

**Technical Reference Material****24 October 2013****Low Bandwidth Motion Imagery - *Technologies***

## 1 Purpose

This Motion Imagery Standards Board (MISB) Technical Reference Material (TRM) provides information on technologies relevant to the delivery of motion video, audio, and metadata to users whose information channel is characterized as low bandwidth consistent with the MISB MISM Levels L2, L1, and L0. This document does not mandate or recommend particular instantiations of technologies; other MISB documents exist for that purpose. Rather, this document is a survey of pertinent technologies to video over low-bandwidth channels, and should serve as an informative guide when implementing low-bandwidth motion imagery systems. It indicates where MISB approved technologies are appropriate, and provides an aid in appreciating tradeoffs and current industry practices.

## 2 Acronyms & Terms

AAC	Advanced Audio Coding
H.264/AVC	ITU-T Recommendation H.264   ISO/IEC 14496-10
AVC	Advanced Video Coding
HE-AAC	Advanced Audio Coding - High Efficiency AAC Profile
ISMA	Internet Streaming Media Alliance
MP4	MPEG-4 File Format
MPEG2 TS	MPEG2 Transport Stream Protocol
RTP	Real time Transport Protocol
RTCP	Real time Control Protocol
RTSP	Real time Streaming Protocol
SRTP	Secure Real time Protocol
SIP	Session Initiation Protocol
SDP	Session Description Protocol
SVC	Scalable Video Coding
TPED	Tasking, Processing, Exploitation & Dissemination
TPPU	Tasking, Posting, Processing and Using
Xon2	Compression type (MPEG2, H264) on MPEG2 Transport Stream

### 3 Reference

- [1] SMPTE ST 336:2007, Data Encoding Protocol Using Key-Length-Value
- [2] ISMA 2.0, Internet Streaming Media Alliance Implementation Specification, Apr 2005
- [3] ITU-T Rec. H.264 (04/2013), Advanced Video Coding for Generic Audiovisual Services
- [4] RFC 768, User Datagram Protocol, 28 Aug 1980
- [5] RFC 3550, RTP: A Transport Protocol for Real-Time Applications, Jul 2003
- [6] RFC 3551, RTP Profile for Audio and Video Conferences with Minimal Control, Jul 2003
- [7] RFC 3984, RTP Payload Format for H.264 Video, Feb 2005
- [8] RFC 2326, Real Time Streaming Protocol (RTSP), Apr 1998
- [9] RFC 3640, RTP Payload Format for Transport of MPEG-4 Elementary Streams, Nov 2003
- [10] RFC 6597, RTP Payload Format for Society of Motion Picture and Television Engineers (SMPTE) ST 336 Encoded Data”, Apr 2012
- [11] MISB EG 1001, Audio Encoding in MPEG-2 TS, May 2010
- [12] MISB EG 0803, Delivery of Low Bandwidth Motion Imagery, Apr 2008
- [13] MISB RP 0101, Use of MPEG-2 Systems Streams in Digital Motion Imagery Systems
- [14] MISB RP 0804, Real Time Protocol for Full Motion Video, May 2010
- [15] MISB TRM 1006, Key-Length-Value (KLV) Users Guide, Jun 2010

### 4 Introduction

Numerous methods exist for delivering Motion Imagery (MI) including: video-over-IP, video-over-cable, video-over-satellite, and terrestrial broadcast video among others. Numerous technology choices are available in preparing and publishing video: for example, MPEG-2, H.264, and VC-1 compression; Real Time Protocol and MPEG Transport Stream carrier protocols; and QuickTime, and ASF file formats to name several. Metadata increases the value of a MI asset, and so its structure, bandwidth, and relationship to motion imagery are important considerations. The question of how much extra bandwidth does metadata consume is particularly important in low bandwidth communications. Tradeoffs in the quality of the video and the quantity of metadata when meeting a particular channel bandwidth constraint are typically necessary. This technical reference material reviews relevant motion imagery technologies in the context of military applications where motion imagery is to be delivered to an end user over a low bandwidth or low-bit-rate communications channel. Such environments are sensitive to quality of service, cost of deployment and maintenance, user adaptability in spatial/temporal/fidelity resolution, and requested media mix.

Providing Situational Awareness (SA) information to edge users is typically done using narrow band data channels. Users armed with client devices may have motion imagery reception capabilities, but the low data rate channel may compromise the reception and quality of the motion imagery. This worsens with significant amounts of associated metadata. These channels may also be shared with other users, susceptible to errors from interference, and not guaranteed to meet any particular level of service.

## 4.1 *Situational Awareness: with respect to Motion Imagery*

“Situational Awareness” embodies broad concepts, but a suitable definition of situational awareness that offers a core theme is:

**Situational Awareness:** 1. Knowledge and understanding of the current situation which promotes timely, relevant, and accurate assessment of (friendly, enemy, and other) operations within the battlespace in order to facilitate decision making; 2. An informational perspective and skill that foster an ability to determine quickly the context and relevance of events as they unfold.

While broad this definition serves to illuminate the essence of what SA is: *timely and relevant information communicated within the battlespace*. Oftentimes, a SA user is the terminal point for MI data and their access to that data is through an extremely low bandwidth channel. The information they receive may come directly from an in-field sensor, such as a UAV, or from a exploited FMV processed further up stream. The FMV data is analogous to a fire hose of “stuff”, while their compromised data stream can be likened to a household garden hose. Methods to manage this information mismatch so that the end user receives meaningful intelligence require smart signal processing and suitable delivery mechanisms. There are two types of motion imagery that factor into battlespace operations: exploitation (PED) quality and *situational awareness* quality motion imagery.

- ❑ “Exploitation quality” motion imagery is the highest quality imagery achievable with a deployed system, including metadata, generally at “high” bit-rates delivered by the collection system to be used for exploitation and analysis by human analysts and automated analysis tools for the **production** of motion imagery products for distribution to end-users. Typically MISP MISM levels 3 and above.
- ❑ “Situational Awareness (SA) quality” motion imagery is a **product** derived from exploitation quality motion imagery, or other sources that is at a bandwidth, compression, resolution, frame rate, and metadata content suitable for distribution over operational networks to field users for the observation of static and dynamic relationship of entities within the field of view of the collection sensor. Typically MISP MISM Level 0, 1, and 2.

Figure 4-1 depicts the relationship between the two types of imagery within the battlespace. In TPED (Tasking, Processing, Exploitation & Dissemination) the motion imagery is held to a quality consistent with collection and sufficient for post analysis (usually highly trained image analysts), while in TPPU (Tasking, Posting, and Processing & Using) the motion imagery and metadata are prepared consistent with network conditions and client devices in the field.

This document is intended to address “situational awareness quality” motion imagery built upon low-data rate, bandwidth-challenged communications channels for compressed video and metadata products between production source and client user.

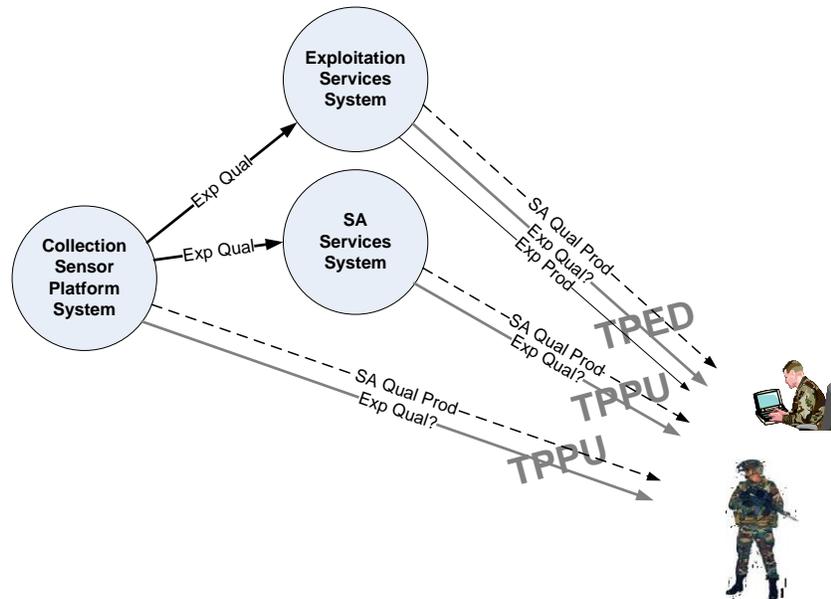


Figure 4-1: Motion Imagery in the Battlespace

## 4.2 Consideration of Issues for SA Motion Imagery System Developers

- Cost*: some field users have need for “zero” footprint SA motion imagery client services
  - No software loads to PC beyond “standard” enterprise load and browser
  - No cost for additional S/W incurred by using organization
- Scale*: some client applications deploy with enhanced imagery and metadata capabilities
- Security*:
  - User authentication / need to know
  - “Certification” of services, network distribution, client apps
    - ◆ Services
    - ◆ DoD-wide
- Versatility*: support streams and files including metadata

## 4.3 Situational Awareness (SA) Full Motion Video Features

- A standardized method for requesting a SA asset from a digital motion imagery provider
- Delivery of motion imagery over low bandwidth links with tolerance for packet loss, latency, and jitter
- Delivery of real time or archived digital motion imagery as stream or file
- Stream control, allowing clients to guide motion imagery delivery

- File storage of media and metadata
- Reduced spatial/temporal resolution and fidelity to meet channel capacity
- Choice in media delivered (video, audio, metadata)

#### **4.4 Questions Guiding Design Choices**

At a high level, streams and files will be issued from a data repository or made available directly from a sensor. A wired or wireless communication path will support transport of the data. A client device will receive the data for display. Beyond this baseline of services there are numerous application-specific issues that will guide technology tradeoffs. For example, is the communication link dedicated to this single use, or shared like the public internet? Is real-time streaming, progressive download, or file transfer required? Can the client request a component media type, such as metadata only, or can only composite MI assets be received? Can the client trade spatial and temporal resolution and bit rate fidelity? Can the application dynamically reduce the bit rate if the network degrades? What is the nominal channel bandwidth? Is the data coming from a sensor in real-time, or non-real-time from an archive?

These questions shape design and technology choices. For instance, a streaming application that uses a shared IP channel that is subject to jitter, packet loss, and varying delay together with client services that afford choice in media components with changes in resolution and fidelity may indicate Real-Time Protocol (RTP) as a preferred protocol, a repackaging of the file for streaming, a streaming server, and simulcast or scalable video coding of the MI asset. A situation where broadcast via microwave transmitter to hand-held client devices might indicate MPEG-2 Transport Stream (TS) as a preferred delivery protocol, a repackaging of the file (if necessary) for MPEG-2 TS, and H.264/AVC encoding of the MI essence to an appropriate spatial-temporal resolution and bit rate.

## **5 Motion Imagery Assets**

The constituent components of a motion imagery asset typically include video, metadata, and audio, although audio is oftentimes optional. In most cases, collected motion imagery will be of a greater resolution and fidelity than a bandwidth-challenged network can support. This collected data may need to be further compressed through a post-processing stage of decompression, spatial/temporal resolution scaling or fidelity reductions, and recompression. Care must be exercised that signal loss is minimized throughout this recoding process.

### **5.1 Video**

#### **5.1.1 MISB Recommendation of H.264/AVC**

The MISB recommends H.264/AVC for low-bandwidth distribution of video. In particular, MISM Levels L1.2, L1.1, L1.0 as specified in RP 9720e correspond directly to H.264/AVC Levels L1.2, L1.1, and L1.0. The MISB also recommends H.264/AVC for all collected motion imagery. **For this reason it can be appreciated that adherence to H.264/AVC across dissemination platforms will ensure interoperability, reuse, and quality of motion imagery assets.**

### 5.1.2 Merits of H.264/AVC

H.264/AVC, also known as MPEG4 Part 10, is displacing MPEG-2 as the preferred coding standard in the commercial world. Coding efficiency is roughly twice that of MPEG-2, and although the H.264/AVC codec is more complex than MPEG-2 microelectronics continues to accommodate the added complexity.

There are several compelling advantages in using H.264/AVC. For one, it is a standard that continues to grow in improved compression efficiency and greater flexibility. As an example, the Scalable Video Coding (SVC) standard, which affords various spatial, temporal, and fidelity levels within one coded bit stream, is founded on AVC and is a natural candidate technology for use in delivering video over bandwidth-challenged links. SVC is poised to become a successor to AVC as it addresses both the wide latitude of bandwidth variations found in IP video, and also the myriad of screen resolutions of consumer client devices. In other words, SVC is a good match to video delivered over IP. Secondly, H.264/AVC is beginning to displace MPEG-2 in the commercial broadcast industry, which means that more robust tools and products will be available when building systems.

## 5.2 Metadata

Metadata provides the context to digital motion imagery adding great value to a MI asset. The metadata used in our community follows the Key-Length-Value (KLV) construct, where the *Key* denotes a specific referenced item in the dictionary, the *Length* defines the length of the metadata data value, and the *Value* is the data itself. A brief description of KLV constructs is given below. For a more in-depth review of KLV see MISB TRM 1006.

### 5.2.1 Structure

In order to strike a balance between flexibility and bandwidth efficiency a number of pack constructs for KLV are utilized. SMPTE 336M-2007 defines several pack constructs; however, only two afford bandwidth efficiency: local sets and variable length packs. Also, the MISB Common Metadata Set defines two valid SMPTE 336M-2007 pack constructs providing additional flexibility: floating length and truncated packs.

### 5.2.2 Background: Summary Definition of SMPTE 336M-2007 Local Set

Refer to SMPTE 336M-2007 for complete definition of local sets. A Key Length Value (KLV) local set is defined as a number of data items that are grouped so that the length of the keys for each item can be reduced. A local set is identified by a 16-byte Universal Label (UL) key, and the items in the set are defined via a local set registry.



**Figure 5-1: Example of Local Set Structure**

Local sets are valuable in applications where the endpoints have a specific set of elements to share that can be defined in a standard or recommended practice (RP) agreed upon by all users. The use of short keys provides bandwidth efficiency without sacrificing the ability to selectively include data elements, as you would when using a fixed/variable length construct.

### 5.2.3 Background: Summary Definition of standard 336M Variable Length Pack

Refer to 336M for complete definition of variable length packs. A variable length pack is similar to a local set in that one key defines a number of data items. Here, though, a grouping of data elements is defined in a specific order so that the keys for individual elements can be removed. Elements with higher priority are ordered first followed by those of lesser importance. Each item in a variable length pack consists of a length field and a value field. A variable length pack is identified by a 16-byte UL key, and the items in the set are defined via a variable length pack registry.

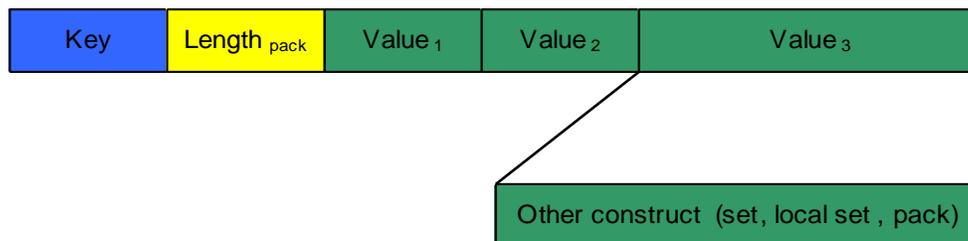


**Figure 5-2: Example of Variable Length Pack Structure**

Variable length packs are valuable in applications where the endpoints have a specific set of elements to share defined in a standard or recommended practice (RP) agreed upon by all users. Variable length packs can provide greater bandwidth efficiency than local sets if most of the data elements occur at a fairly common frequency, or the number of elements is small enough such that the overhead incurred in setting the length of empty elements is less than defining a local set key for all elements.

### 5.2.4 Background: Summary Definition of Common Metadata Set Floating Length Pack

Refer to MISB RP0701 for the complete definition of a floating length pack. A floating length pack is identical to a standard fixed length pack except the last elements' length is allowed to "float". This allows the pack to contain another pack construct within it.



**Figure 5-3: Example of Floating Length Pack Structure**

### 5.2.5 Background: Summary Definition of Common Metadata Set Truncated Pack

Refer to MISB RP0701 for the complete definition of a truncated pack. A truncated pack is identical to a standard fixed length pack except the length of the pack can be reduced so only a portion of the values in the pack are used. This allows items not frequently used to be placed at the end of the pack and ignored when the length doesn't include them.

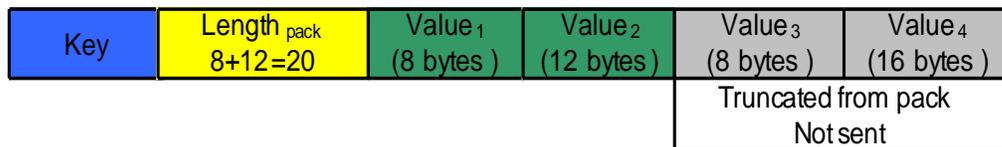


Figure 5-4: Example of Truncated Pack Structure

### 5.2.6 Choosing the Right Metadata Construct

The choice in a metadata construct is dependent on the application. There is on-going debate regarding the use of XML as a metadata type in low-bandwidth MI. The merits of XML are that it is a human readable format and that it facilitates ready exchange of data between applications because many tools exist for such purposes. Cursor on Target is defined in XML, and the MISB has defined a one-to-one mapping between the XML format and KLV for CoT. XML inherently has a greater overhead and XML schemas differ which greatly challenges adherence to interoperability. Work on reducing XML's overhead (binary XML) and the development of appropriate XML schemas for military application continues, and the ultimate utility of XML may lie as a global metadata descriptor of content rather than description at the more granular frame-level of MI essence.

## 5.3 Audio

Content delivered over bandwidth-challenged networks may not include audio, but if required then MPEG-1 Layer II, MPEG-2 Layer II, and MPEG-2 AAC-LC are recommended (see MISB EG 1001).

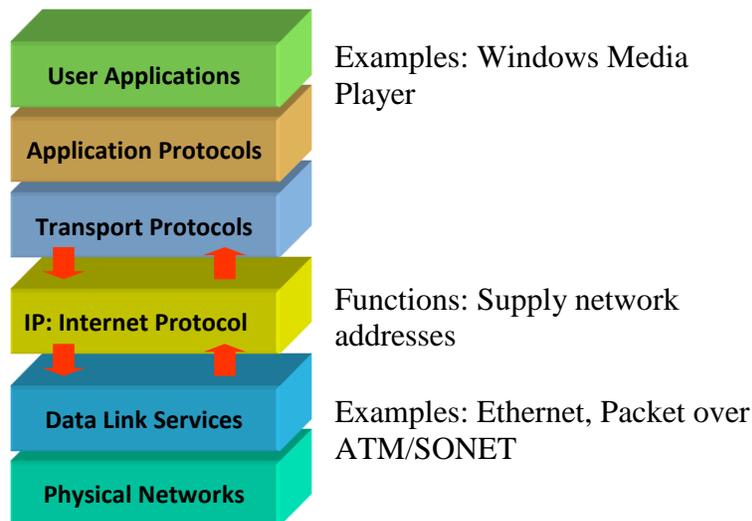
## 6 Preparing Content for Delivery

### 6.1 IP Basics

Internet Protocol (IP) depends on other software and hardware to function, and other software likewise depends on IP. IP unto itself is not reliable, but depends on other protocols to provide those functions. Figure 6-1 shows where IP fits into the hierarchy of data communications. Although quality of communications depends on the stack of technologies shown, the discussion here focuses on the layers above IP: the *Transport Protocols* (TCP, UDP), the *Application Protocols* (RTP, RTCP, RTSP, HTTP), and (to a lesser extent) the *User Applications* in the movement of motion imagery across IP.

The choice of a protocol is largely determined by whether the content is downloaded (like a file) or streamed. If the content is streamed, the choice will then depend on whether the steaming is progressive download or real-time; and for real-time streaming whether the stream is to be unicast or multicast. Finally, the method of transport protocol will be dictated by whether the channel is private (or dedicated) or shared (or public).

In file download HTTP is used to request a file through a browser hyperlink, and the content is transferred over TCP for guaranteed delivery, albeit in non-real-time. In unicast streaming RTCP and RTSP provide player feedback for QoS and content control respectively, while the actual content is delivered over RTP/UDP. These protocols are further described next.



**Figure 6-1: Data Communications Stack**

## 6.2 Mapping Essence to Packets

All encoded essence—video or audio—flows as one continuous stream from its respective encoding source. This encoded format, however, is not appropriate for IP delivery. Each encoded essence raw stream—called an Elementary Stream (ES) must first be segmented into packets (PES for Packetized Elementary Stream) suitable to a protocol for IP delivery. This process, called *encapsulation*, takes an elementary stream, formats it into packets, and adds appropriate headers and other information required to comply with the specific protocol used.

Transport protocols control the transmission of data packets in concert with IP. The two most common protocols are UDP (User Datagram Protocol), and TCP (Transmission Control Protocol). Data formatted according to UDP is referred to as UDP/IP, and TCP/IP when formatted according to TCP. Time-sensitive data, such as video and audio, rely on UDP for delivery because UDP will not request a resend of lost packets, which would stall smooth delivery and waste unnecessary bandwidth. UDP does not guarantee delivery—it is a “best efforts” protocol. The IP header of 20 bytes and the UDP header of 8 bytes together accompany every packet for a per-packet overhead of 28 bytes. Total allowable packet size is set by the transmission medium; as an example Ethernet the packets must be less than 1500 bytes.



**Figure 6-2: UDP/IP Packet**

### 6.3 Established MISB Standards

The MISB identifies standards—both commercial and those specific to the MISB—to ensure interoperability for the sharing of motion imagery assets. The standards embodied in MISB-compliant systems have proven sufficient for interchange and delivery across robust links, where channel capacity is sufficient to sustain high-fidelity motion imagery with a guaranteed quality of service. These standards include MPEG-2 and H.264/AVC compression carried within an MPEG-2 Transport Stream—referred to as Xon2, where X is the compression type.

### 6.4 MPEG-2 TS and RTP

#### 6.4.1 Overview

MPEG-2 Transport Stream was intended for the transmission of synchronized audio, video and data primarily for cable, terrestrial, and satellite television. Support for simultaneous transmission of multiple programs and error resilience in wireless transmission were primary design requirements. The multiplexed stream, small packet size, and optional Forward Error Correction make it ideal for RF Transmission and private networks. MPEG-2 TS will be a better choice than RTP for broadcast applications. Mobile TV, aimed to deliver television quality video over handheld devices such as cell phones and PDA's, is principally a broadcast modality and the specifications by those supplying mobile TV services identify MPEG2-TS as a protocol.

“RTP (Real-Time Protocol) provides a flexible framework for delivery of real-time media, such as audio and video, over IP networks. Its core philosophies—application-level framing and the end-to-end principle—make it well suited to this unique environment. Application-level framing comes from the recognition that there are many ways in which an application can recover from network problems, and that the correct approach depends on both the application and the scenario in which it is being used. In some cases it is necessary to retransmit an exact copy of the lost data. In others, a lower-fidelity copy may be used, or the data may have been superseded, so the replacement is different from the original. Alternatively, the loss can be ignored if the data was of only transient interest. These choices are possible only if the application interacts closely with the transport. The philosophy of application-level framing implies smart, network-aware applications that are capable of reacting to problems. The end-to-end principle implies that intelligence is at the endpoints, not within the network.” *RTP: Audio and Video for the Internet -- Colin Perkins.*

RTP was created for carriage of real time imagery, audio, and other time-sensitive data over IP networks. RTP provides over MPEG2-TS:

- ❑ Optimization of user experience under varying network conditions (better resiliency)
- ❑ Client control of session via RTSP (pause, rewind, fast forward, etc.)
- ❑ Flexibility in sending individual component media streams (video or just metadata)
- ❑ Improved services allowing additional streams containing user-specific or content-specific data to be streamed and synchronized

Quite simply, MPEG-2 TS and RTP were designed for different purposes, so choosing one over another is purely a matter of application. In the delivery of motion imagery over bandwidth-limited channels either can be used; however, there will be tradeoffs in the quality of content received, the degree of control a user has over that content, the flexibility in requesting a subset of the original media package, processing at the server, and overhead.

MPEG-2 TS is greatly leveraged in the broadcast industry and finds great utility to this day. On the other hand, RTP is an industry standard for carrying video and audio over the internet, and is even preferred in some situations where TS can be used. Both are found in high-QoS managed IP networks. Tools and understanding in applying either are readily available.

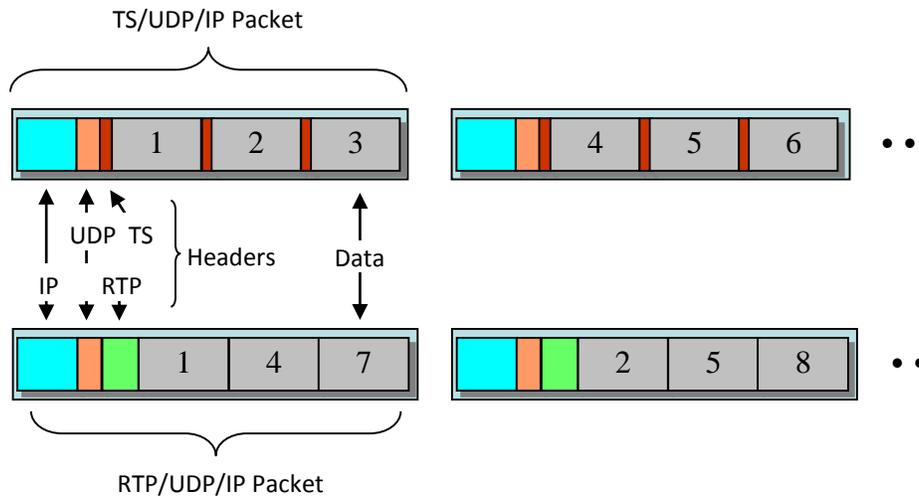
In the commercial broadcast community where reliability of stream delivery is critical and often takes precedence over latency, RTP is recommended as a carrier when transporting MPEG-2 TS packets over UDP. In this case, an RTP header precedes a number of TS packets, which as a group are then packaged into a UDP datagram. RTP serves to provide timing effective in correcting packet jitter, and sequencing effective in detecting out of order and lost packets.

#### 6.4.2 Header Overhead

It is perfectly legitimate to deliver MPEG2-TS either directly over UDP/IP or over RTP/UDP/IP, although the DoD specifies UDP/IP as the carrier for MPEG-2 TS. The additional header bytes required for the various protocols are shown in Table 6-1. Figure 6-3 shows the ordering of the headers in a packet.

Packetization	Bytes in Header
IP	20
UDP	8
RTP	12
TS	4

**Table 6-1: Header size for various protocols**



**Figure 6-3: Transport Stream over RTP and Native RTP. Illustrates optional reordering in RTP for error resiliency**

### 6.4.3 Bidirectional Feedback

Both RTP and MPEG-2 TS are unidirectional protocols. Both can send data over UDP/IP. Any bidirectional aspect to RTP is provided by a second protocol called RTCP (Real Time Control Protocol). RTCP provides bidirectional communications between the sender and the receiver. It allows the sender to provide information to the receiver such as how many bytes and packets have been sent, and it allows the receiver to provide information to the sender such as how many packets were lost, and a measure of the packet arrival jitter. RTCP is important in synchronizing media streams, such as for lip sync, at the receiver as it carries important time reference information.

### 6.4.4 Packetization

In RTP, each media type (video, audio, metadata) is sent as a separate RTP stream. Multiplexing is done by the network layer. Timestamps in the RTP headers aid to re-synchronize the streams at the decoder. Sequence numbers in the RTP header allow the receiver to detect packet loss. In MPEG-2 TS, all media is multiplexed into one composite data stream. Timestamps in the PES headers are recovered to synchronize the streams at the decoder. Continuity counts in the TS packets allow the decoder to detect TS packet loss.

Separate RTP streams permit a client to request a subset of the media types, thereby saving considerable bandwidth. For MPEG-2 TS, on the other hand, a client must either accept the composite media package or the server must demultiplex the transport stream into individual media components and then re-multiplex the requested subset, which is not efficient.

### 6.4.5 Error Resilience

Both MPEG-2 TS and RTP over IP rely on UDP. UDP packets have a CRC (Cyclic Redundancy Code) code. If there are bit errors, the CRC check fails and the entire packet is discarded. There is no difference between the two protocols at this level.

RTP packets tend to be variable length based on the content. This allows the RTP packets to start on picture or slice boundaries, and to only contain that information. Should a packet get lost, the loss is limited to that information alone, and if slices are used, the decoder can resynchronize on the next RTP packet. As a step towards ensuring resiliency packets in RTP can be shuffled to prevent transmission errors from impacting adjacent media data. Since reshuffling introduces delay upon reconstruction at the receiver, this must be taken into consideration as it impacts overall stream latency.

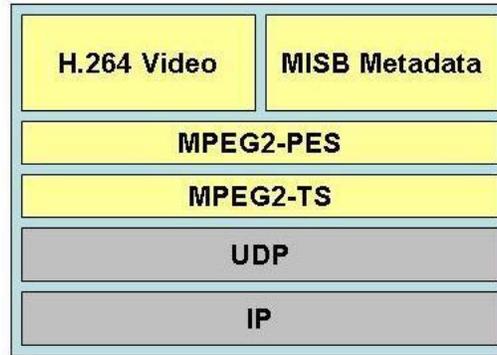
Typically, the number of MPEG-2 TS packets that are placed into a UDP packet is controllable up to seven TS packets per UDP packet (for a maximum transmission unit (MTU) of 1500 bytes). These TS packets can be system, video, audio, or data. With a fixed number of TS packets in a UDP packet the loss of one UDP packet may impact TS packets from one video frame, from the end of the previous frame and the beginning of the next, from video and audio, or other combinations. Fewer TS packets placed in a UDP packet will minimize the effect of packet loss, but at the expense of stream overhead. Also, a variable number of TS packets can be placed in a UDP packet to approximate the resiliency of RTP. Forward Error Correction (FEC) can be applied to MPEG2 TS at the expense of additional overhead and latency.

### 6.4.6 Synchronization

MPEG-2 TS synchronizes the decoder to the encoder through the PCR (Program Clock Reference), which is based on a 27 MHz reference clock that is present at the encoder. This assures that the decoder is in lock step with the encoder, and that no frames need to be dropped or repeated during playback. RTP has a similar facility using the RTCP, which links the RTP packet timestamp and NTP (Network Time Protocol) or other wall clock.

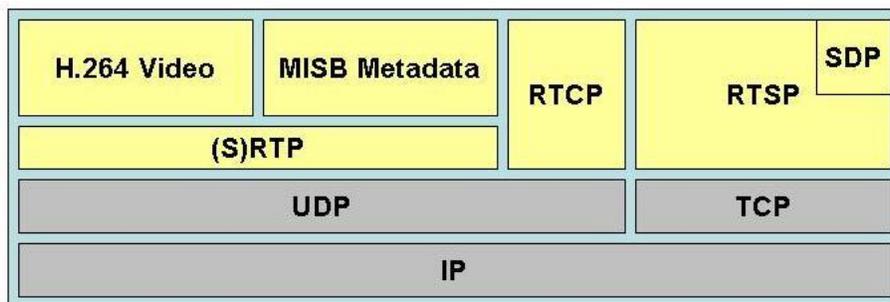
## 6.5 Other Supporting Technologies

Figure 6-4 illustrates various MI technologies for IP delivery. Figure 6-4a denotes the technologies used in delivering video-over-IP with the MPEG-2 Transport Stream. Video and metadata encoded elementary streams are segmented into packets (MPEG-2 PES) and then mapped into MPEG-2 TS, which is then delivered over UDP/IP.



**Figure 6-4a. Video over IP – MPEG2 TS**

In Figure 6-4b, video and metadata are delivered over RTP. One can see that UDP rides above IP as a protocol, and similarly, RTP rides one layer above UDP. Secure Real Time Protocol (SRTP) defines a profile of RTP, intended to provide encryption, message authentication and integrity, and replay protection to the RTP data in both unicast and multicast applications. SRTP adds a trailer with cryptographic metadata to the end of the RTP payload, which contains information for encryption and/or authentication for the RTP payload. As would be expected more overhead is incurred in delivering SRTP than RTP. Otherwise, the operation of SRTP is identical to RTP. RTSP (Real Time Streaming Protocol) provides a means for users to control the communications session. Much like a VCR remote control, RTSP allows a user to skip, rewind, and fast forward, etc. media. RTSP is a companion protocol to RTP, and is often carried like RTCP—on TCP/IP.



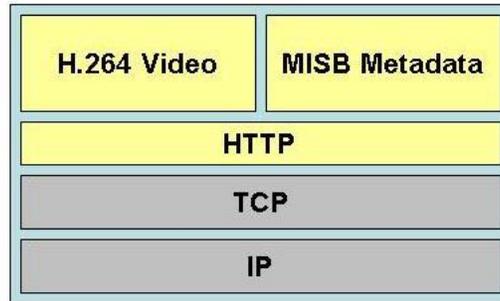
**Figure 6-4b. Video over IP – Native RTP**

Session Initiation Protocol (SIP) is a signaling protocol used to set up sessions. Devices that wish to communicate, for example a sensor sending data to a client, use SIP to communicate about how to address one another, how they are each configured, and the types of streams they wish to send and receive. SIP typically relies on a Proxy Server outside the actual transaction path for the media to host information about requesting devices, such as location and security. As a client moves on and off the network, SIP can provide the sender/receiving applications information on how they can reconnect.

Finally, Session Description Protocol (SDP) operates within SIP but at the endpoints for devices to describe their capabilities and the types of stream they are supporting. SDP can specify the media type (video, metadata, etc.), the encoding format (MPEG-2, H.264/AVC, etc.) and the

transport protocol used (RTP, etc.) Once both ends agree on the specifics, then communication can occur.

Figure 6-4c denotes the technologies employed in an HTTP file download (or progressive video download).



**Figure 6-4c. Video over IP – HTTP Download**

## 6.6 Web Servers versus Streaming Servers

Web servers, which rely on HTTP (Hypertext Transfer Protocol), download content over IP in a file transfer mode—they do not stream content per se. HTTP is a response/request protocol between a client and a server, and is used in most web applications. Web servers offer no facility to control the delivery of the stream, so that if network congestion is high the delivery speed will be low; if the network capacity is high, the packets may arrive in bursts.

Newer methods to deliver streamed video over the internet do use HTTP coupled with short segments of video. These adaptive streaming protocols rely on TCP/IP with various data rate encoded versions of the original content stored on a standard web server. As network conditions change the client can request an encoded version consistent with current channel bandwidth.

Streaming servers process multimedia data under timing constraints and support interactive control functions such as pause/resume, fast forward, rewind. The streaming servers are responsible for serving video, audio, slides and other components in a synchronous fashion. Streaming offers functions like real-time flow control, intelligent stream switching, and interactive media navigation. They are also designed to serve many streams simultaneously; for example, the QuickTime streaming server can serve 4,000 simultaneous streams. Streaming servers rely on hint tracks, which are additional tracks created to control the stream, as pointers to the RTP information needed to serve the relevant media chunks.

To access a stream, the client player issues a request to the streaming server to stream a file. Typically, this is done by requesting content that is posted with a hyperlink on a website. This directs the server to issue the client an address and the filename of the content. The client sends this information back over RTSP to the streaming server, which then delivers the content over RTP.

A streaming server is required to support RTP/RTSP.

## 7 Architecture

Choice of whether to use MPEG-2 TS or RTP as the transport vehicle will be a function of the intended application and channel capacity/reliability. The constraints of the system and the functional environment will ultimately dictate the compression type, metadata format, transport method, and security type/control.

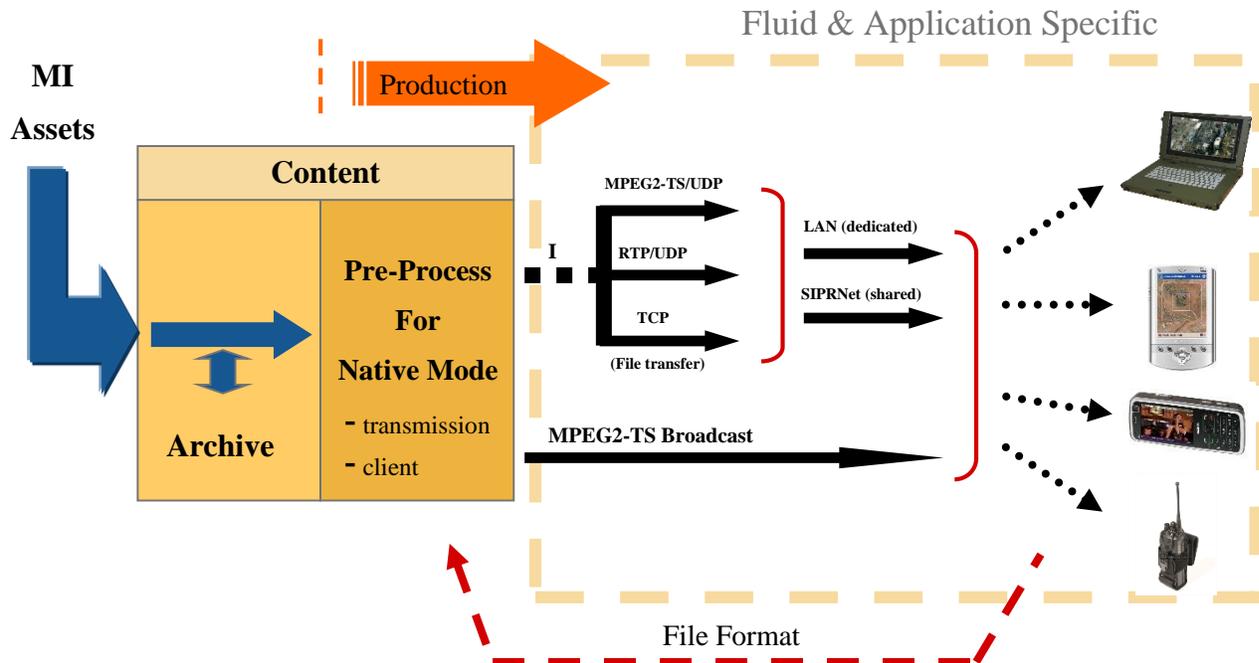
Important considerations from a systems design perspective include:

- The overall topology of the network
- Quality of service and expected (or required) user experience
- Live real-time versus “real time” video-on-demand capabilities
- CONOPS sensitive to stream latency
- The client device and its overall capabilities, such as stream control

Choice in a client will guide the selection of best compression for that player and how the data should be formatted. Aside from file format re-wrapping, which can be done without impacting the encoded content, care must be taken when transcoding one encoded video type to another. Generally, transcoding requires a decoding of the compressed signal back to its uncompressed state followed by subsequent compression to the target encoded format. The lower the resolution and quality of the encoded source video the greater the degradation imposed on the transcoded output.

There is *no* loss, however, in changing a files format or its transport carrier. For example, if a MPEG-2 TS stream is received directly from a platform but is best delivered to the client device over RTP, then the video, audio, and metadata elementary streams can be de-multiplexed from the MPEG-2 transport stream and remapped into RTP. Similarly, if content is archived in MXF and the client requires an MP4 format the content can be first unwrapped and then rewrapped into the MP4 format for final delivery over RTP.

Figure 7-1 indicates options when dispensing a motion imagery asset directly from a sensor or an archive. A pre-process step may be needed to change the asset to another form, such as an encoding type, bit rate, transport protocol, or player file format. Client and network dynamics will determine the type of preprocessing necessary. Networks that are well managed with a guaranteed quality-of-service (QoS)—denoted by the “LAN” designation offer the greatest flexibility in choosing a transport protocol. Generally, a less reliable shared network (perhaps like the SIPRNet) will require feedback in regulating the network to improve the received experience. Either transport could be used in a LAN application, but if stream control is required then RTP would be the preferred choice. RTP is likewise the preferred choice for shared networks.



**Figure 7-1: Archive to Client Device Network Choices**

Mentioned here only briefly is the option to *broadcast* media not over IP but rather by RF (Radio Frequency) similar to that in commercial television broadcasting. Standards such as DVB-H and MediaFLO for Mobile TV offer the capability to broadcast video to cell phones, and do so within a MPEG-2 transport stream protocol. Backchannel control of the server by the client may be provided through traditional cell phone links. These systems are now being deployed for mobile TV to the cell phone, but offer interesting potential for government application as well.

## 8 Final Thoughts

The identification of SA as a “product” can now be better appreciated. Native MI assets in most cases will be subject to a “publication stage” prior to delivery to the end user. But the technologies that drive the publication stage are application specific in that a given system configuration will exhibit its own networking capabilities, constraints, and client receivers. The application itself will drive technology choices.

Most networks can be classified as either circuit-based or packet-based. Traditional circuit-based telephony provides a dedicated channel for each user. This type of network offers reliable transmission, and hence can be used to deliver MPEG-2 TS or RTP streams if the information is packetized and sent according to the IP communications model. Packet-based networks, such as the internet, are often shared amongst a number of services. The “public” nature of the network creates issues that may prevent reliable delivery of MPEG-2 TS streams. In these cases, a protocol like RTP may provide better optimization and resilience to such issues because of the corresponding QoS feedback control afforded by RTCP.