XEP-0425: Moderated Message Retraction

JC Brand
mailto:jc@opkode.com
xmpp:jc@opkode.com

2023-03-02
Version 0.3.0

Status	Type	Short Name
Experimental	Standards Track	message-moderation

This specification defines a method for groupchat moderators to retract messages of other users.

Legal

Copyright

This XMPP Extension Protocol is copyright © 1999 – 2024 by the XMPP Standards Foundation (XSF).

Permissions

Permission is hereby granted, free of charge, to any person obtaining a copy of this specification (the ”Specification”), to make use of the Specification without restriction, including without limitation the rights to implement the Specification in a software program, deploy the Specification in a network service, and copy, modify, merge, publish, translate, distribute, sublicense, or sell copies of the Specification, and to permit persons to whom the Specification is furnished to do so, subject to the condition that the foregoing copyright notice and this permission notice shall be included in all copies or substantial portions of the Specification. Unless separate permission is granted, modified works that are redistributed shall not contain misleading information regarding the authors, title, number, or publisher of the Specification, and shall not claim endorsement of the modified works by the authors, any organization or project to which the authors belong, or the XMPP Standards Foundation.

Warranty

NOTE WELL: This Specification is provided on an ”AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE.

Liability

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall the XMPP Standards Foundation or any author of this Specification be liable for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising from, out of, or in connection with the Specification or the implementation, deployment, or other use of the Specification (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if the XMPP Standards Foundation or such author has been advised of the possibility of such damages.

Conformance

This XMPP Extension Protocol has been contributed in full conformance with the XSF’s Intellectual Property Rights Policy (a copy of which can be found at <https://xmpp.org/about/xsf/ipr-policy> or obtained by writing to XMPP Standards Foundation, P.O. Box 787, Parker, CO 80134 USA).

Contents

1 Introduction **1**

2 Discovering support **1**

3 Use Case **2**

3.1 Success case 2

3.2 Error case 3

4 Tombstones **3**

5 Business Rules **5**

6 Security Considerations **5**

7 IANA Considerations **5**

8 XMPP Registrar Considerations **5**

8.1 Protocol Namespaces 5

8.2 Protocol Versioning 5

9 XML Schema **6**

1 Introduction

Occasionally, a Multi-User Chat (XEP-0045) [1] moderator might wish to moderate certain groupchat messages by, for example, retracting them from the groupchat history as part of an effort to address and remedy issues such as message spam, indecent language for the venue or exposing private third-party personal information.
This XEP is designed in such a way to also allow other potential moderator actions, such as message editing, but these actions fall out of the scope of this XEP and we will only be considering message retractions.
Due to the federated nature of XMPP and as with any content moderation tool, the moderation request can **only be considered as a hint** and clients which don't support message moderation are not obligated to enforce any such request.

2 Discovering support

If a groupchat supports moderated message retraction, it MUST specify the 'urn:xmpp:message-moderate:1' feature in its service discovery information features as specified in Service Discovery (XEP-0030) [2] and the Entity Capabilities profile specified in Entity Capabilities (XEP-0115) [3].

Listing 1: Client requests information from a Multi-User Chat (MUC)

```
<iq type='get'
    from='romeo@montague.example/orchard'
    to='room@muc.example.com'
    id='info1'>
  <query xmlns='http://jabber.org/protocol/disco#info'/>
</iq>
```

Listing 2: The MUC advertises support for moderation

```
<iq type='result'
    to='romeo@montague.example/home'
    from='room@muc.example.com'
    id='info1'>
  <query xmlns='http://jabber.org/protocol/disco#info'>
...
    <feature var='urn:xmpp:message-moderate:1'/>
...
  </query>
</iq>
```

[1] XEP-0045: Multi-User Chat <https://xmpp.org/extensions/xep-0045.html>.
[2] XEP-0030: Service Discovery <https://xmpp.org/extensions/xep-0030.html>.
[3] XEP-0115: Entity Capabilities <https://xmpp.org/extensions/xep-0115.html>.

3 Use Case

Consider a situation where a user sends a message that may be deemed inappropriate by a groupchat moderator. The MUC will append a Unique and Stable Stanza IDs (XEP-0359) [4] stanza ID to the message before relaying it to all participants.

Listing 3: The user's client sends a message and the MUC adds a stanza id.

```
<message type="groupchat" from="room@muc.example.com/oldhag" to='
    room@muc.example.com/macbeth' id='inappropriate-1'>
  <body>DM me for free magic potions!</body>
  <stanza-id xmlns='urn:xmpp:sid:0' id='stanza-id-1' by='room@muc.
      example.com'/>
</message>
```

The moderator sees the message and indicates to their client that the message should be retracted. The client then sends an IQ stanza to the MUC service, requesting that the message be retracted.
An optional <reason/> element may be included inside the <moderate/> element, to provide a human-readable explanation why the message was retracted.

Listing 4: The moderator's client sends an IQ stanza to the MUC service

```
<iq type='set' to='room@muc.example.com' id='retract-request-1'>
  <moderate id="stanza-id-1" xmlns='urn:xmpp:message-moderate:1'>
    <retract xmlns='urn:xmpp:message-retract:1'/>
    <reason>This message contains inappropriate content for this forum
        </reason>
  </moderate>
</iq>
```

3.1 Success case

If the moderator is allowed to moderate the message, the MUC will send a message stanza to all participants (including the moderator), indicating that the message has been retracted and by whom.
This message will use Message Retraction (XEP-0424) [5] to indicate that it is a retraction, and will refer to the retracted message by its XEP-0359 Stanza ID. Care must be taken to use the stanza ID that was added by the MUC, i.e. which has a "by" attribute set to the MUC JID, and not some other stanza ID. The client MUST check that the correct stanza ID was used and ignore the stanza if not.
Inside the <retract/> element is a <moderated/> element indicating that the retraction was a moderator action, as well as an optional reason.

[4] XEP-0359: Unique and Stable Stanza IDs <https://xmpp.org/extensions/xep-0359.html>.
[5] XEP-0424: Message Retraction <https://xmpp.org/extensions/xep-0424.html>.

Listing 5: The MUC sends a moderation message to all participants

```
<message type="groupchat" id='retraction-id-1' from='room@muc.example.
    com' to="room@muc.example.com/macbeth">
  <retract id="stanza-id-1" xmlns='urn:xmpp:message-retract:1'>
    <moderated by='room@muc.example.com/macbeth' xmlns='
        urn:xmpp:message-moderate:1'>
      <occupant-id xmlns="urn:xmpp:occupant-id:0" id="
          dd72603deec90a38ba552f7c68cbcc61bca202cd" />
    </moderated>
    <reason>This message contains inappropriate content for this forum
        </reason>
  </retract>
</message>
```

The MUC also responds with an IQ "result" stanza.

Listing 6: The MUC responds to the IQ stanza

```
<iq type='result' to='juliet@montague.example/home' id='retract-
    request-1'></iq>
```

3.2 Error case

In case the message could for some reason not be retracted, the MUC responds with an IQ stanza of type error.

Listing 7: The MUC responds with an error stanza

```
<iq type='error' to='room@muc.example.com/macbeth' id='retract-request
    -1'>
<error type='modify'>
  <forbidden xmlns='urn:ietf:params:xml:ns:xmpp-stanzas'/>
  <text xmlns='urn:ietf:params:xml:ns:xmpp-stanzas'>
    Only moderators are allowed to moderate other participants'␣
        messages
␣␣</text>
</error>
</iq>
```

4 Tombstones

A Message Archive Management (XEP-0313) [6] service MAY replace the contents of a message, that was retracted due to moderation, with a 'tombstone' similar to the one described in

[6] XEP-0313: Message Archive Management <https://xmpp.org/extensions/xep-0313.html>.

Message Retraction (XEP-0424) [7].
The archiving service replaces the message contents (i.e. the <body/> and any related elements which might leak information about the original message) with a <retracted/> element which in turn contains a <moderated/> element.
Note the past tense in the element names, e.g. ”retracted” instead of just ”retract”, to indicate that the retraction has already been executed and that this is therefore a tombstone and not an action that should be applied.
The <moderated/> element SHOULD have a ’by’ attribute specifying the JID of the moderating entity. There might however be cases where it’s preferable to not specify which JID was responsible for the moderation, for example when the moderation happens in an automated manner.
If the message was sent in a semi-anonymous MUC, the occupant id from Anonymous unique occupant identifiers for MUCs (XEP-0421) [8] SHOULD be included for the moderator in the <moderated/> element and for the original (retracted) message author in the <message/> element.

Listing 8: Retracted groupchat message tombstone in a MAM result

```
<message id='aeb213' to='room@muc.example.com/macbeth'>
  <result xmlns='urn:xmpp:mam:2' queryid='f28' id='stanza-id-1'>
    <forwarded xmlns='urn:xmpp:forward:0'>
      <delay xmlns='urn:xmpp:delay' stamp='2019-09-20T23:18:41Z'/>
      <message type="groupchat" from="room@muc.example.com/oldhag" to=
          "room@muc.example.com/macbeth" id="message-id-1">
        <occupant-id xmlns="urn:xmpp:occupant-id:0" id="
            ef73b09de9c90a38ba552f7c68cbeef1dca24429" />
        <retracted stamp='2019-09-20T23:19:12Z' xmlns='
            urn:xmpp:message-retract:1'>
          <moderated by="witch@shakespeare.example" xmlns="
              urn:xmpp:message-moderate:1">
            <occupant-id xmlns="urn:xmpp:occupant-id:0" id="
                dd72603deec90a38ba552f7c68cbcc61bca202cd" />
          </moderated>
          <reason>This message contains inappropriate content for this
              forum</reason>
        </retracted>
      </message>
    </forwarded>
  </result>
</message>
```

[7] XEP-0424: Message Retraction <https://xmpp.org/extensions/xep-0424.html>.
[8] XEP-0421: Anonymous unique occupant identifiers for MUCs <https://xmpp.org/extensions/xep-0421.html>.

5 Business Rules

A moderator MUST NOT send a moderation request for a message with non-messaging payloads. For example, a moderator MUST NOT moderate a roster item exchange request or a file transfer part.
In MUCs, only moderation messages (not tombstones, but messages containing the <moderate/> element) received from the MUC service itself are legitimate, all other such messages MUST be discarded.
The MUC or other service that supports message retraction SHOULD prevent further distribution of a moderated retracted message and the archiving service MAY replace the retracted message with a tombstone.

6 Security Considerations

There can never be a guarantee that a moderated message will appear as such in all clients. Clients should therefore, when possible, inform users that no such guarantee exists.
To prevent message spoofing, it's very important to check that the moderation message comes from the MUC service (as explained in the Business Rules section).

7 IANA Considerations

None.

8 XMPP Registrar Considerations

8.1 Protocol Namespaces

The XMPP Registrar [9] includes 'urn:xmpp:message-moderate:1' in its registry of protocol namespaces (see <https://xmpp.org/registrar/namespaces.html>).

- urn:xmpp:message-moderate:1

8.2 Protocol Versioning

If the protocol defined in this specification undergoes a revision that is not fully backwards-compatible with an older version, the XMPP Registrar shall increment the protocol version

[9] The XMPP Registrar maintains a list of reserved protocol namespaces as well as registries of parameters used in the context of XMPP extension protocols approved by the XMPP Standards Foundation. For further information, see <https://xmpp.org/registrar/>.

number found at the end of the XML namespaces defined herein, as described in Section 4 of XEP-0053.

9 XML Schema

```
<?xml version='1.0' encoding='UTF-8'?>

<xs:schema
    xmlns:xs='http://www.w3.org/2001/XMLSchema'
    targetNamespace='urn:xmpp:message-moderate:1'
    xmlns='urn:xmpp:message-moderate:1'
    elementFormDefault='qualified'>

  <xs:element name='moderate'>
    <xs:attribute name='id' type='xs:string' use='required'/>
  </xs:element>

  <xs:element name='moderated'>
    <xs:complexType>
      <xs:simpleContent>
        <xs:extension base='xs:string'>
          <xs:attribute name='by' type='xs:string'/>
        </xs:extension>
      </xs:simpleContent>
    </xs:complexType>
  </xs:element>

</xs:schema>
```