ETSI GS CIM 019 V1.1.1 (2022-08)

cross-cutting Context Information Management (CIM); handling of provenance information in NGSI-LD

Disclaimer

The present document has been produced and approved by the cross-cutting Context Information Management (CIM) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
DGS/CIM-0019

Keywords
data, digital signature, provenance,
reliability, trust

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Content

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Group Specification (GS) has been produced by ETSI Industry Specification Group (ISG) cross-cutting Context Information Management (CIM).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Executive summary

The present document specifies a mechanism for embedding W3C® Data Integrity digital signatures into NGSI-LD Entities.

Introduction

In the most generic scenario of a NGSI-LD [i.2] ecosystem, Entities from Context Providers are sent, through multiple Context Brokers, to Clients. In this scenario, the context information creator is the Context Provider, which is trusted by the Clients.

When an Entity typically contains multiple Attributes, it is important to guarantee that these values will not be altered through all its cycles, so that a Client, without further contact with the Context Provider, can be sure of the integrity.

The preferred solution in both literature and industry, to the data integrity problem, is the implementation of a digital signature system.

1 Scope

The present document designs a solution to verify integrity and to precisely evaluate attribution and authenticity of NGSI-LD [i.2] Context Information, throughout its lifecycle. It defines technical means for enabling a chain of trust from Context Providers to Context Consumers, by embedding verifiable credentials into NGSI-LD documents, leveraging the W3C® Data Integrity methodology for digital signatures.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] W3C® Draft Community Group Report 13 April 2021: "RDF Dataset Canonicalization".

NOTE: Available at https://json-ld.github.io/rdf-dataset-canonicalization/spec.

[2] W3C® Final Community Group Report 22 July 2022: "Data Integrity 1.0".

NOTE: Available at https://www.w3.org/community/reports/credentials/CG-FINAL-data-integrity-20220722.

[3] IETF RFC 8785: "JSON Canonicalization Scheme (JCS)".

NOTE: Available at https://datatracker.ietf.org/doc/html/rfc8785.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI GR CIM 007 (V1.1.1): "Context Information Management (CIM); Security and Privacy".

NOTE: Available at https://www.etsi.org/deliver/etsi_gr/CIM/001_099/007/01.01.01_60/gr_CIM007v010101p.pdf.

[i.2] ETSI GS CIM 009 (V1.5.1): "Context Information Management (CIM); NGSI-LD API".

NOTE: Available at https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.05.01_60/gs_CIM009v010501p.pdf.

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

Atomic Entity: digitally signed NGSI-LD Entity with only one Attribute

Client: shorthand for NGSI-LD Context Consumer

Context Provider: NGSI-LD Context Source or NGSI-LD Context Producer

Derivation Process: process that transforms NGSI-LD Attributes into Sealed Attributes

NGSI-LD Attribute: reference to both an NGSI-LD Property and to an NGSI-LD Relationship

NGSI-LD Context Broker: architectural component that implements all the NGSI-LD interfaces

NGSI-LD Context Consumer: agent that uses the query and subscription functionality of NGSI-LD to retrieve context information

NGSI-LD Context Producer: agent that uses the NGSI-LD context provision and/or registration functionality to provide or announce the availability of its context information to an NGSI-LD Context Broker

NGSI-LD Context Source: source of context information which implements the NGSI-LD consumption and subscription (and possibly provision) interfaces defined by the present document

NGSI-LD Entity: informational representative of something that is supposed to exist in the real world, physically or conceptually

NGSI-LD Property: description instance which associates a main characteristic, i.e. an **NGSI-LD Value**, to either an NGSI-LD Entity, an NGSI-LD Relationship or another NGSI-LD Property and that uses the special *hasValue* property to define its target value

Reconstruction Process: opposite process of the Derivation Process

Sealed Attribute: NGSI-LD Attribute with "ngsildproof" sub-property

verification method: method that can be used together with a process to independently verify a proof

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

JCS	JSON Canonicalization Scheme
JSON	JavaScript Object Notation
JSON-LD	JSON Linked Data
JWS	JSON Web Signature
JWS/CT	JSON Web Signature Clear Text
LD	Linked Data
NGSI	Next Generation Service Interfaces
NGSI-LD	NGSI Linked Data
RDF	Resource Description Format

4 Requirements

ETSI's Industry Specification Group on cross-cutting Context Information Management (ISG CIM) is tasked with specifying security structures for, among other things, ensuring data integrity and provenance of NGSI-LD Entities throughout the typical workflow from data sources, which are the creators of the context information/data, to a federation of Context Providers and NGSI-LD Context Brokers, to NGSI-LD Context Consumers.

Critical requirements for the integrity of data flowing within the NGSI-LD ecosystem, are (adapted from [i.1]):

- INT-1: The NGSI-LD Context Consumers should be able to determine that data integrity has been preserved.

This requirement dictates that data integrity for NGSI-LD Entities is recommended, but not mandatory, when assembling NGSI-LD Entities.

- INT-2: Verification of integrity shall be independent of syntactical re-ordering that may occur when serializing NGSI-LD Entities between peers.

- INT-3: Verification of integrity should be independent of the NGSI-LD serialization format itself, i.e. serialization formats should not strip verification information.

Information for verification of integrity is transported within NGSI-LD Entities, when they are serialized, as specified in clause 5. This requirement acknowledges that some of the output formats supported in NGSI-LD (e.g. the simplified representation, see clause 4.5.4 of [i.2]) may strip information that is vital to verification of integrity.

- INT-4: Preservation of data integrity shall not rely on the Clients trusting the relaying intermediate Context Providers or NGSI-LD Context Brokers, but solely the creators.

5 Specification

5.1 Fulfilling requirements

5.1.0 Foreword

For the sake of brevity and clarity, the terms Entity, Attribute, Property and Context Broker (or simply Broker, all of them capitalized) are used interchangeably with NGSI-LD Entity, NGSI-LD Attribute, NGSI-LD Property and NGSI-LD Context Broker, respectively.

The scenario used throughout the present document is the generation of Entities from Context Providers that are then sent, through multiple Context Brokers, to Clients. In this scenario, without loss of generality, the context information creator is the Context Provider, which is thusly trusted by the Clients.

In this scenario, where an Entity typically contains multiple Attributes, it is important to guarantee that these values will not be altered through all its cycles, so that a Client, without further contact with the Context Provider, can be sure of the integrity.

The preferred solution in both literature and industry, to the data integrity problem, is the implementation of a digital signature system.

A digest file of the Entity, cryptographically encoded with the signer private key, bound with it, guarantees the non-corruption of data (integrity) and the association to a specific private key (provenance).

Thus, using a digital signature system fulfils requirements INT-1 and INT-4 described in clause 4.

But cryptographic operations like hashing and signing depend on the fact that the target data does not change during serialization, transport, or parsing.

In the NGSI-LD ecosystem, every time a Context Broker receives Entities, it stores them in terms of the underlying Property Graph structure. On request, the Broker will serialize the Entity, generating its JSON-LD structure anew, in order to share it or send it to Clients. The new structure, though semantically equivalent, can be very different in terms of formatting and ordering of the underlying JSON key+value pairs.

The solution is the implementation of a canonicalization algorithm. Canonicalization is the process of transforming an input dataset to a normalized dataset. Any two input datasets that contain the same information, regardless of their arrangement, will be transformed into identical normalized dataset. This process is sometimes also called normalization.

ISG CIM is thus seeking to apply JSON canonicalization algorithms to serialized JSON-LD data, prior to digitally signing it, in order to fulfil requirement INT-2 described in clause 4.

Table 5.1.0-1 summarizes the status of various JSON canonicalization algorithms.

Table 5.1.0-1: Canonicalization algorithms

Specification Name	Group	Specification Status	Comments
JCS: JSON Canonicalization Scheme	IETF RFC 8785 [3]	Not an Internet Standards Track specification	- Builds on the strict serialization methods for JSON primitives defined by ECMAScript (https://en.wikipedia.org/wiki/ECMAScript), constraining JSON data to the Internet JSON (I-JSON) subset, and by using deterministic property sorting. - Good fit for JSON format. - Array elements are not managed by the algorithm, thus rearranging them will invalidate any digital signature on them. - Simple. Possible implementation with JWS standard, through JWS/CT specification (not yet a published standard).
RDF Dataset Canonicalization	W3C® Credentials Community Group; W3C® RDF Dataset Canonicalization and Hash Working Group	It is not a W3C® Standard nor is it on the W3C® Standards Track	- An algorithm for normalizing RDF datasets such that comparing the differences between sets of graphs, digitally sign them, or generate short identifiers for graphs via hashing algorithms is possible. - Good fit for JSON-LD format. - Array elements can be reordered without invalidating signature. - More complex. - Supported by W3C® Data Integrity specification [2] (not yet Standard) and W3C® Verifiable Credentials standard.

The RDF Dataset Canonicalization [1] is based on Resource Description Framework (RDF), an abstract model with several serialization formats.

The implementation of the RDF Dataset Canonicalization inside the NGSI-LD ecosystem fulfils the INT-2 requirement described in clause 4.

5.1.1 Overview of W3C® Data Integrity specification

The W3C® Data integrity specification [2] describes mechanisms for ensuring the authenticity and integrity of structured digital documents using cryptography.

In order to produce a verifiable digital proof, it supports the usage of different canonicalization algorithms, so that both detection of tampering with the integrity of data and, at the same time, re-ordering or the structured document, is possible.

Following the W3C® Data Integrity specification [2], it is possible to create a data integrity "`proof`" element, which is a set of attributes that represent a digital proof and all parameters required to verify it.

A data integrity proof contains, at least, the following attributes:

- `type:` which indicates the specific type of digital signature used. It is defined as "a specified set of cryptographic primitives bundled together into a cryptographic suite for the purposes of safety and convenience, by cryptographers for developers. A proof type typically consists of a canonicalization algorithm, a message digest algorithm, and a specific corresponding proof algorithm";
- `proofPurpose`: a parameter that ensures that the digital proof is used for the reason it was created for;
- `verificationMethod:` a set of parameters required to independently verify the proof;
- `created`: date and time of the proof generation;
- `proofValue`: the value of the encoded hash.

The verification process is possible through the access to a so-called controller document, a set of data that specify the relationship between a controller, the entity who can change the controller document, and other data sets such as a public cryptographic key.

Whoever wants to verify the data integrity proof shall ensure that a verification method is bound to a specific controller, by going from the verification method attribute in the proof to the controller document, ensuring that this also contains the same verification method and the same proof purpose.

The following signature suites (i.e. verification methods and digital signature types) are contemplated in W3C® Data integrity specification [2]: eddsa-2022, nist-ecdsa-2022, koblitz-ecdsa-2022, rsa-2022, pgp-2022, bbs-2022, eascdsa-2022, ibsa-2022, and jws-2022.

Both JSON Canonicalization Scheme and RDF Dataset Canonicalization are supported by the W3C Data Integrity specification.

5.2 Data integrity and provenance for NGSI-LD

5.2.0 Foreword

Adoption of a W3C® Data Integrity signature mechanism that is based on an RDF Dataset Canonicalization (for example the Ed25519Signature2020 proof type, which produces a verifiable digital proof by canonicalizing the input data using the RDF Dataset Canonicalization algorithm and then digitally signing it using an Ed25519 elliptic curve signature), fulfils requirements INT-1, INT-2 and INT-4, thus guaranteeing data integrity and provenance through the whole NGSI-LD Entity lifecycle.

In order to fulfil the INT-3 requirement, i.e. in order to specify how to serialize and embed the W3C® verifiable digital proof into the NGSI-LD Entity, the following need to be defined and detailed:

- Atomic Entity.
- Sealed Attribute.
- Derivation process.
- Reconstruction Process.