RTP Payload for Haptics

Abstract

This memo describes an RTP payload format for the MPEG-I haptic data. A haptic media stream is composed of MIHS units including a MIHS unit header and zero or more MIHS packets. The RTP payload header format allows for packetization of a MIHS unit in an RTP packet payload as well as fragmentation of a MIHS unit into multiple RTP packets.

Status of This Memo

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

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

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

This Internet-Draft will expire on 18 April 2024.

Copyright Notice

Copyright (c) 2023 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Haptics provides users with tactile effects in addition to audio and video, allowing them to experience sensory immersion. Haptic data is mainly transmitted to devices that act as actuators and provides them with information to operate according to the values defined in haptic effects. The IETF is registering haptics as a primary media type akin to audio and video [I-D.ietf-mediaman-haptics].

The MPEG Haptics Coding standard [ISO.IEC.23090-31] defines the data formats, metadata, and codec architecture to encode, decode, synthesize and transmit haptic signals. It defines the "MIHS unit" as a unit of packetization suitable for streaming, and similar in essence to the NAL unit defined in some video specifications. This document describes how haptic data (MIHS units) can be transmitted using the RTP protocol. This document followed recommendations in [RFC8088] and [RFC2736] for RTP payload format writers.
2. 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].

3. Definition

This document uses the definitions of the MPEG Haptics Coding standard [ISO.IEC.23090-31]. Some of these terms are provided here for convenience.

Actuator: component of a device for rendering haptic sensations.

Avatar: body (or part of body) representation.

Band: component in a channel for containing effects for a specific range of frequencies.

Channel: component in a perception containing one or more bands rendered on a device at a specific body location.

Device: physical system having one or more actuators configured to render a haptic sensation corresponding with a given signal.

Effect: component of a band for defining a signal, consisting of a haptic waveform or one or more haptic keyframes.

Experience: top level haptic component containing perceptions and metadata.

Haptics: tactile sensations.

Keyframe: component of an effect mapping a position in time or space to an effect parameter such as amplitude or frequency.

Metadata: global information about an experience, perception, channel, or band.

Modality: type of haptics, such as vibration, force, pressure, position, velocity, or temperature.

Perception: haptic perception containing channels of a specific modality.

Signal: representation of the haptics associated with a specific modality to be rendered on a device.
4. Haptic Format Description

4.1. Overview of Haptic Coding

The MPEG Haptics Coding standard specifies methods for efficient transmission and rendering of haptic signals, to enable immersive experiences. It supports multiple types of perceptions, including the most common vibrotactile (sense of touch that perceives vibrations) and kinaesthetic perceptions (tactile resistance), but also other, less common perceptions, including for example the sense of temperature or texture. It also supports two approaches for encoding haptic signals: a "quantized" approach based on samples of measured data, and a "descriptive" approach where the signal is synthesized using a combination of functions. Both quantized and descriptive data can be encoded in a human-readable exchange format based on JSON (.hjif), or in a binary packetized format for distribution and streaming (.hmpg). This last format is referred to as the MPEG-I Haptic Stream (MIHS) format and is a base for the RTP payload format described in this document.

4.2. MPEG-I Haptic Stream (MIHS) format

MIHS is a stream format used to transport haptic data. Haptic data including haptic effects is packetized according to the MIHS format, and delivered to actuators, which operate according to the provided effects. The MIHS format has two level packetization, MIHS units and MIHS packets.

MIHS units are composed of a MIHS unit header and zero or more MIHS packets. Four types of MIHS units are defined. An initialization MIHS unit contains MIHS packets carrying metadata necessary to reset and initialize a haptic decoder, including a timestamp. A temporal MIHS unit contains one or more MIHS packets defining time-dependent effects and providing modalities such as pressure, velocity, and acceleration. The duration of a temporal unit is a positive number. A spatial MIHS unit contains one or more MIHS packets providing time-independent effects, such as vibrotactile texture, stiffness, and friction. The duration of a spatial unit is always zero. A silent MIHS unit indicates that there is no effect during a time interval and its duration is a positive number.

A MIHS unit can be marked as "sync" (i.e., independent) or "non-sync" (i.e., dependent). When a decoder processes a sync unit, it resets the previous effects and therefore provides a haptic experience independent from any previous MIHS unit. A non-sync unit is the continuation of previous MIHS units and cannot be independently decoded and rendered without having decoded previous MIHS unit(s). Initialization and spatial MIHS units are always sync units. Temporal and silent MIHS units can be sync or non-sync units.
4.3. MIHS transmission Considerations

The following considerations apply for the streaming of MIHS over RTP:

*A media sender SHOULD set durations with the same value for all non-zero duration MIHS units between initialization MIHS units, to make the decoder more robust to RTP packet loss.

*A haptic silence suppression mechanism SHOULD be used. A sender MAY send the first (or first few) MIHS silent units at the beginning of a haptic silence. A sender SHOULD NOT send subsequent consecutive silent units, to save network resources. Following the reception of a MIHS silent unit, a receiver SHOULD interpret subsequent lost MIHS units as silent MIHS units, until the reception of a non-silent MIHS unit.

TBD: these considerations may need to be re-evaluated depending on the finalization of the haptics coding specifications in MPEG.

5. Payload format for haptics

5.1. RTP header Usage

The RTP header is defined in [RFC3550] and represented in Figure 2. Some of the header field values are interpreted as follows.
Figure 2: RTP header for Haptic.

Payload type (PT): 7 bits. The assignment of a payload type MUST be performed either through the profile used or in a dynamic way.

Time Stamp (TS): 32 bits. A timestamp representing the sampling time of the first sample of the MIHS unit in the RTP payload. The clock frequency MUST be set to the sample rate of the encoded haptic data and is conveyed out-of-band (e.g., as an SDP parameter).

Marker bit (M): 1 bit. The marker bit SHOULD be set to one in the first non-silent RTP packet after period of haptic silence. This enables jitter buffer adaptation and haptics device washout (i.e., reset to a neutral position) prior to the beginning of the burst with minimal impact on the quality of experience for the end user. The marker bit in all other packets MUST be set to zero.

5.2. Payload Header

The RTP Payload Header follows the RTP header. Figure 3 describes RTP Payload Header.

Figure 3: RTP payload header for Haptic.

D (Dependency, 1 bit): this field is used to indicate whether the MIHS unit included in the RTP payload is, when its value is one, dependent (i.e., "non-sync") or, when its value is zero, independent (i.e., "sync"). In case of congestion, a receiver or intermediate node MAY prioritize independent packets over dependent ones, since the non-reception of an independent MIHS unit can prevent the decoding of multiple subsequent dependent MIHS units.

UT (Unit Type, 3 bits): this field indicates the type of the MIHS unit included in the RTP payload. In case of congestion, a receiver or intermediate node MAY prioritize initialization MIHS units over other units, since initialization MIHS units contain metadata used to re-initialize the decoder, and MAY drop silent MIHS units before other types of MIHS units, since a receiver may interpret a missing MIHS unit as a silence. UT field values are listed in Figure 4.

L (MIHS Layer, 4 bits): this field is an integer value which indicates the priority order of the MIHS unit included in the RTP payload, as determined by the haptic sender (e.g., by the haptic...
codec), based on application-specific needs. For example, the sender
may use the MIHS layer to prioritize perceptions with the largest
impact on the end-user experience. Zero corresponds to the highest
priority. In case of congestion, intermediate nodes and receivers
SHOULD use the MIHS layer value to determine the relative importance
of haptic RTP packets.

5.3. Payload Structures

Two different types of RTP packet payload structures are specified.
The single unit payload structure contains a single MIHS unit. The
fragmented unit payload structure contains a subset of a MIHS unit.
The unit type (UT) field of the RTP payload header Figure 4
identifies both the payload structure and, in the case of a single
unit structure, also identifies the type of MIHS unit present in the
payload.

TBD: consider if it would be useful to add aggregation.

<table>
<thead>
<tr>
<th>Unit Type</th>
<th>Payload Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>1</td>
<td>Single</td>
</tr>
<tr>
<td>2</td>
<td>Single</td>
</tr>
<tr>
<td>3</td>
<td>Single</td>
</tr>
<tr>
<td>4</td>
<td>Single</td>
</tr>
<tr>
<td>7</td>
<td>Frag</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization MIHS Unit</td>
</tr>
<tr>
<td>Temporal MIHS Unit</td>
</tr>
<tr>
<td>Spatial MIHS Unit</td>
</tr>
<tr>
<td>Silent MIHS Unit</td>
</tr>
<tr>
<td>Fragmented Packet</td>
</tr>
</tbody>
</table>

Figure 4: Payload structure type for haptic

The payload structures are represented in Figure 4. The single unit
payload structure is specified in Section 5.3.1. The fragmented unit
payload structure is specified in Section 5.3.2.

Figure 5: RTP Transmission mode
5.3.1. Single Unit Payload Structure

In a single unit payload structure, as described in Figure 5, the RTP packet contains the RTP header, followed by the payload header and one single MIHS unit. The payload header follows the structure described in Section 5.2. The payload contains a MIHS unit as defined in ISO/IEC.23090-31.

Figure 6: Single unit payload structure

5.3.2. Fragmented Unit Payload Structure

In a fragmented unit payload structure, as described in Figure 7, the RTP packet contains the RTP header, followed by the payload header, a Fragmented Unit (FU) header, and a MIHS unit fragment. The payload header follows the structure described in Section 5.2. The value of the UT field of the payload header is 7. The FU header follows the structure described in Figure 8.

Figure 7: Fragmentation unit header

FU headers are used to enable fragmenting a single MIHS unit into multiple RTP packets. Fragments of the same MIHS unit MUST be sent in consecutive order with ascending RTP sequence numbers (with no
other RTP packets within the same RTP stream being sent between the first and last fragment). FUs MUST NOT be nested, i.e., an FU MUST NOT contain a subset of another FU.

Figure 8 describes a FU header, including the following fields:

```
+---------------------------+   +-------------------+
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
+---+---+---+---+---+---+---+---+
|FUS|FUE| RSV | UT   |
+---+---+-----------+-----------+
```

Figure 8: Fragmentation unit header

FUS (Fragmented Unit Start, 1 bit): this field MUST be set to 1 for the first fragment, and 0 for the other fragments.

FUE (Fragmented Unit End, 1 bit): this field MUST be set to 1 for the last fragment, and 0 for the other fragments.

RSV (Reserved, 3 bits): these bits MUST be set to 0 by the sender and ignored by the receiver.

UT (Unit Type, 3 bits): this field indicates the type of the MIHS unit this fragment belongs to, using values defined in Figure 4.

6. Payload format parameters

This section specifies the optional parameters. A mapping of the parameters into the Session Description Protocol (SDP) [RFC8866] is also provided for applications that use SDP. Equivalent parameters could be defined elsewhere for use with control protocols that do not use SDP.

6.1. Media type registration

The receiver MUST ignore any parameter unspecified in this memo. This memo defines the 'hmpg' haptic subtype for use with the MPEG-I haptics streamable binary coding format described in [ISO.IEC.23090-31].

*Type name: haptics

*Subtype name: hmpg

*Required parameters: N/A

*Optional parameters are defined in the following section.
6.2. Optional parameters definition

hmpg-ver provides the year of the edition and amendment of ISO/IEC 23090-31 that this file conforms to, as defined in [ISO.IEC.23090-31].

hmpg-profile indicates the profile used to generate the encoded stream as defined in [ISO.IEC.23090-31]. The current possible values are "simple-parametric" and "main".

hmpg-lvl indicates the level used to generate the encoded stream as defined in [ISO.IEC.23090-31].

hmpg-maxlod indicates the maximum level of details to use for the avatar(s). The avatar level of detail (LOD) is defined in [ISO.IEC.23090-31].

hmpg-avtypes indicates, using a coma-separated list, types of haptic perception represented by the avatar(s). The avatar type is defined in [ISO.IEC.23090-31].

hmpg-modalities indicates, using a coma-separated list, haptic perception modalities (e.g., pressure, acceleration, velocity, position, temperature, etc.). The perception modality is defined in [ISO.IEC.23090-31].

hmpg-bodypartmask indicates, using a bitmask, the location of the devices or actuators on the body. The body part mask is defined in [ISO.IEC.23090-31].

hmpg-maxfreq indicates the maximum frequency of haptic data for vibrotactile perceptions (Hz). Maximum frequency is defined in [ISO.IEC.23090-31].

hmpg-minfreq indicates the minimum frequency of haptic data for vibrotactile perceptions (Hz). Minimum frequency is defined in [ISO.IEC.23090-31].

hmpg-dvctypes indicates, using a coma-separated list, the types of actuators. The device type is defined in [ISO.IEC.23090-31].

hmpg-silencesupp indicates whether silence suppression should be used (1) or not (0). The default value shall be 1.

7. SDP Considerations

The mapping of above defined payload format media type to the corresponding fields in the Session Description Protocol (SDP) is done according to [RFC8866].
The media name in the "m=
line of SDP MUST be haptics.

The encoding name in the "a=rtpmap" line of SDP MUST be hmpg.

The clock rate in the "a=rtpmap" line may be any sampling rate, typically 8000.

The OPTIONAL parameters (defined in Section 6.2), when present, MUST be included in the "a=fmtp" line of SDP. This is expressed as a media type string, in the form of a semicolon-separated list of parameter=value pairs.

An example of media representation corresponding to the hmpg RTP payload in SDP is as follows:

```
m=haptics 43291 UDP/TLS/RTP/SAVPF 115
a=rtpmap:115 hmpg/8000
a=fmtp:115 hmpg-profile=1;hmpg-lvl=1;hmpg-ver=2023
```

### 7.1. SDP Offer/Answer Considerations

When using the offer/answer procedure described in [RFC3264] to negotiate the use of haptic, the following considerations apply:

The haptic signal can be sampled at different rates. The MPEG Haptics Coding standard does not mandate a specific frequency. A typical sample rate is 8000Hz.

The parameter 'hmpg-ver' indicates the version of the haptic standard specification. If it is not specified, the initial version of the MPEG Haptic Coding specification SHOULD be assumed, although the sender and receiver MAY use a specific value based on an out-of-band agreement. The parameter 'hmpg-profile' is used to restrict the number of tools used (e.g., the simple-parametric profile fits enable simpler implementations than the main profile). If it is not specified, the most general profile "main" SHOULD be assumed, although the sender and receiver MAY use a specific value based on an out-of-band agreement. The parameter 'hmpg-lvl' is used to further characterize implementations within a given profile, e.g., according to the maximum supported number of channels, bands, and perceptions. If it is not specified, the most general level "2" SHOULD be assumed, although the sender and receiver MAY use a specific version based on an out-of-band agreement.

Other parameters can be used to indicate bitstream properties as well as receiver capabilities. The parameters 'hmpg-maxlod', 'hmpg-avtypes', 'hmpg-bodypartmask', 'hmpg-maxfreq', 'hmpg-minfreq', 'hmpg-dvctypes', and 'hmpg-modalities' can be sent by a sender to reflect the characteristics of bitstreams and can be set by a receiver to reflect the nature and capabilities of local actuator...
devices, or a preferred set of bitstream properties. For example, different receivers may have different sets of local actuators, in which case these parameters can be used to select a stream adapted to the receiver. In some other cases, some receivers may indicate a preference for a set of bitstream properties such as perceptions, min/max frequency, or body-part-mask, which contribute the most to the user experience for a given application, in which case these parameters can be used to select a stream which include and possibly prioritizes those properties.

The parameter 'hmpg-silencesupp' can be used to indicate sender and receiver capabilities or preferences. This parameter indicates whether silence suppression should be used, as described in Section 4.3.

7.2. Declarative SDP considerations

When haptic content over RTP is offered with SDP in a declarative style, the parameters capable of indicating both bitstream properties as well as receiver capabilities are used to indicate only bitstream properties. For example, in this case, the parameters hmpg-maxlod, hmpg-bodypartmask, hmpg-maxfreq, hmpg-minfreq, hmpg-dvctypes, and hmpg-modalities declare the values used by the bitstream, not the capabilities for receiving bitstreams. A receiver of the SDP is required to support all parameters and values of the parameters provided; otherwise, the receiver MUST reject or not participate in the session. It falls on the creator of the session to use values that are expected to be supported by the receiving application.

8. Congestion control consideration

The general congestion control considerations for transporting RTP data apply to HMPG haptics over RTP as well [RFC3550].

It is possible to adapt network bandwidth by adjusting either the encoder bit rate or by adjusting the stream content (e.g., level of detail, body parts, actuator frequency range, target device types, modalities). It is also possible, using the layer field of the RTP payload header, to allocate MIHS units to different layers based on their content, to prioritize haptic data contributing the most to the user experience.

9. Security Considerations

TBD

10. IANA Considerations

A new media type will be registered with IANA; see Section 6.
11. References

11.1. Normative References


11.2. Informative References


Authors' Addresses

Hyunsik Yang
InterDigital
United States of America

Email: hyunsik.yang@interdigital.com

Xavier de Foy
InterDigital
Canada

Email: xavier.defoy@interdigital.com