                    The Atom "deleted-entry" Element
                    draft-snell-atompub-tombstones-18

Abstract

   This specification adds mechanisms to the Atom Syndication Format
   which publishers of Atom Feed and Entry documents can use to
   explicitly identify Atom entries that have been removed.

Status of this Memo

   This Internet-Draft is submitted to IETF 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 http://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 January 21, 2013.

Copyright Notice

Table of Contents

1. Introduction

   Atom [RFC4287] is an XML-based document format that describes lists
   of related information known as "feeds". Feeds are composed of a
   number of items, known as "entries", each with an extensible set of
   attached metadata. The primary use case that Atom addresses is the
   syndication of Web content such as weblogs and news headlines to Web
   sites as well as directly to user agents.

   In the base Atom format, when an entry is removed from a feed, but a
   consumer has already received and processed that entry, perhaps
   adding it to a local cache or display, there is no mechanism for
   determining that the entry has been removed. This specification adds
   a mechanism to the Atom Syndication Format that publishers of Atom
   Feed and Entry documents can use to explicitly identify Atom entries
   that have been removed. Atom consumers can use that information to
   adjust such things as their document cache and user interfaces.

2. Notational 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].

   This specification uses XML Namespaces [W3C.REC-xml-names-19990114]
   to uniquely identify XML element names. It uses the following
   namespace prefix for the indicated namespace URI;

     "at": "http://purl.org/atompub/tombstones/1.0"

3. The at:deleted-entry element

   The at:deleted-entry element represents an Atom Entry that has been
   removed.

```
   deletedEntry =
     element at:deleted-entry {
       atomCommonAttributes,
       attribute ref { atomUri },
       attribute when { atomDateConstruct },
       ( element at:by { atomPersonConstruct }?
       & element at:comment { atomTextConstruct }?
       & element atom:link { atomLink }*
       & element atom:source { atomSource }?
       & anyElement* )
     }
```

The at:deleted-entry element MUST contain a ref attribute whose value
specifies the value of the atom:id of the entry that has been
removed.

The at:deleted-entry element MUST contain a when attribute whose
value is an [RFC3339] "date-time" specifying the instant the entry
was removed. An uppercase "T" character MUST be used to separate
date and time, and an uppercase "Z" character MUST be present in the
absence of a numeric time zone offset

The at:deleted-entry element MAY contain one at:by element used to
identify the entity that removed the entry. The at:by element is an
Atom Person Construct as defined by Section 3.2 of [RFC4287].

The at:deleted-entry element MAY contain one at:comment element whose
value provides additional, language-sensitive information about the
deletion operation. The atom:comment element is an Atom Text
Construct as defined by Section 3.1 of [RFC4287].

The at:deleted-entry element MAY contain any number of atom:link
elements as specified by Section 4.2.7 of [RFC4287].

The at:deleted-entry element MAY contain one atom:source element as
defined by Section 4.2.11 of [RFC4287]. Within the context of an at:
deleted-entry element, the atom:source element is intended to allow
the aggregation of at:deleted-entry element from different feeds
while retaining information about an at:deleted-entry's source feed.
When an at:deleted-entry element appears in a Feed document other
than it's source Feed or when an at:deleted-entry element that has a
source Feed document is used in the context of a Deleted Entry
Document, it MUST contain an atom:source element.

An Atom feed MAY contain any number of at:deleted-entry elements, but
MUST NOT contain more than one with the same combination of ref and
when attribute values.

```
   <feed xmlns="http://www.w3.org/2005/Atom"
         xmlns:at="http://purl.org/atompub/tombstones/1.0">
     ...
     <!-- Minimal deleted-entry -->
     <at:deleted-entry
       ref="tag:example.org,2005:/entries/1"
       when="2005-11-29T12:11:12Z"/>

     <!-- Extended deleted-entry -->
     <at:deleted-entry
       ref="tag:example.org,2005:/entries/2"
       when="2005-11-29T12:11:12Z">
       <at:by>
         <name>John Doe</name>
         <email>jdoe@example.org</email>
       </at:by>
       <at:comment>Removed comment spam</at:comment>
     </at:deleted-entry>
     ...
   </feed>
```

An Atom feed MAY contain atom:entry elements and at:deleted-entry
elements sharing the same atom:id value. In such cases, the
implication is that the particular atom:entry has either been
published to the feed and then subsequently removed; or that a
previously removed entry has been republished to the feed following a
previous deletion. To determine which condition applies, the
processor needs to compare the value of the at:deleted-entry
element's when attribute to the value of the corresponding atom:entry
element's atom:updated value:

o If the when attribute specifies a value equal to or more recent
   than that specified by the atom:updated element, the atom:entry is
   assumed to have been published then subsequently removed.
   Processors SHOULD ignore the older atom:entry element.

o If the when attribute specifies a value older than that specified
   by the atom:updated element, the atom:entry is assumed to have
   been republished to the feed following a prior removal.
   Processors SHOULD ignore the older at:deleted-entry element.

Publishers of feeds containing at:deleted-entry elements should note
that the at:deleted-entry element is advisory in nature only and may
be ignored by Atom processors. The presence of an at:deleted-entry
element does not guarantee that the atom:entry to which it is
referring will no longer be available. For example, if an entry was
published to a feed document that was published and processed
yesterday by an aggregator application, then is subsequently deleted
today, with a corresponding at:deleted-entry element added to the
feed as a signal that the entry was deleted, there is no guarantee

that the aggregator application will pay any attention to the at:
deleted-entry element during subsequent processing operations.

Elements and attributes from any XML vocabulary MAY be used within an
at:deleted-entry element. Processors encountering such markup MUST
NOT stop processing or signal an error. It might be the case that
the Processor is able to process the foreign markup correctly and
does so. When unknown markup is encountered as a child of at:
deleted-entry, Processors MAY bypass the markup and any textual
content and MUST NOT change their behavior as a result of the
markup's presence.

This specification allows the use of IRIs [RFC3987] in precisely the
same manner specified in Section 2 of [RFC4287].

Any element defined by this specification MAY have an xml:base
attribute [W3C.REC-xmlbase-20010627]. When xml:base is used, it
serves the function described in section 5.1.1 of [RFC3986],
establishing the base URI (or IRI) for resolving any relative
references found within the effective scope of the xml:base
attribute.

Any element defined by this specification MAY have an xml:lang
attribute, whose content indicates the natural language for the
element and its descendents. Requirements regarding the content and
interpretation of xml:lang are specified in XML 1.0
[W3C.REC-xml-20040204], Section 2.12.

4. Deleted Entry Document

A "Deleted Entry Document" represents exactly one at:deleted-entry
element outside the context of an Atom feed. It's root is the at:
deleted-entry element.

namespace at = "http://purl.org/atompub/tombstones/1.0"
start = at:deleted-entry

Deleted Entry Documents are specified in terms of the XML Information
Set, serialized as XML 1.0 [W3C.REC-xml-20040204] and identified with
the "application/atomdeleted+xml" media type. Deleted Entry
Documents MUST be well-formed XML. This specification does not
define a DTD for Deleted Entry Documents, and hence does not require
them to be valid (in the sense used by XML).

5. **Digital Signatures**

   The at:deleted-entry element MAY have an Enveloped Signature, as
   described by XML-Signature and Syntax Processing
   [W3C.REC-xmldsig-core-20020212].

   Processors MUST NOT reject an at:deleted-entry containing such a
   signature because they are not capable of verifying it; they MUST
   continue processing and MAY inform the user of their failure to
   validate the signature.

   In other words, the presence of an element with the namespace URI
   "http://www.w3.org/2000/09/xmldsig#" and a local name of "Signature"
   as a child of the document element MUST NOT cause an Processor to
   fail merely because of its presence.

   Section 6.5.1 of [W3C.REC-xmldsig-core-20020212] requires support for
   Canonical XML [W3C.REC-xml-c14n-20010315]. However, many
   implementers do not use it because signed XML documents enclosed in
   other XML documents have their signatures broken. Thus, Processors
   that verify signed at:deleted-entry elements MUST be able to
   canonicalize with the exclusive XML canonicalization method
   identified by the URI "http://www.w3.org/2001/10/xml-exc-c14n#", as
   specified in Exclusive XML Canonicalization
   [W3C.REC-xml-exc-c14n-20020718].

   Intermediaries such as aggregators may need to add an atom:source
   element to an at:deleted-entry that does not contain its own atom:
   source element. If such an entry is signed, the addition will break
   the signature. Thus, a publisher of individually-signed at:deleted-
   entry's should strongly consider adding an atom:source element to
   those elements before signing them. Implementers should also be
   aware of the issues concerning the use of markup in the "xml:"
   namespace as it interacts with canonicalization.

   Section 4.4.2 of [W3C.REC-xmldsig-core-20020212] requires support for
   DSA signatures and recommends support for RSA signatures. However,
   because of the much greater popularity in the market of RSA versus
   DSA, Atom Processors that verify signed Atom Documents MUST be able
   to verify RSA signatures, but do not need be able to verify DSA
   signatures. Due to security issues that can arise if the keying
   material for message authentication code (MAC) authentication is not
   handled properly, Atom Documents SHOULD NOT use MACs for signatures.

6. **Encryption**

   The root of a Deleted Entry Document (the at:deleted-entry element)

MAY be encrypted, using the mechanisms described by XML Encryption
Syntax and Processing [W3C.REC-xmlenc-core-20021210].

Section 5.1 of [W3C.REC-xmlenc-core-20021210] requires support of
TripleDES, AES-128, and AES-256. Processors that decrypt Deleted
Entry Documents MUST be able to decrypt with AES-128 in Cipher Block
Chaining (CBC) mode.

Encryption based on [W3C.REC-xmlenc-core-20021210] does not ensure
integrity of the original document. There are known cryptographic
attacks where someone who cannot decrypt a message can still change
bits in a way where part or all the decrypted message makes sense but
has a different meaning. Thus, Processors that decrypt Deleted Entry
Documents SHOULD check the integrity of the decrypted document by
verifying the hash in the signature (if any) in the document, or by
verifying a hash of the document within the document (if any).

When a Deleted Entry Document is to be both signed and encrypted, it
is generally a good idea to first sign the document, then encrypt the
signed document. This provides integrity to the base document while
encrypting all the information, including the identity of the entity
that signed the document. Note that, if MACs are used for
authentication, the order MUST be that the document is signed and
then encrypted, and not the other way around. Further, if MACs are
used along with a symmetric encryption algorithm, the same key SHOULD
NOT be used in the generation of the MAC and the encryption.

7. Security Considerations

As specified in [RFC4287], Atom processors should be aware of the
potential for spoofing attacks where an attacker publishes atom:entry
or atom:deleted-entry elements using the same atom:id values as
entries from other Atom feeds. An attacker may attempt to trick an
application into believing that a given entry has either been removed
from or added to a feed. To mitigate this issue, Atom processors are
advised to ignore at:deleted-entry elements referencing entries that
have not previously appeared within the containing Feed document and
should take steps to verify the origin of the Atom feed before
considering the entries to be removed.

The at:deleted-entry element can be encrypted and signed using
[W3C.REC-xmlenc-core-20021210] and [W3C.REC-xmldsig-core-20020212],
respectively, and are subject to the security considerations implied
by their use.

Digital signatures provide authentication and message integrity with
proof of origin. Encryption provides data confidentiality.

An application supporting the use of digitally signed atom:entry and
at:deleted-entry elements should be aware of the potential issues
that could arise if a at:deleted-entry element indicating the
deletion of an atom:entry element has been signed using a different
key than what was used to sign the atom:entry, or when an unsigned
at:deleted-entry is used to indicate the deletion of a signed atom:
entry. Either case can potentially indicate a form of spoofing
attack. Processors must take steps to verify the validity of the at:
deleted-entry element.

8. IANA Considerations

A Deleted Entry Document, when serialized as XML 1.0, can be
identified with the following media type:

   Type name: application
   Subtype name: atomdeleted+xml
   Required parameters: None
   Optional parameters: "charset" : This parameter has semantics
   identical to the charset parameter of the "application/xml" media
   type as specified in [RFC3023].
   Encoding considerations: Identical to those of "application/xml"
   as described in [RFC3023], Section 3.2.
   Security considerations: As defined in this specification. In
   addition, as this media type uses the "+xml" convention, it shares
   the same security considerations as described in [RFC3023],
   Section 10.
   Interoperability considerations: There are no known
   interoperability issues.
   Published specification: This specification.
   Applications that use this media type: Undefined. As an extension
   to the Atom Syndication Format ([RFC4287]), this specification may
   be used within any application that uses the Atom Format.
   Additional information:
      Magic number(s): As specified for "application/xml" in
      [RFC3023], Section 3.2
      File extension(s): .atomdeleted
      Macintosh file type code(s): TEXT
   Person & email address to contact for further information: James M
   Snell <jasnell@us.ibm.com>
   Intended usage: COMMON
   Restrictions on usage: None.
   Author: James M Snell <jasnell@us.ibm.com>
   Change controller: IESG

9. Acknowledgements

   The author gratefully acknowledges the feedback from the members of
   the Atom Publishing Format and Protocol working group during the
   development of this specification.

10. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC3023] Murata, M., St. Laurent, S., and D. Kohn, "XML Media
              Types", RFC 3023, January 2001.

   [RFC3339] Klyne, G., Ed. and C. Newman, "Date and Time on the
              Internet: Timestamps", RFC 3339, July 2002.

   [RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform
              Resource Identifier (URI): Generic Syntax", STD 66,
              RFC 3986, January 2005.

   [RFC3987] Duerst, M. and M. Suignard, "Internationalized Resource
              Identifiers (IRIs)", RFC 3987, January 2005.

   [RFC4287] Nottingham, M., Ed. and R. Sayre, Ed., "The Atom
              Syndication Format", RFC 4287, December 2005.

   [W3C.REC-xml-20040204]
              Yergeau, F., Maler, E., Sperberg-McQueen, C., Paoli, J.,
              and T. Bray, "Extensible Markup Language (XML) 1.0 (Third
              Edition)", World Wide Web Consortium FirstEdition REC-xml-
              20040204, February 2004,
              <http://www.w3.org/TR/2004/REC-xml-20040204>.

   [W3C.REC-xml-c14n-20010315]
              Boyer, J., "Canonical XML Version 1.0", World Wide Web
              Consortium Recommendation REC-xml-c14n-20010315,
              March 2001,
              <http://www.w3.org/TR/2001/REC-xml-c14n-20010315>.

   [W3C.REC-xml-exc-c14n-20020718]
              Reagle, J., 3rd, D., and J. Boyer, "Exclusive XML
              Canonicalization Version 1.0", World Wide Web Consortium
              Recommendation REC-xml-exc-c14n-20020718, July 2002,
              <http://www.w3.org/TR/2002/REC-xml-exc-c14n-20020718>.

   [W3C.REC-xml-names-19990114]

            Hollander, D., Bray, T., and A. Layman, "Namespaces in
            XML", World Wide Web Consortium FirstEdition REC-xml-
            names-19990114, January 1999,
            <http://www.w3.org/TR/1999/REC-xml-names-19990114>.

   [W3C.REC-xmlbase-20010627]
            Marsh, J., "XML Base", World Wide Web Consortium
            FirstEdition REC-xmlbase-20010627, June 2001,
            <http://www.w3.org/TR/2001/REC-xmlbase-20010627>.

   [W3C.REC-xmldsig-core-20020212]
            Solo, D., Reagle, J., and D. Eastlake, "XML-Signature
            Syntax and Processing", World Wide Web Consortium
            FirstEdition REC-xmldsig-core-20020212, February 2002,
            <http://www.w3.org/TR/2002/REC-xmldsig-core-20020212>.

   [W3C.REC-xmlenc-core-20021210]
            Eastlake, D. and J. Reagle, "XML Encryption Syntax and
            Processing", World Wide Web Consortium Recommendation REC-
            xmlenc-core-20021210, December 2002,
            <http://www.w3.org/TR/2002/REC-xmlenc-core-20021210>.

Author's Address

   James M Snell

   Email: jasnell@us.ibm.com
   URI: http://ibm.com