SAML V2.0 Subject Identifier Attributes Profile Version 1.0

Working Draft 02

13 September 2017

Technical Committee:
OASIS Security Services (SAML) TC

Chair:
Thomas Hardjono (hardjono@mit.edu), M.I.T.

Editor:
Scott Cantor (cantor.2@osu.edu), Internet2

Additional artifacts:
None

Related work:
This specification is related to:
- eduPerson Object Class Specification (201602)

Declared XML namespaces:
None

Abstract:
This specification standardizes two new SAML Attributes to identify security subjects, as a replacement for long-standing inconsistent practice with the `<saml:NameID>` and `<saml:Attribute>` constructs, and to address recognized deficiencies with the SAML persistent NameID format.

Status:
This Working Draft (WD) has been produced by one or more TC Members; it has not yet been voted on by the TC or approved as a Committee Draft (Committee Specification Draft or a Committee Note Draft). The OASIS document Approval Process begins officially with a TC vote to approve a WD as a Committee Draft. A TC may approve a Working Draft, revise it, and re-approve it any number of times as a Committee Draft.

This Working Draft is being developed under the RF on RAND Terms Mode of the OASIS IPR Policy, the mode chosen when the Technical Committee was established. All members of the TC should be familiar with this document, which may create obligations regarding the disclosure and availability of a member's patent, copyright, trademark and license rights that read on an approved OASIS specification. For information on whether any patents have been disclosed that may be essential to implementing this specification, and any offers of patent licensing terms, please refer to the Intellectual Property Rights section of the TC's web page (https://www.oasis-open.org/committees/security/ipr.php).

Any machine-readable content (Computer Language Definitions) declared Normative for this Work Product must also be provided in separate plain text files. In the event of a discrepancy between such plain text file and display content in the Work Product's prose narrative document(s), the content in the separate plain text file prevails.

URI patterns:
Initial publication URI:
http://docs.oasis-open.org/security/saml-subject-id-attr/v1.0/csd01/saml-subject-id-attr-v1.0-csd01.odt
Permanent "Latest version" URI:
http://docs.oasis-open.org/security/saml-subject-id-attr/v1.0/saml-subject-id-attr-v1.0.odt

Table of Contents

1 Introduction

1.1 IPR Policy

This Working Draft is being developed under the RF on RAND Terms Mode of the OASIS IPR Policy, the mode chosen when the Technical Committee was established.

For information on whether any patents have been disclosed that may be essential to implementing this specification, and any offers of patent licensing terms, please refer to the Intellectual Property Rights section of the TC's web page (https://www.oasis-open.org/committees/security/ipr.php).

1.2 Terminology and Notation

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].

Conventional XML namespace prefixes are used throughout the listings in this specification to stand for their respective namespaces as follows, whether or not a namespace declaration is present in the example:

Prefix	XML Namespace	Comments
`saml:`	urn:oasis:names:tc:SAML:2.0:assertion	This is the SAML V2.0 assertion namespace [SAML2Core].
`samlp:`	urn:oasis:names:tc:SAML:2.0:protocol	This is the SAML V2.0 protocol namespace [SAML2Core].
`md:`	urn:oasis:names:tc:SAML:2.0:metadata	This is the SAML V2.0 metadata namespace [SAML2Meta].
`mdattr:`	urn:oasis:names:tc:SAML:metadata:attributes	This is the SAML V2.0 metadata extension for entity attributes namespace [MetaAttr].
`xsd:`	http://www.w3.org/2001/XMLSchema	This namespace is defined in the W3C XML Schema specification [XMLSCHEMA-2].

1.3 Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997. http://www.ietf.org/rfc/rfc2119.txt.

[SAML2Core] *Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0.* Edited by Scott Cantor, John Kemp, Rob Philpott, Eve Maler. 15 March 2005. OASIS Standard. http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf

[MetaAttr] *SAML V2.0 Metadata Extension for Entity Attributes Version 1.0*. Edited by Scott Cantor. 4 August 2009. OASIS Committee Specification. http://docs.oasis-open.org/security/saml/Post2.0/sstc-metadata-attr-cs-01.pdf. Latest version: http://docs.oasis-open.org/security/saml/Post2.0/sstc-metadata-attr.pdf.

[SAML2Errata] *SAML V2.0 Errata.* Edited by Scott Cantor. 1 May 2012. OASIS Approved Errata. http://docs.oasis-open.org/security/saml/v2.0/errata05/os/saml-v2.0-errata05-os.pdf. Latest version: http://docs.oasis-open.org/security/saml/v2.0/sstc-saml-approved-errata-2.0.pdf

[SAML2Meta] *Metadata for the OASIS Security Assertion Markup Language (SAML) V2.0.* Edited by Scott Cantor, Jahan Moreh, Rob Philpot, Eve Maler. 15 March 2005. OASIS Standard. http://docs.oasis-open.org/security/saml/v2.0/saml-metadata-2.0-os.pdf

[XMLSCHEMA-2] *XML Schema Part 2: Datatypes Second Edition*. Paul V. Biron, A. Malhotra, Editors. W3C Recommendation. October 28, 2004. http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/. Latest version: http://www.w3.org/TR/xmlschema-2/.

1.4 Non-Normative References

[eduPerson] Internet2, “eduPerson Object Class Specification (201602)”, February 2016. http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201602.html.

[RFC4648] Josefson, S., “The Base16, Base32, and Base64 Data Encodings”, RFC 4648, October 2006. http://www.ietf.org/rfc/rfc4648.txt.

[SAML2Prof] *Profiles for the OASIS Security Assertion Markup Language (SAML) V2.0.* Edited by John Hughes, Scott Cantor, Jeff Hodges, Frederick Hirsch, Prateek Mishra, Rob Philpot, Eve Maler. 15 March 2005. OASIS Standard. http://docs.oasis-open.org/security/saml/v2.0/saml-profiles-2.0-os.pdf

2 Motivation

2.1 Problem Statement

Identification of subjects in security protocols and applications has a fraught history of inconsistent syntax, bugs, terrible but deeply cemented practices such as misuse of email addresses, vertical market-specific approaches, and failure to precisely communicate intended semantics and constraints. These problems lead to overly complex burdens on both asserting and relying parties to supply and consume a variety of different identifiers in different formats, many of which work poorly with off the shelf applications. Much of this is self-inflicted fragmentation due to the constant tension between fixing problems with new solutions and avoiding them to gain scale.

SAML itself has its origins in a design philosophy that tried to avoid breaking new ground in this area, and instead attempted to design for generality, which is valuable, but did not ease adoption due to a lack of guidance. SAML also complicates itself by providing an optional, singly-appearing construct for identification (the `<saml:NameID>` element) *and* a more general multiply-appearing `<saml:Attribute>` construct that inherently overlap.

This, together with inconsistent technical precision by implementers and deployers, creates complexity. Deployment experience has shown that use of the NameID feature is confusing in many implementations. It also, through its presence in the SAML Single Logout protocol, potentially appears in web access logs, leading to the added complexity of encryption when privacy is a consideration.

There is a general consensus by most federated identity practitioners around a few common requirements:

- Identifiers should be as stable as possible and should never have a risk of reassignment to different subjects due to the lack of tight synchronization[1] inherent between loosely-coupled systems.
- Opaque (i.e., superficially random) identifiers are inherently more stable than name-based identifiers or email addresses in many organizations.
- Identifiers should be compact and simple to handle and manipulate.
- The ability to clearly express the scope of an identifier's uniqueness and enforce policy around the issuers permitted to supply an identifier is crucial to federated systems and the lack of such policy has led to widely-publicized breaches.

Another requirement perhaps more common to education and research is the ability for different asserting parties to issue the same identifier. This is facilitated by ensuring the scope of an identifier is part of its value and not implicit in a protocol-specific value specific to an asserting party.

SAML does not define an identifier that meets all of these requirements well. It does standardize a kind of NameID termed "persistent" that meets some of them in the particular case of so-called "pairwise" identification, where an identifier varies by relying party. It has seen minimal adoption outside of a few contexts, and fails at the "compact" and "simple to handle" criteria above, on top of the disadvantages inherent with all NameID usage.

Pairwise identification helps meet certain privacy and regulatory requirements, but does not address many common use cases that demand cross-system correlation without the friction of complex linking protocols and the involvement of the data subject.

In addition, it has come to light that many, if not most, applications have a predisposition to handle identifiers case-insensitively, partly due to a long-standing, though factually untrue, assumption that e-mail address mailbox names are case-insensitive data. SAML's "persistent" NameID definition explicitly requires case-sensitive handling, making them impossible to use safely with such applications without resorting to additional layers of profiling. Note that any other specification promulgating such identifiers is potentially unsafe in combination with such applications and should be used with caution.

1 It's worth noting that SAML actually defines a protocol for managing changes to NameID values, but it has seen very little adoption, further demonstrating the lack of value of NameID usage.

For all these reasons, this profile attacks these problems using a clean-slate approach that abandons existing practice instead of attempting to layer more profiling and out of band agreements on top of existing solutions, an approach that has seemingly reached its breaking point.

2.2 Relationship to Existing Work

Clean slate notwithstanding, this profile is based on a thorough review of practice within the higher education sector, which has seen extensive adoption of SAML and partially-successful efforts to standardize subject identification and avoid the “email address” trap that most of the technical world fell into many years ago.

Among the significant work in this space, the [eduPerson] schema includes a number of identifier attributes, some widely adopted and some less so. This profile is particularly influenced by:

- Experience with the SAML “persistent” NameID construct and the eduPersonTargetedID attribute.
- The eduPersonPrincipalName and eduPersonUniqueId attributes, the former successful but deeply flawed, the latter less successful but more consciously defined.
- Success with DNS domain-based scoping of values and managing policy around their use in SAML.
- Challenges in the adoption of profiles required to accommodate the limitations of widely deployed identifiers.

Portions of this specification are borrowed liberally from the [eduPerson] specification in a deliberate desire to remain consistent with the formulation of the eduPersonUniqueId attribute.

3 SAML V2.0 Subject Identifier Attributes Profile Version 1.0

3.1 Required Information

Identification: urn:oasis:names:tc:SAML:profile:subject-id

Contact information: security-services-comment@lists.oasis-open.org

Description: Given below.

Updates: None.

3.2 Overview

This profile defines a pair of SAML Attributes providing for unique identification of security subjects (generally but not exclusively people). One is designed for general use as a correlatable identifier, and the other is a pairwise identifier suitable for more specialized use.

Both Attributes are limited to a single value when expressed in SAML assertions and other constructs. They may be mapped to and form other technical forms (e.g., LDAP) but this profile does not include such mappings.

In the terminology used in this profile:

- "asserting party" refers to a SAML entity, uniquely identified by an entityID, that issues assertions containing one or both of these Attributes
- "relying party" refers to one or more SAML entities, each uniquely identified by an entityID, that receive assertions containing one or both of these Attributes

In addition, this profile defines a signaling mechanism for a Service Provider to express its subject identification requirements via SAML metadata [SAML2Meta], by means of the `<mdattr:EntityAttributes>` extension [MetaAttr]. This allows Identity Providers to unambiguously understand the requirements of the service and facilitates deployment profiles that wish to mandate support for one or both of these Attributes, while maintaining appropriate privacy expectations.

3.3 Standard Subject Identifier

For standard identification of subjects, the following SAML Attribute is defined:

Name: `urn:oasis:names:tc:SAML:attribute:subject-id`

NameFormat: `urn:oasis:names:tc:SAML:2.0:attrname-format:uri`

This is a long-lived, non-re-assignable, omni-directional identifier suitable as a globally-unique external key by applications.

3.3.1 Syntax and Handling

This Attribute, when appearing as a SAML `<saml:Attribute>` element, MUST contain exactly one `<saml:AttributeValue>` element, whose `xsi:type` SHOULD be absent or if present MUST BE bound to the XML Schema `xsd:string` data type [XMLSCHEMA-2].

Any leading or trailing whitespace present in the `<saml:AttributeValue>` element's content is not significant and MUST be stripped by the relying party prior to evaluation or comparison.

The value consists of two substrings (termed a “unique ID” and a “scope” in the remainder of this definition) separated by an `@` symbol (ASCII 64) as an inline delimiter.

The unique ID consists of from 1 to 127 characters, all either alphanumeric or the equals sign (ASCII 61). The first character MUST be alphanumeric.

The scope consists of 1 to 127 alphanumeric, hyphen (ASCII 45), or period (ASCII 46) characters. The first character MUST be alphanumeric. The scope deliberately resembles, and typically may be, a DNS domain name, but is drawn from a more limited character set due to case folding considerations, and no attempt is made to limit the allowable grammar to legal domain names (e.g., it allows consecutive periods).

The ABNF grammar is therefore:

```
<value> = <uniqueID> "@" <scope>
<uniqueID> = (ALPHA / DIGIT) 0*126(ALPHA / DIGIT / "=")
<scope> = (ALPHA / DIGIT) 0*126(ALPHA / DIGIT / "-" / ".")
```

Value comparison MUST be performed case-insensitively (that is, values that differ only by case are the same, and refer to the same subject). It is RECOMMENDED that alphabetic characters be in lower-case when expressing and storing values.

3.3.2 Semantics and Practices

A value (the unique ID and scope together) MUST be bound to only one subject, but the same unique ID given a different scope may refer to the same or (far more likely) a different subject.

The relationship between an asserting party and a scope is an arbitrary one and does not reflect any assumed relationship between a scope in the form of a domain name and a domain found in a given SAML entityID.

A value MUST NOT be assigned to more than a single subject over its lifetime of use under any circumstances. The unique ID should therefore be constructed in a fashion that reduces the probability of non-technical or political considerations leading to a violation of this requirement, and any such violation should be treated as a potential security risk to the relying parties to which the value may have been given.

Relying parties should not treat this identifier as an email address for the subject as it is unlikely (though not precluded) for it to be valid for that purpose. Most organizations will find that existing email address values will not serve well as values for this Attribute.

The unique ID should not change as a result of a change to any other data associated with the subject (e.g., name, email address, age, organizational role).

A given value MUST identify the same subject regardless of the context of use and for which relying parties to which the Attribute is given. It is therefore to be assumed by relying parties that receive a given value that the same subject has been identified.

Note that, policy permitting, a given value could be provided by any asserting party, and the requirement still holds: identical values correspond to the same subject. While it will be common in many deployments to limit values with a given scope to a single asserting party, this is ultimately left to the discretion of the relying party and the use case.

Inevitably, a single subject may be identified simultaneously by multiple values, but this should be minimized to the extent possible.

3.4 Pairwise Subject Identifier

For pairwise identification of subjects, the following SAML Attribute is defined:

Name: `urn:oasis:names:tc:SAML:attribute:pairwise-id`

NameFormat: `urn:oasis:names:tc:SAML:2.0:attrname-format:uri`

This is a long-lived, non-re-assignable, uni-directional identifier suitable as a unique external key specific to particular applications. Its value for a given subject depends on the relying party to whom it is given, preventing unrelated systems from using it as a basis for correlation.

3.4.1 Syntax and Handling

The requirements for this Attribute are identical to those described in Section 3.3.1. That is, values of this Attribute are indistinguishable, lacking context, from the other.

3.4.2 Semantics and Practices

Given a particular relying party, a value (the unique ID and scope together) MUST be bound to only one subject, but the same unique ID given a different scope may refer to the same or (far more likely) a different subject. The same value provided to different relying parties MAY refer to different subjects, and indeed that is the primary distinguishing characteristic of this identifier Attribute.

The relationship between an asserting party and a scope is an arbitrary one and does not reflect any assumed relationship between a scope in the form of a domain name and a domain found in a given SAML entityID.

A value MUST NOT be assigned to more than a single subject over its lifetime of use under any circumstances. The unique ID should therefore be constructed in a fashion that reduces the probability of non-technical or political considerations leading to a violation of this requirement, and any such violation should be treated as a potential security risk to the relying parties to which the value may have been given.

The value MUST NOT be reversible by a relying party into a non-pairwise identifier for the subject through ordinary effort.

Relying parties should not treat this identifier as an email address for the subject as it is unlikely (though not precluded) for it to be valid for that purpose. Most organizations will find that existing email address values will not serve well as values for this Attribute.

The unique ID should not change as a result of a change to any other data associated with the subject (e.g., name, email address, age, organizational role).

Assuming a particular scope, a given subject MUST be identified with a different, though consistent, unique ID for each relying party to which a value is provided; however, the relationship between relying parties and SAML entities is not defined by this profile and is interpreted from the perspective of the asserting party. While it would be typical for an Identity Provider to base its notion of a relying party boundary on a single Service Provider's entityID, that is not specifically required by this profile. The boundary MAY be larger or even smaller, at the Identity Provider's discretion or as addressed by additional profiles.

While it will be common in many deployments to limit values with a given scope to a single asserting party, this is ultimately left to the discretion of the relying party and the use case. It is unspecified by this profile whether a given value provided by two or more asserting parties correspond to the same subject. This would depend on out of band arrangements made between the parties. But, in such cases, the "standard" subject identifier defined in Section 3.3 is likely to be a much better choice.

3.4.3 Strategies

Supporting pairwise identifiers typically involves either the generation and storage of random values, or the computation of reproducible values that can be produced on demand but need not be stored. This profile does not require any specific approach, but implementers should be aware that some techniques for computing values may result in an unacceptable risk of case conflicts. For example, a salted hash over a seed identifier together with a relying party identifier produces a "safe" generated value, but becomes unsafe when encoded in Base64 [RFC4648] (and the allowable character set is defined in part to preclude this choice). However, encoding hashes in Base32 [RFC4648] is a safe choice, and the equals sign is included in the allowable character set to accomodate this.

3.4.4 Differences from "persistent" NameIDs

This Attribute is a direct replacement for the `urn:oasis:names:tc:SAML:2.0:nameid-format:persistent` NameID Format defined in SAML [SAML2Core]. There are obvious syntactic differences, in a deliberate attempt at simplification. The XML syntax and data "triple" are replaced with a simpler id/scope pair encoded into a string, and the awkward use of a URI to qualify the value is replaced with a simpler, shorter, and more flexible approach that more easily emulates the email address syntax required by many applications, and decouples identifier scoping from SAML entity naming.

One functional gap is the interoperable mechanism of SAML "affiliations" to group entities for the purpose of targeting pairwise identifiers to multiple Service Providers, which was baked into the SAML protocol. It has been left out of this profile due to the general lack of adoption by implementers or deployers in the

intervening years since the publication of the standard. Were there demand, it could be incorporated into a future revision of this work.

3.5 Considerations for SAML Profiles

The Attributes defined in this profile are designed to be used in conjunction with any SAML profiles that support the use of SAML Attributes, though its predominant expected use is with the various SAML authentication profiles [SAML2Prof] such as the Browser SSO and Enhanced Client and Proxy profiles.

3.5.1 Requirements Signaling

In the event that SAML metadata [SAML2Meta] is used, a relying party MUST express its identifier requirements by including an `<mdattr:EntityAttribute>` extension [MetaAttr] in its metadata containing the following Attribute:

Name: `urn:oasis:names:tc:SAML:profile:subject-id`

NameFormat: `urn:oasis:names:tc:SAML:2.0:attrname-format:uri`

This Attribute, MUST contain exactly one `<saml:AttributeValue>` element, whose `xsi:type` SHOULD be absent or if present MUST BE bound to the XML Schema `xsd:string` data type [XMLSCHEMA-2].

The value MUST be one of the following, signaling the corresponding requirement:

- `subject-id`
  - The relying party requires the standard identifier Attribute defined in Section 3.3.
- `pairwise-id`
  - The relying party requires the pair-wise identifier Attribute defined in Section 3.4.
- `none`
  - The relying party does not require any subject identifier and is designed to operate without a specific user identity (e.g., with authorization based on non-identifying data).
- `any`
  - The relying party will accept any of the identifier Attributes defined in this profile but requires at least one.

This profile does not define specific normative behavior on the part of asserting parties in response to this metadata, but it is expected that other profiles will do so in the future.

3.5.2 NameID Considerations

While the Attributes defined in this profile have as a goal the explicit replacement of the `<saml:NameID>` element as a means of subject identification, it is certainly possible to compose them with existing NameID usage provided the same subject is being identified. This can also serve as a migration strategy for existing applications.

In addition, some profiles such as the Single Logout Profile [SAML2Prof] require the use of a `<saml:NameID>` element, which implies the earlier use of a NameID. In such cases, it is RECOMMENDED that the `urn:oasis:names:tc:SAML:2.0:nameid-format:transient` NameID Format be used.

4 Conformance

4.1 Conformance Clause 1: Asserting Party Implementations

An asserting party implementation conforms to this specification if it can be configured to produce the two identifier Attributes conforming to the normative requirements in Sections 3.3 and 3.4.

4.2 Conformance Clause 2: Relying Party Implementations

A relying party implementation conforms to this specification if it can be configured to consume neither, either, and both of the two identifier Attributes conforming to the normative requirements in Sections 3.3 and 3.4.

If the relying party implementation provides a mechanism for generation and/or publication of SAML metadata [SAML2Meta], then it MUST support the inclusion of the extension defined in Section 3.5.1.

Appendix A Acknowledgments

The following individuals have participated in the creation of this specification and are gratefully acknowledged:

Contributors to the InCommon Deployment Profile Working Group

Appendix B Revision History

Revision	Date	Editor	Changes Made
WD 01	30 Aug 2017	Scott Cantor	Initial draft
WD 02	13 Sep 2017	Scott Cantor	Added considerations for other profiles