Workgroup: Network Working Group
Internet-Draft:
draft-polli-id-digest-algorithms-02
Published: 2 December 2021
Intended Status: Experimental
Expires: 5 June 2022
Authors: R. Polli
Digital Transformation Department, Italian Government

The "id-" prefix for Digest Algorithms

Abstract

This document defines the "id-" prefix for digest-algorithms used in the Digest Fields. This prefix explicits that the computed checksum value is independent from Content-Encoding.

Note to Readers

RFC EDITOR: please remove this section before publication

Discussion of this draft takes place on the HTTP working group mailing list (ietf-http-wg@w3.org), which is archived at https://lists.w3.org/Archives/Public/ietf-http-wg/.

The source code and issues list for this draft can be found at https://github.com/ioggstream/draft-polli-id-digest-algorithms.

Status of This Memo

This Internet-Draft is submitted 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 https://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 5 June 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction

The [DIGEST] defines a way to convey a checksum of a representation-data as specified in [SEMANTICS].

As the representation data depends on the value of Content-Encoding, it is useful to convey the checksum value of a representation without any content coding applied.

This proposal introduces the id- prefix to specify that the provided digest-algorithm value is computed on the representation-data without any content coding applied.

1.1. Notational Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here. These words may also appear in this document in lower case as plain English words, absent their normative meanings.

This document uses the Augmented BNF defined in [RFC5234] and updated by [RFC7405].

The definitions "representation", "selected representation", "representation data", "representation metadata", and "content" in this document are to be interpreted as described in [SEMANTICS].

The definitions "digest-algorithm" and "representation-data-digest" in this document are to be interpreted as described in [DIGEST].

2. The "id-" prefix for digest-algorithms

A new digest-algorithm to be registered within the HTTP Digest Algorithm Values Registry MUST NOT start with the string id-.

The following two examples show two digest-algorithm names that cannot be registered

id-sha-256
id-sha-512

For every digest-algorithm registered in the HTTP Digest Algorithm Values the associate id- digest-algorithm has the following properties:

* the checksum is computed on the representation-data of the resource when no content coding is applied;

* the checksum is computed according to the original digest-algorithm "Description" field, and uses the same encoding of the original digest-algorithm.

3. Security Considerations

3.1. Disclosure of encrypted content

If the content coding provides encryption features, sending the checksum of unencoded representation can disclose information about the original content.

4. IANA Considerations

Please, add the following text to the "Note" section of the HTTP Digest Algorithm Values.

" For each registered Digest Algorithm, an associated id- algorithm is defined.

The associated representation-data-digest is computed according to Section 2 of this document. "

5. Normative References

[DIGEST] Polli, R. and L. Pardue, "Digest Fields", Work in Progress, Internet-Draft, draft-ietf-httpbis-digest-headers-07, 16 November 2021, <https://www.ietf.org/archive/id/draft-ietf-httpbis-digest-headers-07.txt>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, DOI 10.17487/RFC5234, January 2008, <https://www.rfc-editor.org/info/rfc5234>.

[RFC7405] Kyzivat, P., "Case-Sensitive String Support in ABNF", RFC 7405, DOI 10.17487/RFC7405, December 2014, <https://www.rfc-editor.org/info/rfc7405>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[SEMANTICS] Fielding, R. T., Nottingham, M., and J. Reschke, "HTTP Semantics", Work in Progress, Internet-Draft, draft-ietf-httpbis-semantics-19, 12 September 2021, <https://www.ietf.org/archive/id/draft-ietf-httpbis-semantics-19.txt>.

Appendix A. Examples

A.1. The id-sha-256 digest-algorithm

The following request conveys a brotli encoded json object

```
{"hello": "world"}
```

The Digest computed using the "sha-256" digest-algorithm present in HTTP Digest Algorithm Values is content coding aware, while its associated "id-" digest-algorithm is not.

```
POST /data HTTP/1.1
Content-Type: application/json
Content-Encoding: br
Digest: sha-256=4REjxQ4yrqUVicfSKYNO/cF9zNj5ANbzgDZt3/h3Qxo=,
        id-sha-256=X48E9qOokqqrvdts8nOJRJN3OWDUoyWxBf7kbu9DBPE=

CwGAZiwiAeyJoZWxsbyI6ICJ3b3JsZCJ9Aw==
```

FAQ

RFC Editor: Please remove this section before publication.

Q: Why to convey the checksum independent from the content codings? This is useful to identify and validate a resource downloaded from different sources using different content codings, or to compare a resource with its stored or signed counterpart.

Q: How does it improve the life of checksum providers? If providers use reverse proxies to eg. compress responses, this could invalidate content coding aware checksums. Providing an id-algorithm, allows the digest checksum to be validated.

Code Samples

RFC Editor: Please remove this section before publication.

How can I generate an identity digest value?

The following python3 code can be used to generate digests for json objects using crc32c algorithm. Note that these are formatted as base64. This function could be adapted to other algorithms and should take into account their specific formatting rules.

```
import base64, json, brotli, hashlib

identity = lambda x: x

def digest(item, content_coding=identity, algorithm=hashlib.sha256) -> b
    json_bytes = json.dumps(item).encode()
    content_encoded = content_coding(json_bytes)
    checksum = algorithm(content_encoded)
    return base64.encodebytes(checksum.digest())


item = {"hello": "world"}

print("sha-256 digest value for a br-coded representation: ",
    digest(item, content_coding=brotli.compress)
)

print("id-sha-256 digest value for a br-coded representation: ",
    digest(item, content_coding=identity)
)
```

Acknowledgements

This specification was born from a thread created by James Manger and the subsequent discussion here https://github.com/httpwg/http-extensions/issues/885.

Change Log

RFC Editor: Please remove this section before publication.

Since draft-polli-id-digest-algorithms-01

*Include id-sha-256 and id-sha-512.

Author's Address

Roberto Polli
Digital Transformation Department, Italian Government
Italy

Email: robipolli@gmail.com