

cbor
Internet-Draft
Intended status: Standards Track
Expires: 4 September 2025

C. Ams端 ss
3 March 2025

Packed CBOR: Table set up by reference
draft-amsuess-cbor-packed-by-reference-04

Abstract

Packed CBOR is a mechanism for transforming Concise Binary Object Representation (CBOR) data into a more compact form. This document introduces a means for setting up its tables by means of dereferenceable identifiers, and introduces a pattern of using it without sending long identifiers.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at <https://chrysn.codeberg.page/packed-by-reference/draft-amsuess-cbor-packed-by-reference.html>. Status information for this document may be found at <https://datatracker.ietf.org/doc/draft-amsuess-cbor-packed-by-reference/>.

Discussion of this document takes place on the CBOR Working Group mailing list (<mailto:cbor@ietf.org>), which is archived at <https://mailarchive.ietf.org/arch/browse/cbor/>. Subscribe at <https://www.ietf.org/mailman/listinfo/cbor/>.

Source for this draft and an issue tracker can be found at <https://codeberg.org/chrysn/packed-by-reference>.

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 4 September 2025.

Copyright Notice

Copyright (c) 2025 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	2
2. Setting up the tables by reference	3
2.1. Count vs. content of source	4
2.1.1. Not all known entries are used	4
2.1.2. Unknown entries are used evolution of sources	4
2.2. Setup with skipped indices	5
2.3. Example	6
3. Nested table setups	6
3.1. Example of nested table setup	7
4. Security Considerations	7
5. IANA Considerations	7
5.1. CBOR Tags Registry	7
6. References	8
6.1. Normative References	8
6.2. Informative References	8
Appendix A. Change log	9
Acknowledgments	9
Author's Address	9

1. Introduction

Packed CBOR [I-D.ietf-cbor-packed] is a mechanism for transforming Concise Binary Object Representation (CBOR, [RFC8949]) data into a more compact form that can be used without a decompression step.

It depends on packing tables, which can be set up through different means: they can come from the CBOR item's context, be populated in the item itself, or use newly defined CBOR tags. This document defines such a tag that uses dereferenceable identifiers to set up a table, and introduces a pattern of using it without sending long identifiers.

2. Setting up the tables by reference

CBOR tag TBD213 is defined with semantics similar to tags TBD113 and TBD1113 from [I-D.ietf-cbor-packed] in that it sets up tables around a rump.

```
Packed-By-Reference = #6.<tbd213>([count, source, rump])
rump = any
source = CRI / ~uri
count = (count-shared-and-argument //      ; similar to tag 113
         count-shared, count-argument )    ; similar to tag 1113

count-shared-and-argument = uint
count-shared = uint
count-argument = uint

tbd213 = 213      ; preliminary value, see IANA considerations
```

The items inserted by the tables are not given explicitly, but picked out of tables known by their identifier given as source. Such a source needs to represent two lists of CBOR items, one for each kind of tables (one for shared item, one for arguments). The tag prepends some number of items out of those source lists to the tables that are used to unpack the rump.

The identifier is given as a URI string (as defined in [RFC3986]) or equivalently as a CRI (as defined in [I-D.ietf-core-href]). Later iterations of this document may introduce additional options.

```
// If the stand-in concept of [I-D.bormann-cbor-yang-standin] is
// generalized, the source item may become the raw list of tables,
// possibly disallowing the CRI and URI variants. Given that tags
// 113 and 1113 are capable of expressing cases where the source
// tables are present, tag TBD213 should then be used by using a
// dereferencing stand-in in the source position. When the source
// identifier is dereferenceable, all considerations from
// [I-D.bormann-t2trg-deref-id] apply. (Simplifying: No dereferencing
// at runtime -- the recipient either knows it already or treats it as
// unknown).
```

If the same number of items is prepended to both tables, their count is given as a single number; otherwise, the numbers are given separately.

Encoders SHOULD use the most compact form of count, and SHOULD pick the lowest count(s) sufficient to encode the items contained in the rump. When those conflict, they may prioritize either. If the source supports evolution of sources (see Section 2.1.2), disregarding that recommendation may pose an interoperability hazard.

2.1. Count vs. content of source

The count encoded for the number of table entries given in a document will often mismatch with the number of entries the receiver of a document knows to be present in the given source.

2.1.1. Not all known entries are used

If the encoded count is less than the number of known entries, this merely expresses that the originator of the document did not use the higher numbers. When a document's tables are populated from multiple sources, encoding the smallest possible count is useful because the table indices used throughout the document stay small and can thus be encoded concisely.

2.1.2. Unknown entries are used evolution of sources

If the encoded count is larger than the number of known entries, this indicates that the document may contain references that the receiver does not know. This can happen when a source has been evolved compatibly to contain more entries, compared to when the receiver learned of the source definition. Source entries beyond the receiver's knowledge stay unpopulated in the receiver's tables, but still shift existing entries to higher indices.

Some CBOR protocols come with elements that support isolation of processing errors. For example, a CRI that uses unknown extensions is regarded as "unprocessable" (Section 5.2.1 of [I-D.ietf-core-href]). It cannot be resolved, is unequal to any other CRI (unless they are identical), but does not inhibit the processing of its surrounding document.

In such protocols, references to unpopulated table entries can be tolerated as described in Section 2.1 of [I-D.ietf-cbor-packed]. Care has to be taken around processing tag TBD1112: If that tag is produced in the course of unpacking, comparisons for identity are not reliable. Similarly, if the unpacking mechanism provides access to the serialized form of the unprocessable entity, identity comparisons are only reliable if the items being compared have the same table setup applied.

```
// Protocols may also pre-populated entries with values that are
// reserved in the protocol and specified to be ignored at reception.
// Later, when the entries are specified, concrete values take their
// places. This has roughly the same effect, but is harder to
// describe. (This paragraph may be removed later unless it is found
// to be particularly useful).
```

Protocols that do not support error isolation need a way to negotiate the understood set of sources and table entries.

2.1.2.1. Evolution beyond adding items

The content of tables may be altered in more ways than just adding entries that were previously unpopulated. Such changes are NOT RECOMMENDED, because while they can be done in a compatible way, providing criteria for this are out of scope of this document.

```
// If a later version of this document uses stand-in values more
// actively, this section will need to be revisited: In that case,
// the tables may be part of the outer source, and then those would
// grow internally.
```

2.2. Setup with skipped indices

If a large number of items at the beginning of the source tables would not be used, there is an additional four-argument form of count that defines a number of items in the source tables that are skipped before selecting items into the table. This allows keeping the indices low and therefore compact.

```
count //= (
    skip-shared, count-shared, skip-argument, count-argument
)
```

```
skip-shared = uint
skip-argument = uint
```

Source tables should be designed in such a way that commonly used items are at the start to minimize the necessity for the four-argument form.

2.3. Example

Suppose the URI "tag:example.com,2023:byref" defines the items ["price", "category", "author", "title", "fiction", 8.95, "isbn"] in both tables. Then the example in figure 3 of [I-D.ietf-cbor-packed] can be written as:

```
213([7, "tag:example.com,2023:byref"
  [{"store": {
    "book": [
      {simple(1): "reference", simple(2): "Nigel Rees",
        simple(3): "Sayings of the Century", simple(0): simple(5)},
      {simple(1): simple(4), simple(2): "Evelyn Waugh",
        simple(3): "Sword of Honour", simple(0): 12.99},
      {simple(1): simple(4), simple(2): "Herman Melville",
        simple(3): "Moby Dick", simple(6): "0-553-21311-3",
        simple(0): simple(5)},
      {simple(1): simple(4), simple(2): "J. R. R. Tolkien",
        simple(3): "The Lord of the Rings",
        simple(6): "0-395-19395-8", simple(0): 22.99}],
    "bicycle": {"color": "red", simple(0): 19.95}}}]])
```

Assuming that the underlying CBOR protocol defines that unknown keys on goods may be ignored, an older receiver that only knows the first 5 entries of the source tables could still process the document, but would be missing all ISBNs and the price of one item.

3. Nested table setups

Documents that use tables from multiple sources can easily spend many bytes on listing source identifiers. A pattern that reduces the verbosity while staying unambiguous are nested table setups, where the outer tables are extended to contain additional identifiers.

In this pattern, tables are set up in two stages:

The outer stage contains the CRIs or URIs that may later be used as source values. (It may also contain other items). The inner stage is set up using tag TBD213, and the source given is a packed reference.

All table inputs can be evolved orthogonally as described in Section 2.1.2. If an unspecified entry is used as a source, the whole source content is considered unspecified.

3.1. Example of nested table setup

In this example, the initial table set up is provided by the media type, and contains these items:

```
* 0: "This class has students with the following names"
* 100: "tag:example.com,2023:english-names.txt"
* 101: "tag:example.com,2023:german-names.txt"

213([5, 6(42) / outer item 100 /,
    213([2, 6(45) / outer item 101, currently item 105 /,
        [simple(11) / outer item 0, currently item 11,
            "This class has students with the following names" /,
            simple(0) / item 0 of german-names, "Franz" /,
            simple(2) / item 0 of english-names, currently item 2, "George" /,
            simple(1) / item 1 of german-names, "Fritz" /,
            simple(7) / item 5 of english-names, currently item 7, "Jack" /
        ]))])
```

Note that a constrained implementation of a decoder may not even have the fully expanded form of the URIs or CRIs available; it may only be capable of using these table entries in the source position and then find the shipped source lists.

4. Security Considerations

General security considerations from [I-D.ietf-cbor-packed] and [I-D.bormann-t2trg-deref-id] apply. In particular, any security implications of different applications disagreeing about what tables are implied by a media type apply likewise to situations when different applications disagree about the tables from a specified source.

5. IANA Considerations

5.1. CBOR Tags Registry

In the registry "CBOR Tags", IANA is requested to allocate one tag:

```
* Tag: 213
* Data item: Array [count(s), source, rump]
* Semantics: "Packed CBOR: table setup"
* Reference: This document
```

6. References

6.1. Normative References

- [I-D.bormann-t2trg-deref-id]
Bormann, C. and C. Amsss, "The "dereferenceable identifier" pattern", Work in Progress, Internet-Draft, draft-bormann-t2trg-deref-id-05, 3 March 2025, <<https://datatracker.ietf.org/doc/html/draft-bormann-t2trg-deref-id-05>>.
- [I-D.ietf-cbor-packed]
Bormann, C. and M. Gtschow, "Packed CBOR", Work in Progress, Internet-Draft, draft-ietf-cbor-packed-14, 3 March 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-cbor-packed-14>>.
- [I-D.ietf-cbor-update-8610-grammar]
Bormann, C., "Updates to the CDDL grammar of RFC 8610", Work in Progress, Internet-Draft, draft-ietf-cbor-update-8610-grammar-06, 24 June 2024, <<https://datatracker.ietf.org/doc/html/draft-ietf-cbor-update-8610-grammar-06>>.
- [I-D.ietf-core-href]
Bormann, C. and H. Birkholz, "Constrained Resource Identifiers", Work in Progress, Internet-Draft, draft-ietf-core-href-18, 3 February 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-core-href-18>>.
- [RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, DOI 10.17487/RFC3986, January 2005, <<https://www.rfc-editor.org/rfc/rfc3986>>.
- [RFC8949] Bormann, C. and P. Hoffman, "Concise Binary Object Representation (CBOR)", STD 94, RFC 8949, DOI 10.17487/RFC8949, December 2020, <<https://www.rfc-editor.org/rfc/rfc8949>>.

6.2. Informative References

[I-D.bormann-cbor-yang-standin]

Bormann, C. and M. Matjka, "Stand-in Tags for YANG-CBOR",
Work in Progress, Internet-Draft, draft-bormann-cbor-yang-
standin-00, 21 February 2024,
<<https://datatracker.ietf.org/doc/html/draft-bormann-cbor-yang-standin-00>>.

Appendix A. Change log

From -03 to -04:

- * Aligned terminology with latest Packed CBOR.

From -02 to -03:

- * Switched from CPA114 to CPA213 to stay out of Carsten's dangerous ASCII region.
- * Add security considerations.
- * Provide an actual introduction.
- * Minor simplifications.

From -01 to -02:

- * Add text on use of unpopulated items, and rationale to count in general.
- * Split 4-argument form into its own subsection
- * Fix erroneous example
- * Augment CDDL with comments and [I-D.ietf-cbor-update-8610-grammar]
- * Add considerations for splitting between loading and importing through stand-ins
- * Write IANA considerations
- * Editorial changes

Acknowledgments

[TBD]

Author's Address

Christian Amsss
Austria
Email: christian@amsuess.com