A Concise Binary Object Representation (CBOR) of DNS Messages (draft-lenders-dns-cbor)

Evaluation of Name Compression

Martine S. Lenders (martine-lenders@tu-dresden.de), Carsten Bormann, Thomas C. Schmidt, Matthias Wählisch
IETF CBOR WG Interim Meeting, 2023-10-18
Outline

Motivation

Name Compression Ideas

Evaluation

Conclusion

Addendum: Flag Rotation
Last Meeting: Compression Ratios & Byte Savings

**Compression ratio**

![Compression ratio CDF](image)

**Byte savings**

![Byte savings CDF](image)
Last Meeting: Compression Ratios & Byte Savings

Compression ratio

Byte savings

CDF

Queries

Responses (w/ question)

Responses (w/o question)

Responses (w/ question, packed)

Responses (w/o question, packed)
Last Meeting: Compression Ratios & Byte Savings

Compression ratio

<table>
<thead>
<tr>
<th>Compression ratio</th>
<th>0.0</th>
<th>2.0</th>
<th>4.0</th>
<th>6.0</th>
<th>8.0</th>
<th>10.0</th>
<th>12.0</th>
<th>14.0</th>
<th>16.0</th>
<th>18.0</th>
<th>20.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Byte savings

<table>
<thead>
<tr>
<th>Savings [bytes]</th>
<th>0</th>
<th>150</th>
<th>300</th>
<th>450</th>
<th>600</th>
<th>750</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

 Queries

<table>
<thead>
<tr>
<th>Responses (w/ question)</th>
<th>Responses (w/o question)</th>
<th>Responses (w/ question, packed)</th>
<th>Responses (w/o question, packed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.0</td>
<td>4.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

DNS

id 63962
opcode QUERY
code NOERROR
flags QR RD RA

;QUESTION
incoming.telemetry.mozilla.org. IN A

;ANSWER
telemetry-incoming.rXX-X.services.mozilla.com. 299 IN CNAME
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.7.5
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.9.102
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.6.6
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.8.214
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.3.214
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.5.226
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.3.170
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com. 60 IN A Y.Y.3.226

;AUTHORITY

⇒ Name compression needed in base format?
Ideas for Name Compression

Discussed @ 2023-10-04 CBOR Interim:

1. DNS-wire-format-style: Reference name components within CBOR object
2. Packed CBOR “lite”: Reference name suffixes in pre-defined table
Idea 1: Reference name components

• Loosely based on Christian’s idea¹:
  • Current draft: `domain-name = tstr .regexp "([^.]\+[.])*[^.]+"`
  • Proposal: `domain-name = (*comp)`
    (cf. core-href)

• Use componified name to also allow for `tag(uint)` in name, i.e., reference:
  
  \[
  \text{comp} = tstr / \#6.t(uint)
  \]

• Reference tag `t(i)`: Use name components from `i`-th `tstr` in CBOR object

• End of name? Stop if object is neither `tstr` nor tag `t` (i.e. not `comp`)

• Evaluated with `t = 7` (Tag length 1+0 bytes) and `t = 48` (Tag length 1+1 bytes)

¹https://mailarchive.ietf.org/arch/msg/cbor/JOHCCBOzC46PrSq-61MMev--mpU/
Idea 1: Reference name components

- Loosely based on Christian’s idea¹:
  - Current draft: \texttt{domain-name = tstr .regexp \texttt{"([\^.]+[\.])\*[\^.]+"}}
  - Proposal: \texttt{domain-name = (*comp)}
    (cf. \texttt{core-href})
  - Use componified name to also allow for \texttt{tag(uint)} in name, i.e., reference:
    \[
    \text{comp} = \text{tstr} / \#6.\text{t(uint)}
    \]

\[
\text{[ ["example", "org", 12, 1],} \]
\[
\text{[ [3600, "coap", "udp", "local"],]} \]
\[
\text{[ [3600, 2, "ns1", \texttt{t(0)}], [3600, 2, "ns2", \texttt{t(0)}],} \]
\[
\text{[ [ \texttt{t(2)}, 3600,} \]
\text{h'20010db8000000000000000000000001' ],} \]
\[
\text{[ [ \texttt{t(5)}, 3600,} \]
\text{h'20010db8000000000000000000000002' ],} \]
\[
\text{[ [ \texttt{t(6)}, 3600,} \]
\text{h'20010db8000000000000000000003535' ] } \]
\]

- Reference tag \texttt{t(i)}: Use name components from \texttt{i-th tstr} in CBOR object
- End of name? Stop if object is neither \texttt{tstr} nor tag \texttt{t} (i.e. not \texttt{comp})
- Evaluated with \texttt{t = 7} (Tag length 1+0 bytes) and \texttt{t = 48} (Tag length 1+1 bytes)

¹https://mailarchive.ietf.org/arch/msg/cbor/JOHCCBOzC46PrSs-61MMev--mpU/
Easy to implement (at least in Python):

```python
from cbor2 import CBORTag

REF_TAG = 7
INITIAL_STATE = ({}, 0)

def encode_name(name: str, state: Tuple[dict, int]) -> List[Union[str, CBORTag]]:
    comps = name.split("."")
    res = []
    for i, comp in enumerate(comps):
        suffix = ".".join(comps[i:])
        if suffix in state[0]:
            res.append(CBORTag(REF_TAG, state[0][suffix]))
        else:
            state[0][suffix] = state[1]
            state[1] += 1
            res.append(comp)
    return res
```

(no need to count occurrences compared to packed CBOR)
Idea 2: Packed CBOR “lite”

• More lightweight version of proposal by Lemogue et al.
  • Use array instead of dict for name_ref

[ ["example.org", "_coap._udp.local",
  216("ns1.")],
 [simple(0), 12, 1],
 [3600, simple(1)]],
 [[3600, 2, simple(2)], [3600, 2, simple(3)]],
 [ [ simple(1), 3600,
     h'20010db80000000000000000000000001' ],
   [ simple(2), 3600,
     h'20010db80000000000000000000000002' ],
   [ simple(3), 3600,
     h'20010db8000000000000000000000003535' ] ]

• Like ; packed=1 but only take names into account
Idea 2: Packed CBOR “lite”

- More lightweight version of proposal by Lemogue et al.
- Use array instead of dict for `name_ref`

```plaintext
[ ["example.org", "+coap._udp.local",
   216("ns1."), 216("ns2."),
   [simple(0), 12, 1],
   [3600, simple(1)]],
  [[3600, 2, simple(2)], [3600, 2, simple(3)]],
  [[ simple(1), 3600,
    h'20010db8000000000000000000001',
    [ simple(2), 3600,
      h'20010db8000000000000000000002'],
    [ simple(3), 3600,
      h'20010db8000000000000000003535']
  ] ]
```

- Like `;packed=1` but only take names into account
**Evaluation: Data Corpus**

<table>
<thead>
<tr>
<th>YourThings¹</th>
<th>IoTFinder²</th>
<th>MonIoTr³</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Collected throughout 2019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 90 consumer IoT devices from 50 vendors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Phones, Tablets, PCs, ... (in IoTFinder &amp; YourThings)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 1.20 million queries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 2.74 million responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 2.07 million w/o mapped query (IoTFinder contains only responses)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 0.66 million w/ mapped query</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Evaluation

- Implementation: https://github.com/netd-tud/cbor4dns
- Applied compression to data corpus
  - Elide query in response if it is present in data corpus
    - draft-04 (as described in draft version 4)
      - unpacked “classic” CBOR+DNS (application/cbor+dns)
      - packed packed CBOR+DNS (application/cbor+dns;packed=1)
  - comp. ref. (Reference components with tag $t$)
    - unpacked With $t = 7$ and $t = 48$
    - packed With $t = 7$ with strings only as shared & as described in draft-04
  - Packed lite
    - unpacked CBOR packed lite (only responses)
    - packed as described in draft-04
Original Results (draft-04)

<table>
<thead>
<tr>
<th>Compression ratio</th>
<th>Byte savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\text{len(wire format)}}{\text{len(CBOR format)}} )</td>
<td>( \text{len(wire format)} - \text{len(CBOR format)} )</td>
</tr>
</tbody>
</table>
Name Compression: Queries

Compression ratio

Compression ratio

Byte savings

Savings [bytes]
Name Compression: Queries

Only small advantage, e.g., with DNS-SD queries

Compression ratio

Byte savings

Savings [bytes]
Name Compression: Responses (w/ question)

Compression ratio

Byte savings
Name Compression: Responses (w/o q.)

**Compression ratio**

<table>
<thead>
<tr>
<th>Compression ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>2.0</td>
</tr>
<tr>
<td>4.0</td>
</tr>
<tr>
<td>6.0</td>
</tr>
<tr>
<td>8.0</td>
</tr>
<tr>
<td>10.0</td>
</tr>
<tr>
<td>12.0</td>
</tr>
<tr>
<td>14.0</td>
</tr>
<tr>
<td>16.0</td>
</tr>
<tr>
<td>18.0</td>
</tr>
<tr>
<td>20.0</td>
</tr>
</tbody>
</table>

**CDF**

- 0.0
- 25
- 50
- 75
- 100
- 125
- 150

**Byte savings**

<table>
<thead>
<tr>
<th>Savings [bytes]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
</tr>
<tr>
<td>1.0</td>
</tr>
</tbody>
</table>
Compression ratio

Byte savings

Comp. ref. better with smaller compression ratios/savings, esp. with 1+0 Tag
(Packed lite introduces small overhead due to extra list)
Packed CBOR: Shared Value vs Inverted Reference for Name Components

Compression ratio

Byte savings

CDF

Write savings [bytes]

unpacked
str. as value ref.
draft-04
Packed CBOR: Shared Value vs Inverted Reference for Name Components

- Small advantage for shared value reference
- No surprise: Similar suffixes in name components not a given
  ⇒ Unnecessary overhead added for random similarities
Conclusion

- Name compression leads smaller messages to other formats
  - Nearly no compression ratio <1.0
- Difference between component referencing and packed lite marginal
- Personal preference for component referencing
  - Conciseness: No added byte overhead due to extra list
  - Simplicity: Encoding easy to implement (i.e. \texttt{packed=1} can stay only for responses)
  - Familiarity: More similar to classic DNS name compression
Addendum: Flag Rotation

- Rotate flags? (QR flag in default flags, DO flag in extended flags are MSB)

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| DO| Z |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

In CBOR: 19 80 00
Rotated CBOR: 01

- Also evaluated: Advantage negligible (most often just those 2 bytes less)