MIMI Content Format

draft-ietf-mimi-content-02

Rohan Mahy, IETF 119, 22–Mar–2024

What's new in -02?

- Replaced the explicit **message ID** with one calculated as the **hash of the ciphertext**
- Removed the timestamp. The protocol document provides a **hub-accepted timestamp** to any fanned out provider. The client might receive this.
- These **deprecate** the **messageId** and **timestamp** shared in the MLS additional authenticated data field from the previous draft (see Slide 11)

- Also need to confirm some changes from -01 version

Attachments via External Content

- Created new binary format modeled on message/external-body (RFC 4483)

```
body.disposition = attachment;
body.contentType = "video/mp4";
body.URL = "https://example.com/storage/bigfile.mp4";
body.size = 708234961;
body.encAlg = 0x0001;    // AES-128-GCM
body.key = "\x21399320958a6f4c745dde670d95e0d8";
body.nonce = "\xc86cf2c33f21527d1dd76f5b";
body.aad = "";
body.expires = 0;
body.description = "2 hours of key signing video";
```

- Sending client encrypts and uploads (usually to local provider); sends external content fields in room
- Receiving client downloads according to their local policy; decrypts
  - Expected that **room policy** will say that attachments are either in the sender's provider's domain or the hub's domain. To prevent the "web bug" see MIMI protocol PR #75.

Sort order of messages

- Strong consensus at last IETF was that the ability of clients to consistently rendering sort order is a required feature of MIMI.
- Currently the draft includes a lastSeen header with the messageId of the last message the sender has seen. If multiple messages arrive with the same lastSeen header, the receiver will list all of those messages in its lastSeen header (“merging” the graph)
- The client still needs to guard against attacks where this header is maliciously generated. There is text to prevent against these attacks in Section 8.1.
- If it is available to the client, it can also use the hub accepted timestamp.
- If you think there is a better way to get this property, speak up.

Mentions

- Currently we use links to im: URLs inside Markdown or HTML

```
body.contentType = "text/markdown;charset=utf-8";
body.content = "Kudos to [@Alice Smith](im:alice-smith@example.com)"
             + "for making the release happen!”;

body.contentType = "text/html;charset=utf-8";
body.content = "<p>Kudos to <a href='im:alice-smith@example.com'>" +
               "@Alice Smith</a> for making the release happen!</p>"
```

- Improved text in Section 8.4. Use your local policy to decide if display text is a valid/consistent representation of the IM URI. (See Slide 10.)
- Please send text if you don’t like what is there!

Encoding

- What **binary** encoding do we want to use and why?
- What are our requirements?
  - fast to parse and encode
  - not too many ways to encode the same thing
  - extensible
  - stable reference
- Schema-less or Schema required?
  - perhaps schema for required elements; schema-less for extensions

- Options
  - TLS presentation language
    - weird syntax
    - no typedefs
    - no parsers for Javascript and some other languages
  - CBOR
    - emphasis on small rather than fast
    - require schema or not?
    - basically requires CBOR playground to design with
  - Roll our own TLV encoding
    - WG should have a very good reason to do this
  - Protobuf version 2
    - wire format not 100% consistent
    - nesting issues
    - stable reference?
  - msgpack
    - not widely implemented

What's left?

- Do we want a separate Subject? Note: just because a messaging service (RCS, Teams) offers something does not need we need to include it.
- Anything else?

Backup slides

Details of External Content encryption

- External content is encrypted:
  - with an **ephemeral symmetric key and nonce**
  - using an IANA-registered Authenticated Encryption with Additional Data (AEAD) algorithm as described in [RFC5116]
    - MUST implement **AES-128-GCM**
  - The key, nonce, and additional authenticated data (aad) values are set to the values used during the encryption.
  - Unless modified by an extension, `aad` is empty.

Text about Mentions in Section 8.4

An IM URI link to a user who has a member client in the MLS group in which the message was sent is considered a mention. Clients may support special rendering of mentions instead of treating them like any other type of link. In Markdown and HTML, the display text portion of a link is considered a rendering hint from the sender to the receiver of the message. The receiver should use local policy to decide if the hint is an acceptable local representation of the user represented by the link itself. If the hint is not an acceptable representation, the client should instead display its canonical representation for the user.

For example, in the first examples, the sender expresses no preference about how to render the mention. In the second example, the sender requests that the mention is rendered as the literal URI. In the third example, the sender requests the canonical handle for Alice. In the fourth example, the sender requests Alice's first name.¶

```
<im:alice-smith@example.com>
[im:alice-smith@example.com](im:alice-smith@example.com)
[@AliceSmith](im:alice-smith@example.com)
[Alice](im:alice-smith@example.com)
```

Removed: Sharing messageId and timestamp with providers

- Content format has a messageId and timestamp chosen by the encrypting client
  - expose a copy of this in the MLS Additional Authenticated Data field (AAD)
  - local or hub provider can reject a provisional message with a timestamp too far in the past or future (ex: one hour)
  - clients are primarily responsible for detecting duplicate messages IDs among messages they have received
  - local or hub provider can reject a message with a duplicate message ID, but are not required to.
    - UUID + owning provider domain.
      - Q: why include owning provider domain?
      - A: owning or hub provider can check if the domain part purports to be from the wrong domain; owning provider can check if user part duplicates prior messages it has a record for; 100% elimination of duplicate messages is not possible in high availability architecture.
- Concerns about extra metadata?
  - There is not a tremendous amount of data that can be gleaned here without access to some decrypted messages, but the domain name could be valuable information for traffic analysis when a provider has a small amount of volume