a size>0 and thus add an ISMACrypExtensionHeader in steps of complete bytes at the above specified position in ISMACrypContextAU. An ISMACryp 2.0 receiver MUST be able to ignore ISMACrypExtensionHeader.

Note: it is possible to compute the access unit count by using the configuration parameters, and the signaled length of the access unit headers. This is because the total bit-length of the AU-headers is given in each packet, and the length of the first AU Header as well as the second and subsequent AU-headers can be computed from the signaled parameters. This is still true with this extended AU header. So we have:

$$\text{AU-count} = (\text{AU-header-length} - \text{first-header-size}) / \text{subsequent-header-size} + 1;$$

Note: this equation does not hold if either CTS or DTS timestamps are present; however, this normally applies only to video, and in that case, the payload format restricts the packet to containing only one AU or a fragment of an AU.

The fields in the ISMACrypContextAU structure have the following meaning:

`AU_is_encrypted (SE)`

An optional single bit field to signal selective encryption. A value of 1 signals that the corresponding access unit is encrypted, a value of 0 means it is not. The presence of this field is configured with the 'ISMACrypSelectiveEncryption' parameter. All fragments of a single access unit MUST have the same value for `AU_is_encrypted`.

Note: if this field is not present, all access units are encrypted.

`Slice_start (SLS)`

This field and the following 'Slice_end' are two optional single bit fields to signal slice boundaries. The presence of these fields is configured with the 'ISMACrypSliceIndication' parameter. 'Slice_start' is an optional single bit field to indicate whether the payload contains the first fragment of a slice. A value of 1 signals that it is the first fragment of a slice or that the slice is not fragmented, a value of 0 means it is not.

`Slice_end (SLE)`

An optional single bit field to indicate whether the payload contains the last fragment of a slice. A value of 1 signals that it is the last fragment of a slice or that the slice is not fragmented, a value of 0 means it is not. As described above, the presence of this field is configured with the 'ISMACrypSliceIndication' parameter.

`Padding_bitcount (P)`

This field contains the number of padding bits at the end of the AU Data Section. The presence and the size of this field are configured with the 'ISMACrypPaddingIndication' parameter.

`Reserved (R)`

The last 2 bits (R) of the first byte are reserved for future use, their values MUST be 0.

`initial_IV`

Contains the initial IV value for the first access unit or fragment contained in the packet. In most cases, this is the only IV in the packet. In some cases such as interleaved media, however, there MAY be an IV per AU. See Section 10 for `initial_IV` definitions for the default transform.

`delta_IV`

This field contains IV data on a per-AU basis when 'ISMACrypDeltaIVLength' is non-zero and the data is interleaved in packet payloads.

`key_indicator`

Contains the key indicator for an access unit when 'ISMACrypKeyIndicatorLength' is non-zero. If 'ISMACrypKeyIndicatorPerAU' is 0, then only the first access unit in a packet has an explicit key indicator value included in the cryptographic context and all subsequent access units SHALL implicitly have the same value for `key_indicator` as the first access unit. If 'ISMACrypKeyIndicatorPerAU' is 1, then a value of `key_indicator` is included in the cryptographic header for each access unit or fragment in the packet. If `AU_is_encrypted` is 0 for an access unit, then the value of this field SHALL be ignored.

The actual IV to be used for each access unit in a packet is computed as follows, with the first access unit in a packet indexed as zero:

```
IV[0]    := AUHeader[0].Initial_IV;           // First AU in packet
IV[N+1] := IV[N] + AUsSize[N]
          + (ISMACrypDeltaIVLength == 0 ? 0 :
             AUHeader[N+1].delta_IV)        // Subsequent
```

Note: The number of access units in a packet is not signaled in this payload format. The number of access units in the packet can be deduced from the access unit header as for unencrypted modes. A packet that has the marker bit cleared in the RTP header contains a fragment that is not the last of an AU. If the marker bit is set, then the packet contains one or more access units or the last fragment of an access unit. The access unit header indicates whether there are two or more access units in the packet. To distinguish between one whole AU and the last fragment, compare the payload data size and the access unit size conveyed in the access unit header. The access unit size will be the size of the whole AU and not the fragment.

Note: In the simple case where there is one AU per packet, or the AUs are contiguous, this structure reduces to signaling a key indicator and an initial IV per packet.

Note: In the case of selective encryption, if the AU is not encrypted, the initial_IV/delta_IV and the key_indicator fields are still present but these values are not needed by the receiver as AU is not encrypted.

Note: It is out of scope for ISMACryp how the key_indicator indicates the decryption key. It is up to the KMS to make this linkage.

7.3.2 *enc-mpeg4-generic* header

The ‘enc-mpeg4-generic’ payload format is a subset of ‘enc-isoff-generic’. The ‘enc-mpeg4-generic’ header differs from the ‘enc-isoff-generic’ header in the usage of the first byte. Only the “AU_is_encrypted” field is used, all other bits are unused and thus “Reserved”. The first byte is only present if “ISMACrypSelectiveEncryption” is used.

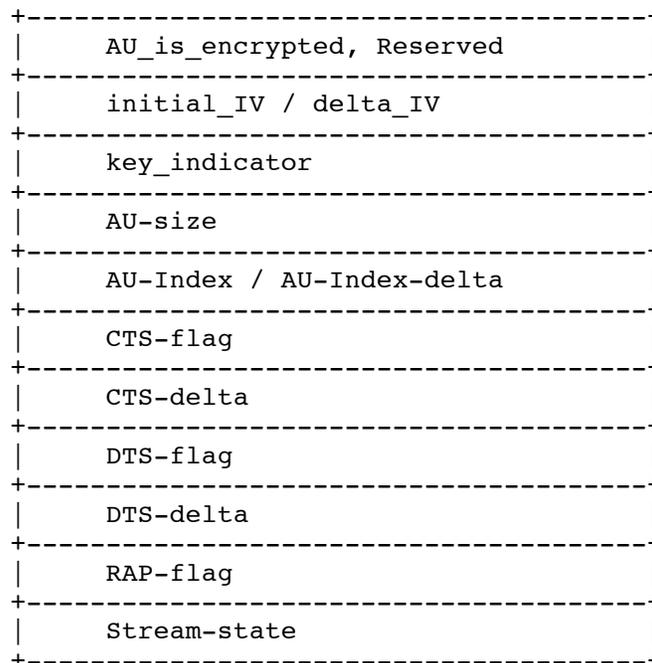
The ISMACrypContextAU is therefore defined as follows:

```
class ISMACrypContextAU(int auNum)
{
    if (ISMACrypSelectiveEncryption)
    {
        bit(1) AU_is_encrypted;
        bit(7) Reserved;
    }
    else AU_is_encrypted = 1;

    if (auNum==0) // First AU in packet?
    {
        unsigned int(ISMACrypIVLength*8)      initial_IV;
        unsigned int(ISMACrypKeyIndicatorLength*8) key_indicator;
    }
    else
    {
        int(ISMACrypDeltaIVLength*8)          delta_IV;
        if (ISMACrypKeyIndicatorPerAU)
            unsigned int(ISMACrypKeyIndicatorLength*8) key_indicator;
    }
}
```

All "Reserved" fields MUST be zero and SHOULD be ignored by the receiver.

The complete corresponding AU header diagram is as follows:



7.4 Encrypted payload modes

7.4.1 enc-isoff-generic mode

The 'enc-isoff-generic' mode supports any media type that can be stored as elementary stream in ISO Media File format based files. In general, the encrypted media data of the samples as described in Section 6 are stored in the Data Section of the RTP packet payload. Two optional features that may be used during packetization are described in this section. These two features apply to fragmentation of Access Units into several fragments and to padding of Access Units to byte boundaries.

Note: For MPEG-4 AVC video [14496-10], two flavors may be used: the generic encrypted "ISO file format" samples (= 'avc1' for AVC video) as described above (including NAL Unit length fields), or the byte stream format '264b' as described below in Section 7.4.2.3 (including start codes).

Note: If it is desired that parameter set NAL units are inserted in the stream for MPEG-4 AVC video [14496-10], they should be inserted in appropriate AU(s) before encryption (the parameter set NALUs have to be concatenated with the video data NALUs). For the generic encrypted "ISO file format" mode ('avc1'), a length code has to be inserted before the parameter set NALU. For the '264b' byte stream mode, a start code has to be inserted.

Note: The enc-isoff-generic format is derived from the format specified in [RFC3640] and only adds signaling of padding bits, slice-start and slice-end to the AU-Headers in this version of the specification. Therefore some features available in codec-specific payload formats may not be supported like e.g. NALU-type signaling as in [RFC3984].

7.4.1.1 Slices

Modern video codecs are able to encode video frames as independently decodable slices. In RTP packetization it is desirable to align packet boundaries with slice boundaries, e.g. for better error resilience.

If the media is encrypted prior to packetization, it is usually not possible for the packetizer to identify the slices boundaries, e.g. by parsing the access unit. However, alignment is possible in these two cases:

1. Packetization and encryption occur together (live encoder, combined encryption-hinter, etc.).
2. The encrypted file uses the SubSampleInformationBoxes to indicate slice boundaries (see Section 6.3).

The optional optimized fragmentation mode aligns slice boundaries with AU fragment boundaries and uses the two fields 'Slice_start' and 'Slice_end' in the AU header section to indicate the first and the last fragment of a slice:

- 'Slice_start' MUST be set to 1 if the payload contains the first fragment of a slice and the payload MUST start with the first byte of the slice.
- 'Slice_end' MUST be set to 1 if the payload contains the last fragment of a slice.
- If a RTP packet carries one or several complete slices, 'Slice_start' and 'Slice_end' MUST be set to 1.

If a slice is greater than the MTU then it SHOULD be fragmented into multiple RTP packets and the payload of the first packet MUST start with the first byte of the first fragment. Each RTP packet MUST carry either one or more complete slice(s) or one fragment of a slice.

7.4.1.2 Padding

Usually the codec handles byte-alignment of Access Units (e.g. all MPEG-4 codecs). In this case, the decoder is able to decode Access Units that are padded at the end of the AU to the next byte boundary. For codecs that do not support this and rely on Access Units that are not byte aligned, the RTP depacketizer has to handle the "de-padding" process. These padding bits need to be stripped after decryption and before decoding. For such codecs the number of padding bits can be signalled in the optional "Padding_bitcount" field in the header section.

7.4.2 *enc-mpeg4-generic mode*

The 'enc-mpeg4-generic' payload mode supports MPEG-4 Audio codecs (AAC, HE-AAC, CELP) and MPEG-4 Video codecs (Part-2 and Part-10/AVC) as used in the ISMA 1.0 and 2.0 specifications.

Note: The formats defined below are also applicable to 'enc-isoff-generic', there is no difference in the packet payload (only in SDP signaling as described in Section 8).

7.4.2.1 Encrypted MPEG-4 audio

Encrypted MPEG-4 audio is handled the same way as unencrypted audio as defined in the mpeg4-generic specification [RFC3640]:

- **MPEG-4 AAC Low and High Bit Rates.** These two modes support encrypted MPEG-4 AAC (including HE-AAC).
- **MPEG-4 CELP CBR and VBR.** These two modes support encrypted constant and variable bit rate MPEG-4 CELP Audio.

7.4.2.2 Encrypted MPEG-4 Part-2 video

This mode defines the payload format for encrypted MPEG-4 Part-2 Video. To enable the carriage of encrypted video, a new mode is defined here for its carriage solely in the context of this encrypted payload format.

This mode is signaled by mode=mpeg4-video. In this mode, the configuration of the payload is as follows:

- SizeLength = 0. Each packet contains only one AU or AU fragment.
- IndexLength = 0 and IndexDeltaLength = 0. No interleaving.
- CTSDeltaLength = 0. No signaling of CTS needed as it is obtained from the RTP timestamp.
- DTSDeltaLength = 22. Video access units may have a DTS that differs from the CTS when B-frames are present, therefore the DTS can be signaled in this mode.
- RandomAccessIndication = 1. The RAP-flag is set to 1 to indicate video I-frames.
- Config. Must be present.

At most one access unit or fragment appears in an RTP packet. Multiple access units in a packet are not allowed. The rules for packetization specified in RFC 3016 are not required in this mode. An access unit may be split anywhere at all, without regard to video packet, video header, or any other boundaries.

For this mode, the AU header takes one of two sizes (not counting the cryptographic metadata):

1. If the DTS is equal to the CTS, the DTS-flag is set to zero. In this case, the AU header consists of only two bits (the DTS-flag followed by the RAP-flag), which is then followed by six padding bits.
2. If the DTS is not equal to the CTS, the DTS-flag MUST be set to 1. In this case, the AU header consists of 24 bits: the DTS-flag, the 22 bit signed integer DTS-delta, and the RAP-flag.

7.4.2.3 Encrypted AVC video bytestream format

This mode defines encrypted AVC video. Unlike AAC and CELP, ISMA 2.0 AVC video uses the payload format defined in RFC 3984 [RFC3984]. To enable the carriage of encrypted AVC video, a new mode is defined here for its carriage solely in the context of this encrypted payload format. This mode MUST not be used with unencrypted video.

The media that is encrypted is transformed before encryption, in order to enable recovery of NAL units even if RTP packets are lost. The length-codes which are stored before each NAL Unit in the standard AVC file format are replaced by start-codes, so that each NAL Unit is now structured in a legal byte-stream format according to Annex B of the AVC specification [14496-10]. This process is safe since the AVC specification [14496-10] requires start-code emulation avoidance (even in length-field oriented applications).

Note: The byte-stream format [14496-10 annex B] allows two types of start-codes: "four bytes 00 00 00 01" and "three bytes 00 00 01". Since the AVC File format [14496-15] does not permit 3-byte length fields, the byte-stream MUST use 4-byte start-codes (to enable easy transformation back into length fields after decryption). In other words, the fields "leading_zero_8bits" and "trailing_zero_8bits" MUST NOT be present and the field "zero_byte" MUST be present in each NAL unit. After decrypting the stream, two consecutive start-codes are needed to discover the length of each NAL unit.

This mode is signaled by mode=avc-video. In this mode, the configuration of the payload is as follows:

- SizeLength = 0. Each packet contains only one AU or AU fragment.
- IndexLength = 0 and IndexDeltaLength = 0. No interleaving.
- CTSDeltaLength = 0. No signaling of CTS needed as it is obtained from the RTP timestamp.
- DTSDeltaLength = 22. Video access units may have a DTS that differs from the CTS (for example, when B-frames are present), therefore the DTS can be signaled in this mode.
- RandomAccessIndication = 1. The RAP-flag is set to 1 to indicate video I-frames.
- Config. Must be present and is the hexadecimal value of AVCDecoderConfigurationRecord [14496-15]. In this structure, LengthSizeMinusOne indicates the length of start-codes and MUST be set to 3 (corresponding to 4-byte start-codes).

At most one access unit or fragment appears in an RTP packet. Multiple access units in a packet are not allowed.

The rules for packetization specified in [RFC3984] are not required in this mode. An access unit may be split anywhere at all, without regard to video packet, video header, or any other boundaries.

For this mode, the AU header takes one of two sizes (not counting the cryptographic metadata):

1. If the DTS is equal to the CTS, the DTS-flag is set to zero. In this case, the AU header consists of only two bits (the DTS-flag followed by the RAP-flag), which is then followed by six padding bits.
2. If the DTS is not equal to the CTS, the DTS-flag MUST be set to 1. In this case, the AU header consists of 24 bits: the DTS-flag, the 22 bit signed integer DTS-delta, and the RAP-flag.

8.0 ISMACryp SDP Signaling

This section defines SDP [RFC2327] signaling for ISMACryp parameters. The presence of ISMACryp parameters in the SDP description is signaled with a new session-level SDP attribute called "ismacryp-compliance" which is described in this section.

8.1 Overview

ISMACryp signaling has session and stream signaling parameters. The stream signaling parameters describe the encryption of the stream.

1. Crypto suite: Identifies the cipher, mode, keylength, authentication algorithm, etc.
2. IV length: Describes the size of the initialization vector in bytes.
3. Key indicator length: Describes the size of the key indicator in bytes.
4. Selective encryption: Indicate whether selective encryption is used for the session or not.
5. Salt key: Initiates the salt key value used, given an additional IV, to generate the AES-CTR counter.
6. Key: A structure that describes the key management system or conveys the key for the stream.

Key indicator length and selective encryption are optional since ISMACryp streams are not required to rotate keys or have unencrypted media frames. The salt key is also optional; if not present, the salt key is directly managed with the same system as the main key.

In addition to the stream signaling parameters, there are two optional transport parameters.

7. Delta IV length: Describes the maximum size of the optional media-frames initialization vector.
8. Key indicator per AU: Indicates key rotation on a media frame basis.

The delta IV length parameter is needed when media frames are interleaved in packets and unneeded otherwise. Key indicator per AU is needed when the stream has multiple keys and the packetizer might rotate a key between two media frames that are in the same packet.

8.2 ISMACryp version identification

The ISMACryp version is explicitly announced in the following mandatory parameter:

```
a=ISMACryp-compliance:<lowest-spec-version>,<authored-to-version>
```

The fields are defined as follows:

- lowest-spec-version: a decimal number, indicating the lowest version number of the ISMACryp specification to which a client can conform, and still decode the content. Clients MUST not decode content with a lowest-spec-version higher than the highest specification version that they implement.
- authored-to-version: the version of the specification against which the content was authored. Ideally the client also implements this version, whereupon the user can be more confident that the content is being completely decoded. A content author may choose to allow clients written to earlier versions of the spec achieve partial decode.

The ISMACryp-compliance parameter MUST only be used in combination with the 'enc-isoff-generic' mode.

Note: The ISMACryp-compliance was not present in the ISMACryp specifications below version 2. Thus, the lowest number that can appear in the above defined two fields is "2.0".

8.3 ISMACryp parameters

The following ISMACryp format specific SDP parameters are common parameters of both modes ‘enc-isoff-generic’ and ‘enc-mpeg4-generic’. All ISMACryp SDP signaling parameters and names are case-insensitive.

Table 8.3.1: ISMACryp fntp parameters

DESCRIPTOR	DEFINED VALUES	DEFAULT
ISMACrypCryptoSuite	AES_CTR_128 ¹	AES_CTR_128 ¹
ISMACrypIVLength	1..17 (8) ²	4 ¹
ISMACrypDeltaIVLength	0..2	0
ISMACrypSelectiveEncryption	0 (False) or 1 (True)	0
ISMACrypKeyIndicatorLength	0..255	0
ISMACrypKeyIndicatorPerAU	0 (False) or 1 (True)	0
ISMACrypSalt	Base64 encoded 64-bit number	0
ISMACrypKey	(type) string	""
ISMACrypKMSID	4CC	
ISMACrypKMSVersion	unsigned int 32	
ISMACrypKMSSpecificData	quoted-string	""

ISMACrypCryptoSuite identifies the default cipher, mode, key length and other descriptors used to describe the encryption of ISMA media (see Section 9). AES-CTR is the default and mandatory-to-implement cipher and mode, see Section 9 of this document.

ISMACrypIVLength describes the byte length of the initialization vector that is conveyed initially in the ISMACryp packet. For the default cipher and mode, this is the BSO value (see Sections 3.2 and 9)

Note: In the ISMACryp 1.1 specification, the maximum length for IV is 8 bytes, since the 8 most significant bytes of the counter are constant (they are the salt key). However, in the OMA specification [OMADCFV2], the 8 most significant bytes of the counter are not required to be constant. Furthermore, OMA's IV counts blocks and not bytes, so ISMACryp needs 4 additional bits to represent the same counter value. Therefore ISMACryp 2.0 extends the maximum allowed IV length to 17 bytes.

ISMACrypDeltaIVLength describes the byte length of the initialization vector, if any, that is conveyed with an individual AU. See Section 9 for the encoding used for the default AES counter value.

ISMACrypSelectiveEncryption declares that the media stream uses selective encryption when it is set to 1, which indicates that the selective encryption bit will appear in the ISMACryp header.

When ISMACrypKeyIndicatorLength is non-zero, a key indicator will appear in the ISMACryp header. ISMACrypKeyIndicatorLength can signal a key indicator field that is 0 to 255 bytes in length.

When ISMACrypKeyIndicatorPerAU is non-zero, a key indicator appears on a per-AU basis.

ISMACrypSalt initializes the salt key with a randomly generated and non-null value that will be used for counter generation in the entire media stream (see Section 9). Unless specifically supplied by the KMS, the default salt value is 0.

ISMACrypKey identifies the key to the receiver. The parameter "type" is either "URI" or "KEY." "URI" is a URI for the key management system, which identifies the key or a location from which to obtain the

¹ See Section 9

² for ‘enc-mpeg4-generic’ the maximum value is ‘8’, for ‘enc-isoff-generic’ the maximum value is ‘17’

key [RFC1630]. Thus, a uniform resource identifier follows the type "uri" as in *ISMACrypKey=(uri)https://example.isma.tv*. "KEY" is an encoding of the key used to decrypt the stream. This encoding is specific to the encryption transform. Section 10 defines the ISMACryp default encoding for ISMACrypKey. When ISMACrypKey is not explicitly signaled, it is an empty string, meaning that there is no key in the SDP, and the key will be delivered from the key management entity (e.g. Conditional Access System or OMA DRM v2).

ISMACrypKMSID is the identification of the KMS used.

ISMACrypKMSVersion is the version of the KMS used.

ISMACrypKMSSpecificData contains information necessary to the KMS (e.g. a KMS Content ID).

Note: The syntax for the quoted-string carried in "ISMACrypKMSSpecificData" is as follows:

- quoted-string = (<"> *(qdtxt) <">)
- qdtxt = <any TEXT except <">>
- TEXT = <any OCTET except CTLs>
- OCTET = <any 8-bit sequence of data>
- CTL = <any US-ASCII control character (octets 0 - 31) and DEL (127)>

The signaling message SHOULD be authenticated when carrying the above defined parameters and SHOULD be encrypted when the ISMACrypKey parameter appears with an encoded key. In addition to signaling the encryption of the ISMA stream, ISMACryp defines signaling for the authentication of ISMA messages. See Section 9 for the default message authentication (integrity) transform for ISMACryp and its signaling method.

For examples of fntp statements, see Annex C and D.

The ISMACrypKey=(uri) in ISMACrypKey serves as a string identifier for the key. The ISMACrypkey=(key) parameter, however, is the key itself, base-64 encoded, and with an optional lifetime parameter. Since the key is crypto-suite dependent, it is defined by the particular crypto-suite (see Section 10). Despite the above example, there is no reason to specify defaults in the fntp parameters and this practice is NOT RECOMMENDED (e.g., there is no reason for ISMACrypCryptoSuite=AES_CTR_128 since it does not change the configuration and takes up space in the SDP message).

It is also possible to convey several keys in parallel to allow key renewal using an extended key signaling: ISMACrypKey = (key) BASE64(aes-key1||salt1)||lifetime1|KI1,BASE64(aes-key2||salt2)||lifetime2|KI2,BASE64(aes-key3||salt3)||lifetime3|KI3;

Therein, the KI1 is the corresponding key indicator of aes-key1 of a lifetime1 and a salt1, KI2 is the corresponding key indicator of aes-key2 of a lifetime2 and a salt2, KI3 is the corresponding key indicator of aes-key3 of a lifetime3 and a salt3, and so on.

8.4 'enc-isoff-generic' SDP signaling

The SDP signaling of the enc-isoff-generic payload format uses parts of mpeg4-generic signaling, but is not a complete superset to that in mpeg4-generic [RFC3640]:

- The 'mpeg4-generic' *mode* attribute and the parameters for codec identification and configuration MUST NOT be used: *streamType*, *profile-level-id*, *config*, *mode*, *objectType*.
- Instead, the parameters for codec identification and initialization as defined below in section 8.4.1 and 8.4.2 MUST be used: *codec*, *config.xxx*.
- All ISMACryp fntp parameters (see Table 8.3.1) MAY be used.

- The following mpeg4-generic parameters MAY be used : *constantSize*, *constantDuration*, *maxDisplacement*, *de-interleaveBufferSize*, *sizeLength*, *indexLength*, *indexDeltaLength*, *CTSDeltaLength*, *DTSDeltaLength*, *randomAccessIndication*, *streamStateIndication*, *auxiliaryDataSizeLength*

8.4.1 Codec identification

The mandatory parameter “codec” MUST be used for codec identification. The codec identification is identical to the 4CC of the sample entry in the file format. Any character is URL encoded [RFC1738] except for “a”-“z”, “A”-“Z”, “0”-“9” and “-”.

The mandatory codec identification parameter is:

```
codec=4CC
```

Examples:

H.263 video:	codec=s263
AMR audio:	codec=samr
MPEG-4 Video:	codec=mp4v
H.264:	codec=avc1
H.264 byte-stream format:	codec=264b

8.4.2 Codec initialization

Most modern codecs need parameters for their initialization. In the ISO file format [14496-12], these parameters are stored in one or more boxes stored in the *SampleDescriptionBox*.

To carry this configuration, a suite of optional parameters is added, all starting with “config.”:

```
config.xxxx=<value>
```

The fields are defined as follows:

- *xxxx* : the 4CC name of the corresponding box contained in the sample description box. Any character is URL encoded [RFC1738] except for “a”-“z”, “A”-“Z”, “0”-“9” and “-”.
- *<value>* : the content of the box
 - coded in base-64
 - not including the length and 4cc name fields of the box
 - including the version and flags fields (if present).
 - must be present in the *fmtp* line in the same order as the corresponding boxes in the sample description.

Note: You may obtain an example of an ES Descriptor ('esds' box of MPEG-4 codecs) writer/parser by requesting the MPEG reference software code from www.mp4ra.org

Examples:

H264:

```
config.avcC=AULgDf/haANl1QKD8gBAARozjyA; config.btrt=AADFRAAGKiAAg
```

H263:

```
config.d263=VmlWaQAKAA==
```

8.4.3 Optional parameters

The following parameters are optional, the sender MAY decide to use them or not. Receivers, however, MUST support parsing these parameters.

8.4.3.1 ISMACrypSliceIndication

For an optimized support of slices, an optional parameter called "ISMACrypSliceIndication" is added. This parameter indicates whether the `slice_start` and the `slice_end` fields are present in the AU-header.

ISMACrypSliceIndication = "0" | "1"

If ISMACrypSliceIndication is 1, the `slice_start` and the `slice_end` fields must be present in each AU header.

Note: For backwards compatibility of 'enc-isoff-generic' packets with 'enc-mpeg4-generic' receivers, ISMACrypSelectiveEncryption should also be switched on if ISMACrypSliceIndication is switched on, even if it is not intended to use the selective encryption feature (in this case, the `AU_is_encrypted` (SE) field in the ISMACryp sample header should always be set to '1').

8.4.3.2 ISMACrypPaddingIndication

For the support of padding bits, an optional parameter called "ISMACrypPaddingIndication" is added. This parameter indicates whether the `padding_bitcount` field is present in the AU header.

ISMACrypPaddingIndication = "0" | "1"

If ISMACrypPaddingIndication is 1, the `padding_bitcount` field is present and its size is 3 bits.

Note: For backwards compatibility of 'enc-isoff-generic' packets with 'enc-mpeg4-generic' receivers, ISMACrypSelectiveEncryption should also be switched on if ISMACrypPaddingIndication is switched on, even if it is not intended to use the selective encryption feature (in this case, the `AU_is_encrypted` (SE) field in the ISMACryp sample header should always be set to '1').

8.4.3.3 ISMACrypExtensionHeaderLength

This parameter indicates the size of the ISMACrypExtensionHeader in bytes. For the default value of zero the parameter may not be present. In the current version of this specification the value MUST be zero. For forward compatibility to future versions of this specification, receivers MUST support parsing this parameter.

8.5 'enc-mpeg4-generic' SDP signaling

The SDP signaling of the 'enc-mpeg4-generic' payload format is not a subset of 'enc-isoff-generic', but a superset of 'mpeg4-generic'. It differs as described as follows from 'enc-isoff-generic':

- The ISMACryp compliance parameter as defined in section 8.2 MUST NOT be used.
- The 'enc-isoff-generic' parameters for codec identification and initialization as defined in section 8.4.1 and 8.4.2 MUST NOT be used: *codec*, *config.xxxx*.
- Instead, the 'mpeg4-generic' mode attribute and the parameters for codec identification and configuration SHALL be used: *streamType*, *profile-level-id*, *config*, *mode*, *objectType*.
- The optional 'enc-isoff-generic' parameters as defined in Section 8.4.3 MUST NOT be used.

The following parameters may be used for 'enc-mpeg4-generic' in the same way as for 'enc-isoff-generic':

- All <ISMACRYP-PARAMS> defined in Section 8.3 (see Table 8.3.1) MAY be used.

- The following mpeg4-generic parameters MAY be used : *constantSize*, *constantDuration*, *maxDisplacement*, *de-interleaveBufferSize*, *sizeLength*, *indexLength*, *indexDeltaLength*, *CTSDeltaLength*, *DTSDeltaLength*, *randomAccessIndication*, *streamStateIndication*, *auxiliaryDataSizeLength*

Generic SDP signaling:

```
m=<media> <port>[/<number of ports>] <transport> <fmt list>
a=rtpmap:<payload type> <encoding name>/<clock rate>[/<encoding parameters>]
a=fmtp:<payload type> mode=<mode>; <MPEG4-GENERIC-PARMS> <ISMACRYP-PARMS>
```

```
<media> = "audio" | "video"
<transport> = "RTP/AVP" | "RTP/SAVP"
<encoding name> = "enc-mpeg4-generic"
<mode> = "aac-hbr" | "aac-lbr" | "celp-cbr" | "celp-vbr" | "mpeg4-video" | "avc-video"
```

The mode attribute is MANDATORY and is defined in mpeg4-generic [RFC3640] or it is "mpeg4-video" as defined in Section 7.4.2.2 for MPEG-4 video streams, or it is "avc-video" as defined in Section 7.4.2.3 for AVC video streams.

For AVC video the NAL Unit structure MUST be transformed into byte-streams prior to encrypting and back into length fields after decrypting as described in Section 7.4.2.3.

MPEG4-GENERIC-PARMS are OPTIONAL parameters that are defined in mpeg4-generic [RFC3640]. ISMACRYP-PARMS are OPTIONAL and defined in Section 8.3.

Note: the parameters in the fmtp line (mode=<mode>; <MPEG4-GENERIC-PARMS> <ISMACRYP-PARMS>) may appear in any order.

9.0 ISMACryp Encryption (Default) Cryptography Specification

This section describes the default encryption and authentication transform for ISMAEncryption and Authentication (ISMACryp). Future specifications MAY define new encryption and/or authentication transforms to supercede or augment these definitions. Section 9.1 describes the ISMACryp encryption transform and 9.2 describes the ISMACryp message authentication (integrity) transform.

9.1 ISMACryp AES-CTR Encryption Transform

Section 9.1.1 defines the cipher, mode, and key length. Section 9.1.2 defines needed transport packet fields. Section 9.1.3 defines signaling parameters.

9.1.1 Default cipher, mode, and configuration

The AES blocksize is 128 bits and so is the counter, which is defined below. The key length SHALL be 128 bits. The 128-bit key, blocksize, and counter describe the AES-CTR cipher exactly. The AES block cipher encrypts the counter to form a pseudorandom block of 128 bits; successive AES blocks form a stream of AES-encrypted blocks called the "keystream." AES-CTR generates a stream of pseudo-random blocks based on AES encryption; encryption is done by XORing those bytes with plaintext to encipher and XORing those bytes with ciphertext to decipher.

Figure 9.1-1 shows AES in Counter Mode with a 128-bit counter, key, and a 128-bit keystream block. The figure assumes "big-endian" order where the left-most bit or byte is the most significant. The "div" denotes truncated integer division that is effectively a logical right shift of n bits for "div 2^n " with zero bits moved into the shifted positions. The exclusive-or operation, designated by "(*)" in Figure 9.1-1, is bit-wise exclusive-or, which is applied to the keystream and a 128-bit block of plaintext to produce a block of ciphertext, or it is applied to a 128-bit block of ciphertext to produce the original block of plaintext.

Each plaintext byte corresponds to one and only one BSO (see Section 3.2, Glossary) for the stream. The BSO information is included in the packet and called the "IV" or initialization vector. The encryption operation is performed using the BSO whereas decryption is performed using the IV, which is the BSO value of the first byte of payload data. This section generally uses only one term, "IV," and it should be understood to mean the BSO when the operation is encryption. The IV starts at the first byte in the packet payload when the packet does not contain interleaved AUs. When the packet contains interleaved AUs, an IV is associated with each AU.

Given the IV, salt and key, the counter is formed using the IV and a 64-bit salting key, k_s , which is left shifted 64 bits and right-padded with zeros. As shown in Figure 9.1-1, the IV is XORed to salting key k_s . The multiply by 2^{64} effectively shifts k_s left by 64 bits and could be written as " $\ll 64$ " using C programming language notation; the div operation is truncated integer division and could be written as " $\gg 4$." Since the BSO and corresponding IV are stream byte-counters, the div is by 16 since there are 16 bytes (octets) in a 128-bit number, i.e. the div operation yields that keystream block that contains the IV byte. The result of the XOR operation is the counter as shown in Figure 9.1-1.

AES-CTR [AES-CTR] defines the value shown as "counter" in Figure 9.1-1 as a "nonce" meaning that it is used once and only once in the counter space for a given key. A value of the counter MUST NOT be repeated for a given key or else the security of AES-CTR is compromised. The IV carries the unique value that forms the counter. For ISMACryp default encryption, the BSO SHALL be incremented for every byte in the stream regardless of whether that byte is encrypted or selectively unencrypted. The BSO is conveyed as an IV value in an ISMACryp packet (Section 7). The length of the IV is application dependent and MAY be signaled at the start of the session and SHALL remain fixed for the duration of the session (see 9.1.2 and 9.1.3). AES-CTR places no restriction on the nonce/counter to be randomized in any way. Indeed, some scholarly work proves superior security for AES-CTR without assuming that the nonce is randomized (e.g. starts from a random offset and incremented from there modulo the 128-bit blocksize, see [SECURITY]). Some experts assert, however, that the counter needs to be initialized from a random offset within the

counter space to prevent a "key-collision attack" [MF00]. This is the reason that the default mode and cipher adds a salting key (k_s) random offset to the AES counter.

In addition to the counter there is the AES key, which is 128 bits. This key **MUST** be secret and difficult for an adversary to guess (there **MUST** be 128 bits of randomness in a 128-bit key). The key is denoted as "k" in Figure 9.1-1 and is used to encrypt the counter for the encryption and decryption operations, which are symmetric. The AES key **SHOULD NOT** be used to encrypt more than one ISMACryp stream owing to the danger of counter reuse [AES-CTR].

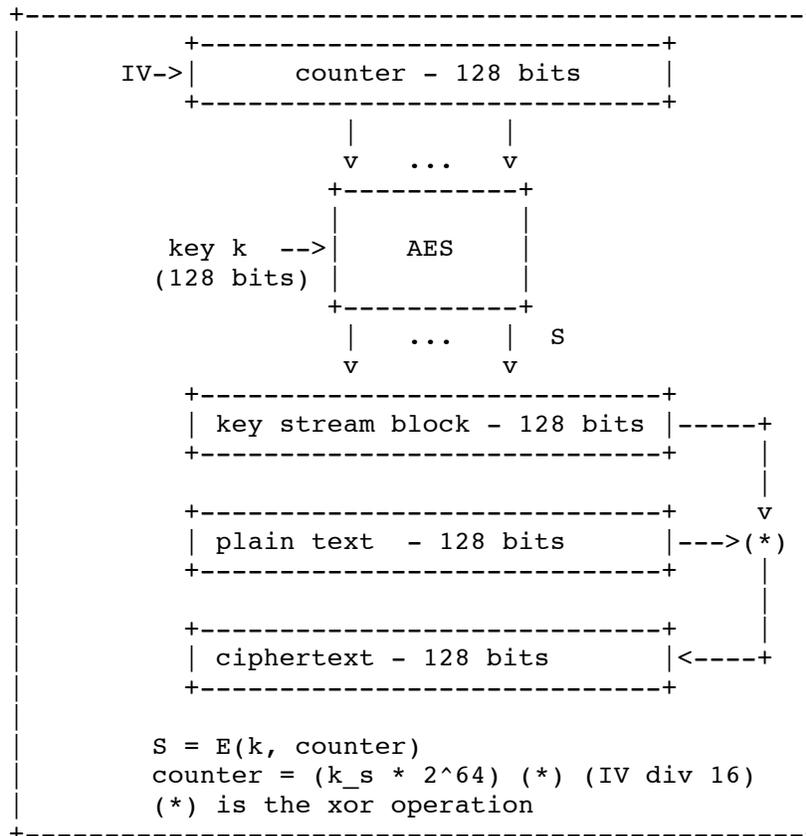


Figure 9.1-1: AES-CTR Mode Encryption

As shown in 9.1-1, the key stream is a sequence of 128-bit blocks (the figure shows just one of them) that are XORed to the plaintext to encrypt and the ciphertext to decrypt. The encrypted byte stream is therefore dependent on the cipher key, the salting key (the non-secret random offset in the counter space) and the counter. With this, we have the following useful properties:

- The keystream may be pre-computed prior to the processing of the cleartext (on the encryption side) or the encrypted text (at the decryption side).
- The data is encrypted in 128-bit blocks, which are cryptographically independent of each other; thus it is not necessary to have a previous or subsequent block of cleartext for encryption or ciphertext for decryption of a given block of data.
- There is zero expansion of the data: Each ciphertext byte is in 1:1 correspondence with the plaintext byte.

- From the preceding bullets, the keystream is "seekable" to any given byte (BSO) of clear or cipher text, where the initial block and byte of the keystream for the packet-payload AU can be computed from the IV as

$$\begin{aligned}\text{keystream-block} &= \text{IV div } 16 \\ \text{keystream-byte} &= \text{IV mod } 16\end{aligned}$$

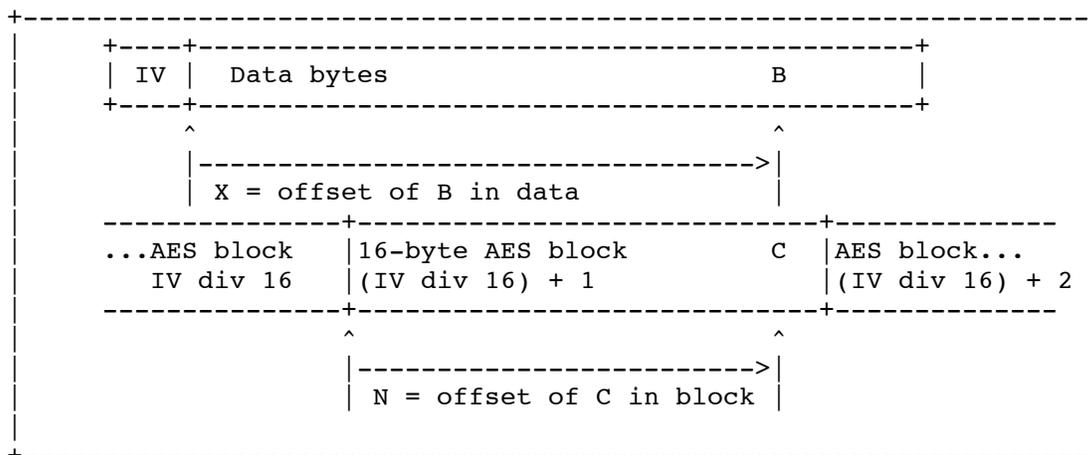


Figure 9-1.2: Locating a keystream byte within the packet payload

Given that "IV div 16" and "IV mod 16" locate the first byte of keystream in a packet payload AU from the IV (or delta IV), figure 9-1.2 shows how any byte within the payload can be matched to the corresponding byte of keystream that was used to encrypt it. Given the counter, the keystream is "seekable" to any given byte of cleartext or ciphertext in the ISMACryp packet payload. The counter is formed from the Initialization vector (IV) or vectors in the packet (one per non-interleaved packet or one per AU in an interleaved packet). In the diagram above, consider byte B which is at offset X after the periodically-supplied initialization vector IV. It is encrypted by XORing it with byte C, which is at offset N in a 128-bit AES keystream block. The counter value for that AES keystream block, and the offset N within that AES keystream block are given by:

$$\begin{aligned}\text{counter} &= k_s * 2^{64} \text{ XOR } ((\text{IV} + \text{X}) \text{ div } 16) \\ N &= (\text{IV} + \text{X}) \text{ mod } 16\end{aligned}$$

The AES-CTR mode decryption has the same structure as its encryption counterpart except that the key stream block is XORed with the cipher text block in order to obtain the plain text.

Note: The IV SHOULD start from zero and MUST be reset before it overflows. Keys SHOULD be changed as well at least once between every IV reset to avoid AES counter reuse. Specifically, for every access unit, the value of the access unit's IV plus the length of the access unit in bytes MUST NOT exceed $(1 \ll (8 * \text{IV_Length}))$. This restriction avoids an ambiguity as to whether the AES counter would continue to increment or would wrap around to zero after the largest possible IV value is reached within the access unit.

9.1.2 Fixed parameters and signaling values

Parameters describing packet fields that are an integral number of bytes are stated in bytes (octets) rather than bits. There are several fixed parameters for the ISMACryp default encryption transform.

1. AES_CTR_128 is the AES-CTR encryption cipher and mode defined in this section with a 128-bit key, a 128-bit blocksize and a 64-bit salting key.
2. The Salt-Key length is 8 bytes.
3. The AES block size is 16 bytes.

4. The AES key length is 16 bytes.

There are a few parameters that MAY be set through signaling.

5. ISMACrypCryptoSuite MAY be set to AES_CTR_128, or it MAY default to this value.
6. ISMACrypIVLength defaults to 4 bytes but MAY be set to any value between 1 byte and 8 bytes. ISMACrypKey MAY carry a URI that identifies a key server when its type is "URI". When its type is "KEY," ISMACrypKey encodes one or several inline ISMACryp decryption key(s) as follows:

$$\text{BASE64}(\text{aes-key}||\text{salt})||\text{lifetime}|KI\{\text{,BASE64}(\text{aes-key}||\text{salt})||\text{lifetime}|KI\}^*$$

Thus, when its type is "KEY," ISMACrypKey is a UTF-8 string with three components. The first component is the concatenated aes-key and salt; "||" is the concatenate operator. The aes-key is 16 bytes and the salt is 8 bytes so "aes-key || salt" is 24 bytes before Base 64 encoding, which results in 32 bytes since Base 64 is a "three in four" encoding scheme. The value MUST be unique and is followed by a "|" and the key lifetime, which is the number of bytes that may be encrypted and decrypted using this master key. When empty, the lifetime defaults to 2^{64} bytes, which is the maximum width of the ISMACrypIVLength, parameter #6 above (the AES-CTR limitation is 2^{64} blocks [SECURITY]). The lifetime is followed by a "|" and the optional key indicator (KI) for this master key. The "|" characters are needed only when the lifetime and/or key indicator is present in the ISMACrypKey parameter.

The following is an example of an ISMACrypKey that is provided with no explicit lifetime (defaults to 2^{64} bytes) and no KI:

```
ISMACrypKey=(key)MTIzNDU2Nzg5MDEyMzQ1Njc4OTAxMjM0
```

The lifetime may be specified as follows:

```
ISMACrypKey=(key)MTIzNDU2Nzg5MDEyMzQ1Njc4OTAxMjM0|2^24
```

In this case, after decrypting 2^{24} bytes of data, the receiver MUST NOT use the key. Either the stream is at an end or the new key is obtained from the KMS. Note that the ISMACrypKey parser MUST be able to evaluate expressions of the form 2^x .

When the ISMACryp stream uses a key indicator, this value MUST be specified as follows.

```
ISMACrypKey=(key)MTIzNDU2Nzg5MDEyMzQ1Njc4OTAxMjM0|4000000|1492
```

In the example above, the lifetime is 4 million bytes and a KI of 1492 is associated with the key.

9.1.3 Transport packetization values

Parameters describing packet fields that are an integral number of bytes are stated in bytes (octets) rather than bits. The only transport packetization value needed by the cipher is an IV input for decryption and a BSO input for encryption, where the IV is a BSO sample that the packetizer adds to an ISMACryp packet. Section 3.2, Glossary, defines the BSO.

An enc-isoff-generic or enc-mpeg4-generic packet contains the "initial IV" value for the first access unit or fragment contained in the packet; in many cases when packet data is not interleaved, the Initial_IV is the only IV in the packet. Its length is determined by the signaling parameter ISMACrypIVLength, and has a default length of 4 bytes. If the media stream is longer than 2^{32} bytes, then there are two options. The first option is to use a longer IV length that is as large as the stream. The second option is to rotate the key before the IV wraps.

When interleaving is used, there SHALL be an IV assigned to each enc-isoff-generic or enc-mpeg4-generic AU. See Section 7 for specifics of how to encode multiple IVs in an enc-isoff-generic or enc-mpeg4-generic packet.

9.2 ISMACryp Message Authentication (Integrity) Transform

The default ISMACryp message authentication (Integrity) transform is SRTP with an HMAC-SHA1 with an 80-bit output tag and a 128-bit master key [RFC3711]. The master key is used with SRTP key derivation function to compute a 160-bit authentication key. This service is signaled using SDP security descriptions [RFC4568]. ISMACryp applications SHOULD NOT signal or use SRTP encryption but MAY signal SRTCP encryption. Here is an example of such SDP signaling:

```
a=crypto:1 AES_CM_128_HMAC_SHA1_80
  inline:azerRTazad1223dsdsfEhtgdjj12ZSzerefigtyt|2^20|1:32
  UNENCRYPTED_SRTP
```

Where

- crypto:1 = we are using master key 1 (
- AES_CM_128_HMAC_SHA1_80 specify the mode (mandatory)
- inline = key (optional)
- 2^20 = lifetime (optional)
- 1:32 = master key indicator (MKI) of 32-bits wide and associates the number 1 (optional)
- UNENCRYPTED_SRTP means that ISMACryp doesn't use SRTP encryption (mandatory)

9.3 The Security of ISMACryp Cryptography

Attacks on the encryption or authentication of ISMACryp media and messages encounter too high a workload to be perceived as a threat to the media confidentiality of an ISMA stream [SECURITY, RFC2104]. ISMACryp signaling is secure against forgery, replay attack, and unauthorized disclosure of an ISMACryp parameter when the SDP signaling uses a data-security protocol such as TLS or IPsec. When an SDP message conveys an enc-isoff-generic or enc-mpeg4-generic parameter, it SHOULD be authenticated and integrity-protected using TLS or IPsec. When an SDP message conveys the enc-isoff-generic/enc-mpeg4-generic ISMACrypKey key parameter, it SHOULD be encrypted. When properly used, ISMACryp is appropriate for government, enterprise, and individual security applications.

There are three key-management risks, however, to the proper use of ISMACryp encryption and message authentication.

1. Counter reuse: Additive stream ciphers share the security properties of the One-Time-Pad encryption system and can disclose information about the plaintext segment when two different plaintexts are encrypted using the same keystream segment (under certain circumstances, the plaintext can be disclosed). It is RECOMMENDED that an ISMACryp encryption key be used for one and only one unidirectional stream: Two or more streams SHOULD NOT use the same ISMACryp key. It is NOT RECOMMENDED that the salting_key be used to ensure that the keystream is unique among streams since it is there to protect against key collision attacks [MF00] and not to make the keystream unique.
2. Key collision: McGrew and Fluhrer [MF00] describe an attack that weakens the ISMACryp key (i.e. reduces the effective bit length) through an attack where the attacker precomputes a large number of keys starting from beginning of the counter space, and looks for key matches based on known plaintext in the stream. This attack is prevented by randomizing the start of the counter space to an unpredictable starting value [MF00]. This random offset is the salting key, k_s, of Section 9, which MUST be unpredictable to the attacker.
3. Key disclosure: When used properly (i.e. avoiding counter reuse) ISMACryp would not be the target of an attacker seeking to get unauthorized access to media-stream plaintext. The easier and oftentimes feasible approach is to attack the key management system. Thus, the protection afforded by ISMACryp for government, enterprise, or individual security is no stronger than the security of the particular key management system and the protection of keys by devices and their users.

There is an additional risk to message confidentiality when there is no authentication: If a cryptanalyst knows the plaintext in a particular position in the stream, and if the attacker can benefit from transforming

the bytes of known plaintext into different values, then the attacker can ensure that those bytes decrypt to the different value rather than the original (known) plaintext. To prevent such an attack from succeeding, an ISMACryp implementation SHOULD use message authentication when the ISMACryp stream traverses an insecure network. When the ISMACryp stream is stored on an insecure host computer, the ISMACryp implementation SHOULD use file authentication techniques.

In addition to organizational or individual security applications, ISMACryp MAY be incorporated into content protection applications to serve as a scrambling mechanism. Consumer devices that process encrypted content generally violate the security assumptions of computers in a government or enterprise security environment since consumer devices routinely suffer key disclosure. Studios, record labels, and other distributors of copyright content that use ISMACryp SHOULD evaluate the robustness and compliance of the key management system that handles ISMACryp keys. This system, however, is outside the scope of ISMACryp.

Note: When an encrypted stream needs to be re-encrypted (decrypted with the old key, encrypted with a new key), decryption and encryption can be done in the reverse order. The encrypted stream can first be XORed with the new key stream and afterwards decrypted (XORed) with the old key stream. This way the content stream is never in the clear.

10.0 Name Assignment and Registration

The following table lists each name assignment that **MUST** be reserved for ISMACryp, its description, and the relevant naming authority.

VALUE	TYPE	DESCRIPTION	AUTHORITY
enc-mpeg4-generic	MIME	ISMACryp1.1 encrypted payload type	IANA
enc-isoff-generic	MIME	ISMACryp 2.0 encrypted generic payload type	IANA
AES_CTR_128	UTF-8	ISMACryp parameter	ISMA
all of Table 8.3.1	UTF-8	ISMACryp parameters	ISMA
enca or encv	4CC	ISO FF 4CC sample description	MPEG
iAEC	scheme-type box	ISO FF ISMACryp default encryption	ISMA
264b	4CC	ISO FF 4CC original format for AVC bytestream format	ISMA
OMA2	4CC	ISO FF 4CC OMA DRM v2 KMS	ISMA
isc2	4CC	ISMACryp 2.0 brand	mp4ra

Note: For usage of the brand “isc2”, see section 6.5.

Annex A: Local Playback (Informative)

This annex provides a walkthrough of how local playback, random access, and editing, can be performed, when the ISMA AES-CTR encryption is used.

1. First, check that the ISMA AES-CTR mode encryption is being used.
2. Then, get the keys needed. Retrieve the KMS indicator and parameters from the sample description, and interface to the KMS to acquire a set of key-indicators/key mappings. This interaction may use the time-offset in the stream.
3. From the sample description, extract the salt, selective-encryption-used, key-indicator-size, and initial-counter-size.
4. Read the sample from disk using standard MP4 means, and place it into a buffer.

```
encrypted := ((selective-encryption-used == 1) ? read-bytes(1) & 0x80 : 0);
if (encrypted == 0x80) then {
    key-indicator := read-bytes( key-indicator-size );
    initial-counter := read-bytes( initial-counter-size );
    decrypt-buffer-bytes( key-map[key-indicator], salt, initial-counter );
}
```

Annex B: Encryption Process Example (Informative)

1) Basic encryption procedure

In this example an (AES) encryptor block size of 3 is used to illustrate the process. The counter starts with a value of 0. By incrementing the counter a byte stream containing "key stream bytes" is generated as indicated in Figure A-1:

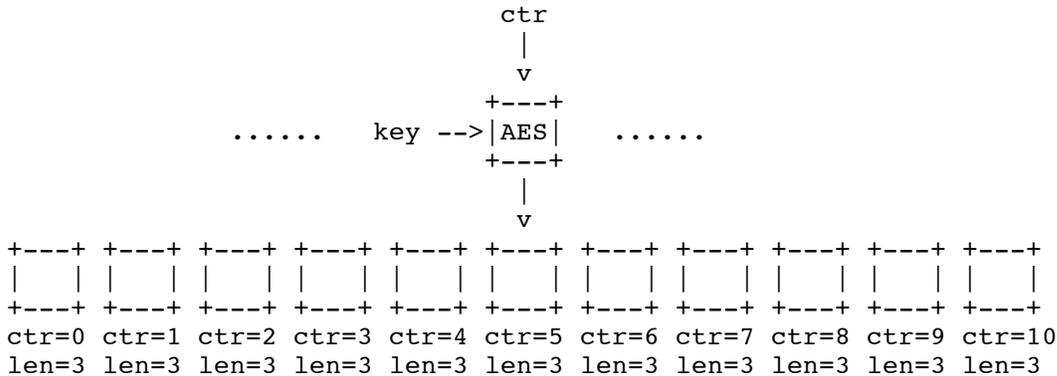
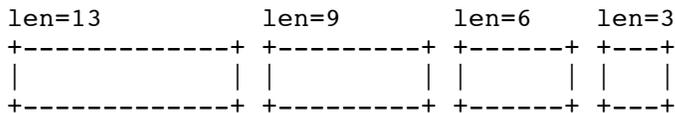


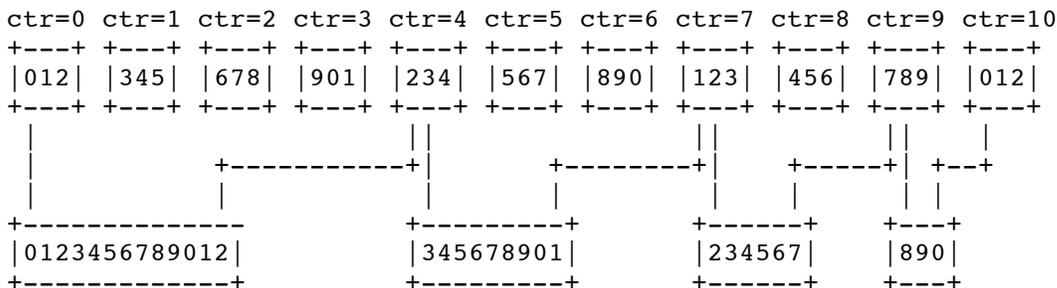
Figure C.3-1: Indicating a high level overview of the key stream generation

To encrypt a plain text byte stream with AES in Counter Mode, the plaintext byte stream is XORed with the key stream to obtain the corresponding cipher-text byte stream. The decryption process is the reverse of the encryption process, i.e., the cipher text byte stream is XORed with the same key byte stream to receive the original plaintext byte stream.

In the following, the plain text is media stream access Units, where, to illustrate the encryption process, 4 access units of different lengths are used:



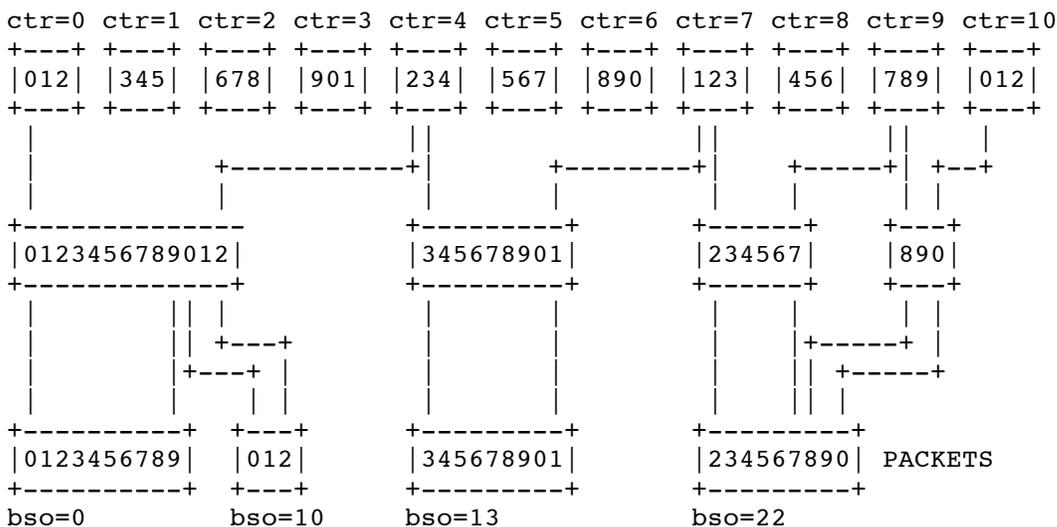
For encryption these access units packets are treated as a single byte stream and are lined up in sequence against the key stream bytes as indicated in the following figure:



Thus the media access units are encrypted by performing the byte-wise XOR of the their bytes with the corresponding key stream bytes. Note that if after encrypting an access unit, if the bytes from the previous encryptor block have not been exhausted, the remaining bytes of that block will carry-over to the beginning

of the next access unit. Thus the encryptor block boundaries are NOT aligned with the access unit boundaries.

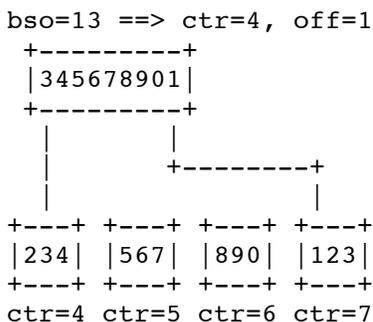
The encrypted access units are now packetized by inserting the "AU byte stream offset" (BSO) into the ISMACryp header (see Section 7). In the example below, the first AU gets fragmented over 2 packets, the second AU goes in as a whole into one packet, and the third and fourth AU are small enough that they may be combined into a single packet:



Having the byte stream offset (bso) in each packet allows the decryption of each packet independent of the other packets. To illustrate this, suppose the third packet, which happens to contain one complete access unit, is received. The byte stream offset (bso) received in the packet is used to derive the counter (ctr) and encryptor block offset (off) as follows:

$$\begin{aligned} \text{ctr} &= \text{bso} / 3 \\ \text{off} &= \text{bso} \% 3 \end{aligned}$$

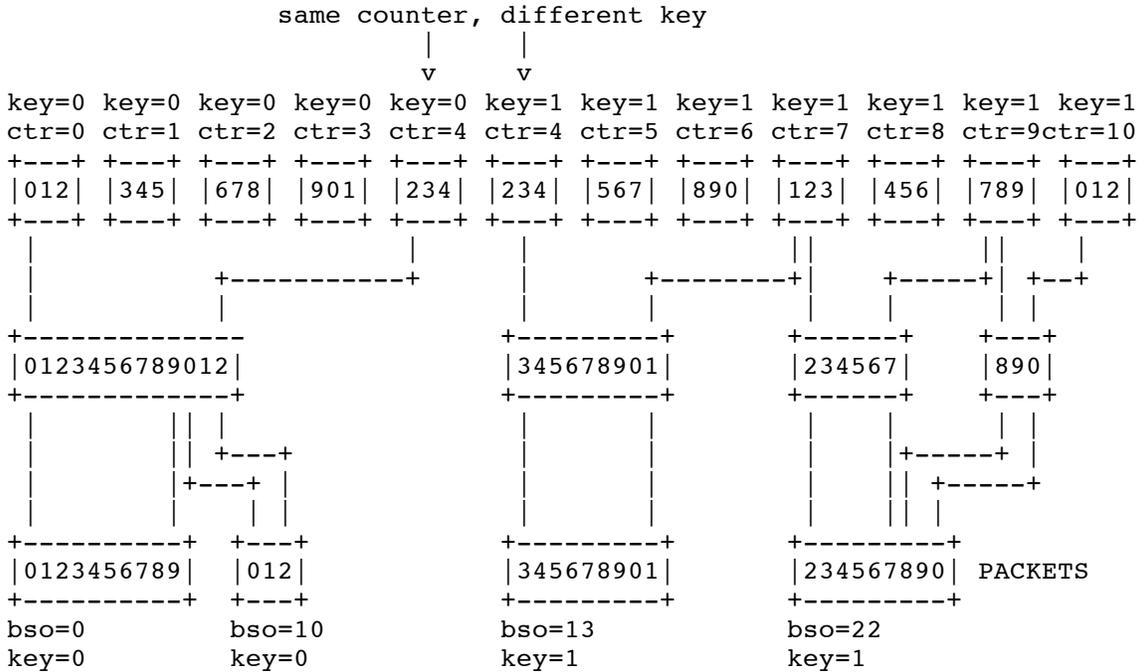
where 3 is the encryptor block size (3 in our example; it is 16 bytes for AES). And that is all is needed to regain synchronization of the data with the key stream. The packet data may be decrypted as follows:



2) Key change

The notion of a key change is illustrated taking the same example from above. Without any loss of generality, assume that the key changes to the next key at the beginning of the second access unit. Note that, in this example, this happens to fall in the middle of an encryptor block. In the case of an encryption

wherein the key remained the same, the last couple of unused key stream bytes from the last encryptor block would be used for the encryption or decryption of the bytes in the next access unit block. But since the key has changed, the key stream block must be regenerated with the same counter, but with the new key, as indicated in the figure below:



For the decryption process, the "bso" and key indicator are sufficient to re-generate the correct key stream and thus, each packet corresponding to an encrypted access is independently decrypted, even with key changes.

Note that the Key Manager can select the key based upon the key indicator, but it is also possible for the Key Manager to determine which key to use based upon the bso (in the example above, key=1 takes effect at bso=13).

3) Selective encryption

Suppose all access units except the 3rd are to be encrypted. It is desired that the 3rd access unit be sent in the clear. During encryption process this is easily accomplished by not performing the XOR for those bytes that have to be left in the clear (marked with XXX below), and, in addition, skipping the unused key stream bytes (marked as UUU below).

Annex C: 'enc-mpeg4-generic' SDP Examples (Informative)

Note about transport type: In media announcement, "SRTP/AVP" MUST be used when either SRTP authentication or SRTP encryption is used. In all other cases, "RTP/AVP" MUST be used (even though the stream is encrypted).

The enc-mpeg4-generic encrypted and authenticated MPEG-4 audio mode

Notes:

- ISMACrypSelectiveEncryption: this example uses selective encryption, so this field is explicitly set to 1 in the fntp line.

Example: (for AAC-hbr mode)

```
m=audio 0 RTP/AVP 96
a=rtpmap:96 enc-mpeg4-generic/22050
a=fntp:96 streamtype=5; profile-level-id=15; mode=AAC-hbr; config=1388;
SizeLength=13; IndexLength=3; IndexDeltaLength=3; ISMACrypSelectiveEncryption=1;
ISMACrypKey=(uri)shttp://talkingHeads.isma.tv
a=mpeg4-esid:1
a=crypto:1 AES_CM_128_HMAC_SHA1_80
  inline:azerRTazad1223dsdsfEhtgdjj12ZSzerefigtyt|2^20|1:32
  UNENCRYPTED_SRTP
```

The enc-mpeg4-generic encrypted MPEG-4 video mode

Notes:

- ISMACrypDeltaIVLength: this parameter is not used since each packet contains only one AU or AU fragment.
- ISMACrypKeyIndicatorPerAU: this parameter is not used since each packet contains only one AU or AU fragment.

Example:

```
m=video 0 RTP/AVP 96
a=rtpmap:96 enc-mpeg4-generic/600
a=fntp:96 streamtype=4; profile-level-id=1; mode=mpeg4-video;
config=000001b0f3000001b50ee040c0cf0000010000000120008440fa28202056a21f;
DTSDeltaLength=22; RandomAccessIndication=1;
ISMACrypKey=(uri)shttp://talkingHeads.isma.tv
a=mpeg4-esid:2
```

The enc-mpeg4-generic encrypted AVC video mode

Notes:

- ISMACrypDeltaIVLength: this parameter is not used since each packet contains only one AU or AU fragment.
- ISMACrypKeyIndicatorPerAU: this parameter is not used since each packet contains only one AU or AU fragment.

Example:

```
m=video 0 RTP/AVP 96
a=rtpmap:96 enc-mpeg4-generic/90000
```

```
a=fmtp:96 streamtype=4; mode=avc-video;
config=0142E00DFFE1000A6742E00D965202C12C8001000468CE3C80; DTSDeltaLength=22;
RandomAccessIndication=1; ISMACrypKey=(uri)shttp://talkingHeads.isma.tv
a=mpeg4-esid:3
```

Interoperability with OMA DRM Version 2.0

Example: CELP-cbr mode

```
m=audio 0 RTP/AVP 96
a=rtpmap:96 enc-mpeg4-generic/16000/1
a=fmtp:96 streamtype=5; profile-level-id=14; mode=CELP-cbr; config=440E00;
constantSize=27; constantDuration=240; ISMACrypSelectiveEncryption =1;
ISMACrypSalt=aXNtYUITTUE; ISMACrypKey=(uri)http://www.rightsserver.org/;
ISMACrypKMSID=OMA2; ISMACrypKMSVersion=512; ISMACrypKMSSpecificData
="content145678@ContentIssuer.com"
a=mpeg4-esid:4
```

Annex D: 'enc-isoff-generic' RTP packetization and SPD examples (Informative)

Note: These examples are here for information; the presence and the value of any SDP parameters for any type of codec should be handled on the client side. However, for common codecs, the following SDP may be used.

Encrypted H.263 video

The SDP lines for an encrypted H.263 video stream would look like:

```
m=video 0 RTP/AVP 96
a=rtpmap:96 enc-isoff-generic/90000
a=fmtp:96 codec=s263; config.d263=VmlWaQAKAA==; DTSDeltaLength=22;
RandomAccessIndication=1
```

In this configuration (assuming no B-frame), the RTP packet would consist of:

- RTP header
- AU-headers-length field (16 bits) = 40
- initial_IV (32 bits) = IV of the AU fragment
- DTS-flag (1 bit) = 0
- RAP-flag (1 bit) = 1 if the AU fragment is part of an I-frame, 0 otherwise
- padding (6 bits) = 0
- fragment of the H.263 AU
- complete AU or fragment of one AU

For a B-frame, the RTP packet would consist of:

- RTP header
- AU-headers-length field (16 bits) = 56
- initial_IV (32 bits) = IV of the AU fragment
- DTS-flag (1 bit) = 1
- DTS-Delta (22 bit) = 2's complement offset from the CTS
- RAP-flag (1 bit) = 1 if the AU fragment is part of an I-frame, 0 otherwise
- fragment of the H.263 AU
- complete AU or fragment of one AU

Note: The marker bit in the RTP header is set to 1 if the RTP packet contains the last fragment of an H.263 access unit.

Encrypted AMR-NB audio

The SDP lines for a selectively-encrypted AMR-NB audio stream (assuming silence detection is not used) would look like:

```
m=audio 0 RTP/AVP 96
a=rtpmap:96 enc-isoff-generic/8000/1
a=fmtp:96 codec=samr; config.damr=VmlWaQEAAQAB;
constantSize=13; constantDuration=160; ISMACrypSelectiveEncryption=1
```

In this configuration, the RTP packet would consist of (assuming SelectiveEncryption):

- RTP header
- AU-headers-length (16 bits) = 32 + 8 * number of AUs in packet
- For the first AU:
 - AU_is_encrypted (1 bit) = 1 if AU is encrypted, 0 otherwise

- Slice-start-flag (1 bit) = 0, unused
- End-start-flag (1 bit) = 0, unused
- Padding_bitcount (3 bits) = 0, unused
- reserved (2 bits) = 0
- initial_IV (32 bits) = IV of the first AU
- For the following AUs:
 - AU_is_encrypted (1 bit) = 1 if AU is encrypted, 0 otherwise
 - Slice-start-flag (1 bit) = 0, unused
 - End-start-flag (1 bit) = 0, unused
 - Padding_bitcount (3 bits) = 0, unused
 - reserved (2 bits) = 0
- One or more encrypted AMR audio frames

Encrypted H.264 video (ISMACryp 1.1 backward compatibility)

The SDP lines for an encrypted H.264 video stream would look like:

```
m=video 0 RTP/AVP 96
a=rtpmap:96 enc-isoff-generic/90000
a=fmtp:96 codec=264b; config.avcC=AULgDf/hAAonQuANqRgsP/SAAQAEKM4eIA==;
config.btrt=AADFRAAGKiAABiog; DTSDeltaLength=22; RandomAccessIndication=1
```

The RTP packet structure is identical to the “avc-video” mode of enc-mpeg4-generic.

Encrypted H.264 video (no ISMACryp 1.1 backward compatibility)

The SDP lines for an encrypted H.264 video stream would look like:

```
m=video 0 RTP/AVP 96
a=rtpmap:96 enc-isoff-generic/90000
a=fmtp:96 codec=avc1;
config.avcC=AULgDf/hAAonQuANqRgsP/SAAQAEKM4eIA==;
config.btrt=AADFRAAGKiAABiog; DTSDeltaLength=22; RandomAccessIndication=1;
ISMACrypSliceIndication=1
```

In this configuration (assuming no B-frame), the RTP packet would consist of (assuming SliceIndication):

- RTP header
- AU-headers-length field (16 bits) = 48
- AU_is_encrypted (1 bit) = 1, unused, always set to one for backward compatibility
- Slice-start-flag (1 bit) = 1, if the fragment is the first fragment of the slice, 0 otherwise
- End-start-flag (1 bit) = 1, if the fragment is the last fragment of the slice, 0 otherwise
- Padding_bitcount (3 bits) = 0, unused
- reserved (2 bits) = 0
- initial_IV (32 bits) = IV of the AU fragment
- DTS-flag (1 bit) = 0
- RAP-flag (1 bit) = 1 if the AU fragment is part of an I-frame, 0 otherwise
- padding (6 bits) = 0
- complete NAL units (slices) or fragment of one NAL unit (slice)

For a B-frame, the RTP packet would consist of:

- RTP header
- AU-headers-length field (16 bits) = 64
- AU_is_encrypted (1 bit) = 1, unused, always set to one for backward compatibility
- Slice-start-flag (1 bit) = 1, if the fragment is the first fragment of the slice, 0 otherwise
- End-start-flag (1 bit) = 1, if the fragment is the last fragment of the slice, 0 otherwise

- Padding_bitcount (3 bits) = 0, unused
- reserved (2 bits) = 0
- initial_IV (32 bits) = IV of the AU fragment
- DTS-flag (1 bit) = 1
- DTS-Delta (22 bit) = 2's complement offset from the CTS
- RAP-flag (1 bit) = 1 if the AU fragment is part of an I-frame, 0 otherwise
- complete NAL units (slices) or fragment of one NAL unit (slice)

Note: The marker bit in the RTP header is set to 1 if the RTP packet contains the last fragment of an AVC access unit.

Encrypted MPEG-4 AAC audio

The SDP lines for an encrypted MPEG-4 AAC audio stream would look like:

```
m=audio 0 RTP/AVP 96
a=rtpmap:96 enc-isoff-generic/48000/2
a=fmtp:96 codec=mp4a; config.esds=
AAAAAAOAGwABAASAEkAVABgAAAE4gAABOIAFgAISEAYBAG==; sizeLength=13; indexLength=3;
indexDeltaLength=3; ISMACrypIVLength=4
```

The RTP packet structure is identical to the “AAC-hbr” mode of enc-mpeg4-generic.

Encrypted MPEG-4 video

The SDP lines for an encrypted MPEG-4 video stream would look like:

```
m=video 0 RTP/AVP 96
a=rtpmap:96 enc-isoff-generic/90000
a=fmtp:96 codec=mp4v; config.esds=
AAAAAM4AAAABDAGEQAlgAAF3AAAAAABSEAAAAGwAAABUO4CAGjwAAAQAAAAEgAMeIruBQsEJBRZ8
GAQIAAAAAAAAAA==; DTSDeltaLength=22; randomAccessIndication=1;
ISMACrypIVLength=4;
```

The RTP packet structure is identical to the “mpeg4-video” mode of “enc-mpeg4-generic”.

Annex E: 'enc-isoff-generic' RTP receiver behavior in case of packet loss (Informative)

If `slice-start-flag` and `slice-end-flag` are not used, when an IP packet is lost, the RTP receiver may drop the entire AU (depending on decoder robustness).

If `slice-start-flag` and `slice-end-flag` are used, when an IP packet is lost, the RTP receiver may drop only the slice which is affected by the packet lost. All others slices may be delivered to decoder.

Example:

- MTU = 800 bytes for the RTP payload
- One Access Unit: 3100 bytes, 5 slices

Slice 1	Slice 2	Slice 3	Slice 4	Slice 5
200	550	150	2000	200

<----- video frame ----->

RTP packet n°1

Headers	Slice 1	Slice 2
	200	550

SLS=1/SLE=1

SLS = slice start flag
SLE = slice end flag

RTP packet n°2

Headers	Slice 3
	150

SLS=1/SLE=1

RTP packet n°3

Headers	Slice 4
	800

SLS=1/SLE=0

RTP packet n°4

Headers	Slice 4
	800

SLS=0/SLE=0

RTP packet n°5

```
+-----+-----+
|Headers|  Slice 4 |
|        |  400   |
+-----+-----+
SLS=0/SLE=1
```

RTP packet n°6

```
+-----+-----+
|Headers|Slice 5 |
|        |  200   |
+-----+-----+
SLS=1/SLE=1
```

Annex F: Interoperability with OMA DRM Version 2.0 (Informative)

This annex provides guidelines on how ISMACryp can be used together with the key and rights management system of OMA DRM v2. It specifies those hooks in the file format and SDP which allow to obtain keys from the OMA DRM KMS for streaming as well as downloading ISMA content.

F.1 Overview

OMA DRMV2.0 is split into four parts: Architecture, Rights Expression Language [OMARELv2], DRM protocols (e.g. Rights Object Acquisition and Management) [OMADRMv2] and DRM Content Format [OMADCFv2]. For more information, the specifications can be consulted from the OMA website (<http://www.openmobilealliance.org>).

Readers familiar with these specifications may have noticed that in OMA DRM DCF v2.0 candidate specification [OMADCFv2], the Packetized DRM Content Format is almost based on ISMACryp format.

F.2 MPEG-4 File Structure

The file format provides a mean to identify when the OMA DRM key management system is used.

F.2.1 Sample description transformation

The Scheme Information Box is only a container box used to carry DRM KMS specific information. The series of contained boxes can be of any type and format, thus it can include OMA specific boxes. To use OMA DRM KMS, this box MUST include exactly:

- a) In first position, one ISMA KMS Box "iKMS" with
 - version = 1
 - KMS ID = OMA2
 - KMS version = 0x00000200
 - KMS URI = OMA DRM v2 right issuer URI
- b) Then one OMA DRM Common Headers Box "ohdr" [OMADCFv2] that specifies the encryption scheme and its parameters and provides information about the Rights Issuer as well.
- c) In third position, one ISMASampleFormat Box "iSFM".
- d) In last position, one ISMACrypSaltBox.

These hooks should be enough to launch the OMA DRM v2.0 Right Object Acquisition Protocol (ROAP) [OMADRMV2] to acquire the license and then to access the associated content. If there is only one salt-key for the stream, the ISMACrypSaltBox defines it.

F.3 Transport Signaling

When using both ISMACryp and OMA DRM v2.0 combined, session and stream signaling parameters must identify these standards. These parameters identify the crypto suite and the structure of the OMA DRM KMS.

Regarding ISMACryp current signaling, this solution SHOULD use the enc-isoff-generic format and its associated generic parameters as a base. In addition, three MANDATORY parameters are required: a salt key, a content identifier parameter and a rights issuer URL parameter.

The SDP fmtp signaling SHALL use enc-isoff-generic as its format, which is case sensitive.

Note: for H.264 streams, the byte-stream format SHOULD not be used as this format is not supported by OMA specification

Table F.3.1: fntp parameters

DESCRIPTOR	Known in [OMADRMV2] as	Defined values for OMA DRM v2	Default value for OMA DRM v2
ISMACrypSalt		<i>Base64 encoded 64-bit number</i>	0
ISMACrypKey	Right Issuer URL	<i>(uri) string</i>	
ISMACrypKMSID		OMA2	OMA2
ISMACrypKMSVersion		0x0000200	0x0000200
ISMACrypKMSSpecificData	ContentID	<i>URI, quoted using <"></i>	""

For examples of fntp statements, see Annex C or D.

Annex G: Use of ISMACryp prior to OMA DRM 2.0 super-distribution (informative)

G.1 Introduction

Assume a mobile operator introduces OMA DRM 2.0 based services over a 3G mobile network, complemented with delivery over a broadcast network, such as DVB-H. Much content is real-time streamed, using RTP, in particular over the broadcast network. The user can store the streamed content on his device and has the option to share the recorded content that he particularly likes with others by sending it to his friends so that they can consume the content on their OMA DRM 2.0 compliant devices.

This concept is called super-distribution, whereby the content is first distributed to the primary users, but where these primary users can distribute it further to their friends, while these friends can further distribute it again to their friends, etc. Of course the friends typically will need to pay for the rights to consume the super-distributed content, but such payment is at the operator's discretion.

For such applications, the use of ISMACryp is very interesting, as ISMACryp supports storage of the content and optional further distribution without any re-encryption. However, for consumption at OMA DRM 2.0 compliant devices, the stored content has to comply with the OMA DRM 2.0 specification. This can be achieved easily by using ISMACryp in a specific manner. This Annex describes the required constraints for the use of ISMACryp so that subsequent OMA DRM 2.0 super-distribution of the content to OMA DRM 2.0 clients can take place without any re-encryption of the content.

G.2 ISMACryp streaming followed by OMA DRM 2.0 super-distribution

In figure G-1, OMA DRM 2.0 super-distribution subsequent to ISMA streaming is depicted. The content to be streamed over RTP is either encrypted in real time or is stored in encrypted form in a file. When streaming, the encrypted content is packetized in RTP packets as specified by ISMACryp and the RTP packets are broadcast to the users. When recording, the user stores the encrypted content in a file; so as to be OMA DRM 2.0 compliant, the file complies with the (P)DCF format specified in OMA DRM 2.0. For random access purposes and various other reasons, it is strongly recommended the streamed content be stored in a PDCF file instead of a DCF file. Once the content is stored in the PDCF file, the user can send the file to the OMA DRM 2.0 clients of his friends, and, if so desired, these friends can distribute the files further to their friends, etc., as depicted in figure G-1.

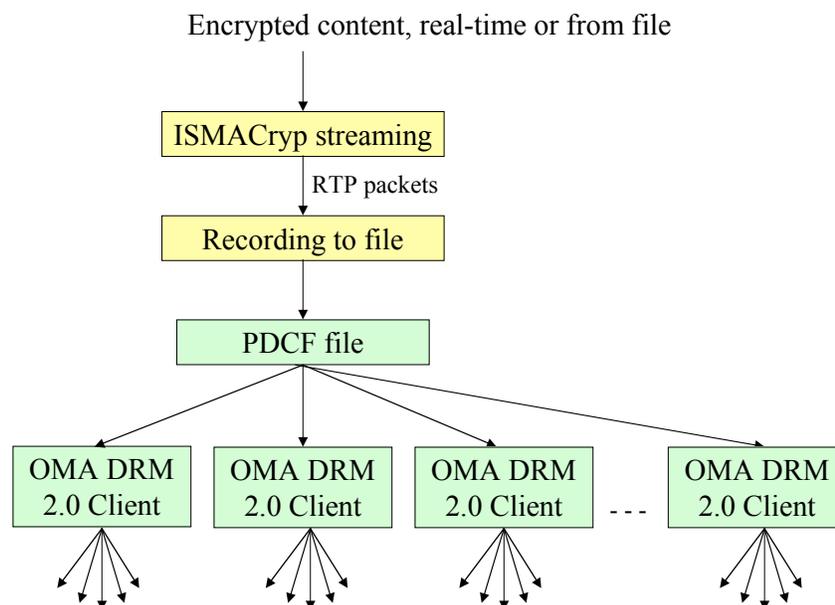


Figure G-1 OMA DRM 2.0 super-distribution subsequent to ISMA streaming

Users receiving super-distributed content can only playback the content if they obtain playback rights from a Rights Issuer. To allow users to contact the Rights Issuer for this purpose, the Common Header Box in the super-distributed PDCF file must contain a Rights Issuer URL. To allow the Rights Issuer to provide the Rights for the content, the super-distributed content in the PDCF file must be identified by a unique content-id that must be provided in the Common Header Box. The content-id may contain timing information (such as date and start / end times) of the recorded content and may be included in the Rights Issuer URL. Next to the content-id also the group-id concept as defined in OMA DRM 2.0 may be used. In case the Rights Issuer URL contains sensitive information, integrity protection on the Rights Issuer URL may be needed, e.g. by using a key derived from the content encryption key. Note also that the Rights Issuer can only provide Rights for the super-distributed content if the Rights Issuer has knowledge of the key that is used to encrypt the content. The format of the Rights Issuer URL, the format of the content-id and the associated issues are application specific and beyond the scope of this Specification, but applications must ensure their proper definition.

G.3 Requirements for ISMACryp streaming

In Annex F, "Interoperability with OMA DRM Version 2.0" it is described how to apply OMA DRM 2.0 on top of the file format specified by ISMACryp, but not how to produce an OMA DRM 2.0 compliant PDCF file from one or more ISMACryp streams. The following should be taken into account to ensure that the encrypted content received after ISMACryp streaming can be stored in OMA DRM 2.0 compliant PDCF files.

1. OMA DRM 2.0 and ISMACryp use AES_CTR_128 as encryption algorithm, which allows the construction of OMA DRM 2.0 compliant streams without any re-encryption, but in DCF the salt key is not supported. When encrypting an ISMACryp stream, then a salt key of zero MUST be used.

Note: in DCF, the initial value of the 128 bit counter is prefixed to the ciphertext, and the counter is incremented for each AES cipherblock by 1 (modulo 2^{128}).

2. OMA DRM 2.0 DCF supports selective encryption, but requires that the selective encryption flag must be transported for each Access Unit; therefore the ISMACrypSelectiveEncryption in SDP must be set to 1.
3. OMA DRM 2.0 PDCF does not support key cycling, hence this feature in ISMACryp must not be used. Consequently, for ISMACryp in SDP the ISMACrypKeyIndicatorLength must be set to 0.
4. The structure of the AU header as used in ISMACryp differs from the AU header in PDCF; therefore the AU header in the ISMACryp RTP packets must be transformed in PDCF compliant AU headers; in this context it should be noted that PDCF allows an AU to consist of a group of samples.
5. In ISMACryp, AUs are encrypted as a sequence of bytes without any requirement for alignment between AUs and AES blocks. In PDCF however a new cipherblock starts at the beginning of each "PDCF Access Unit" (corresponding to a group of one or more samples), and hence, unlike in ISMACryp, "PDCF Access Units" are always AES Block aligned. Thus, ISMACryp uses a byte counter and the concept of a byte stream offset, while PDCF simply uses an AES block counter, whereby each AES block consists of 16 bytes. As a consequence, an ISMACryp stream can only be packetized in an OMA DRM 2.0 compliant PDCF without re-encryption if AES Block alignment is applied within the ISMACryp stream. Therefore ISMACryp must provide the IV information for each "PDCF Access Unit" carried in the RTP payload, and for each "PDCF Access Unit" the IV data must indicate the start of a new AES block, thereby typically introducing a IV discontinuity at the beginning of each "PDCF Access Unit".

G.4 Conclusion

Content that is streamed to devices by means of ISMACryp can be super-distributed to OMA DRM 2.0 compliant devices if the requirements described in this Annex are taken into account. Section G.3 provides the requirements, such as ISMACryp parameter initialization, IV constraints and transformation of headers, which are within the scope of this Specification. However, some other requirements need further definition at application level, and are therefore beyond the scope of this Specification. For this purpose, in summary, applications must ensure that the following is specified:

- A method for unique identification of super-distributed content, so that the Rights Issuer, when so requested by recipients of the super-distributed content, knows for which content to grant rights.
- A method to ensure that the Rights Issuer has knowledge of the key that is used to encrypt the super-distributed content, so as to allow the Right Issuer to provide the correct keys for decryption of the content.
- The format of the Rights Issuer URL in the Common Header Box, so as to allow proper communication between the recipient of the super-distributed content and the Rights Issuer.
- A method to apply integrity protection on the Rights Issuer URL, if needed, to protect sensitive information contained in this URL.

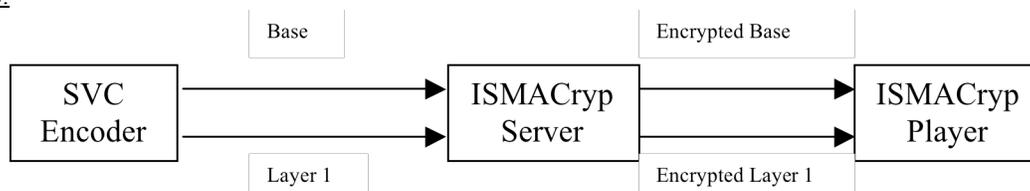
Users can distribute OMA DRM 2.0 protected content without explicit permission, but of course this only makes sense if the content provider / rights issuer is willing and prepared to grant rights to the recipient users to consume the content. Therefore it is desirable that applications provide means to indicate whether OMA DRM 2.0 super-distribution is a meaningful option.

Annex H: 'enc-isoff-generic' SVC protection (Informative)

SVC allows the viewing of a video (Live, VOD ...) at several quality levels according to terminals' properties or available bandwidth in a network. Transcoding is unnecessary and a simple selection of encoded layers provides targeted quality and rate. Scalability may spread out in several paths: P=priority, D=resolution (size of the screen), T=frame rate and Q=SNR.

By conjunction of SVC and ISMACryp, it is also possible to offer various qualities of video depending of the subscription level of users. In this case, each transmission level must have an independent key.

Live:

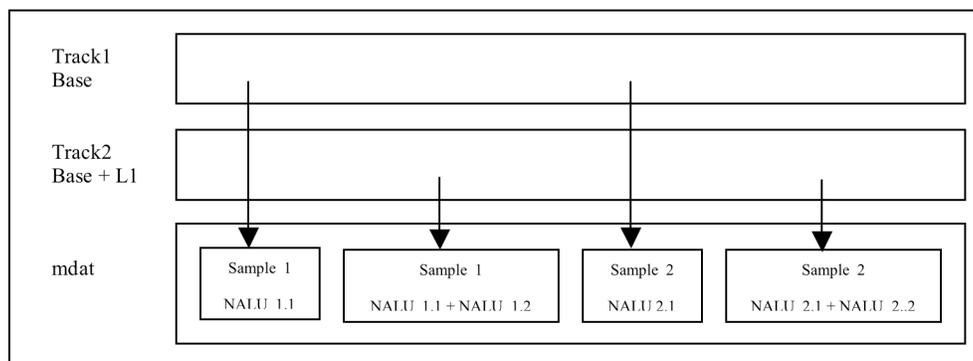


Each stream is encrypted separately at the sample level: NAL unit headers and NAL unit payloads are encrypted. Note that the sample here is different from the Access Unit in the SVC terminology: the sample gathers all the NAL units corresponding to a specific dependency layer, e.g. all the NAL units of a given spatial resolution. Keys can be identical or different on each layer (salt key must be different).

VoD:

During the encryption of an SVC file, the following rules should be respected:

- Extractors and aggregator [SVCFF] must not be encrypted
- Encrypted file must not use any of the following SVC File Format mechanisms: aggregators, extractors, mapping group etc.: (see [SVCFF]). Hence, each track must have its own copy of data (encrypted NAL units cannot be shared between tracks since this is not compatible with ISMACryp encryption)
- A metadata mechanism (extended SubsampleInformationBox, Time Metadata tracks etc.) is recommended to give the streaming server/hinter access to NAL units boundaries and PDTQ.



Note: all Samples are encrypted in this figure.

Bandwidth adaptation:

For bandwidth adaptation, RTP packets may be dropped on a specific layer (layers can be identified by their IP address, their UDP port or their SSRC, or by DSCP field in IP header). However it is not possible to select NAL units independently from their scalability information since that is conveyed in the NAL

header and therefore not accessible (encrypted). By scalability information, we mean the set of the following fields, for short (P,D,T,Q) :

- P (priority_id) : priority information for easy stream manipulation
- D (dependency_id) : indicates the layer, which is characterized by separate motion/prediction information (indicates spatial or CGS layer)
- T (temporal_level) : indicates temporal resolution
- Q (quality_level) indicates the quality refinement layer (FGS or MGS)

However it may be possible to transport the (P,D,T,Q) fields for all NAL units of a RTP packet in the Auxiliary Data Section.