

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

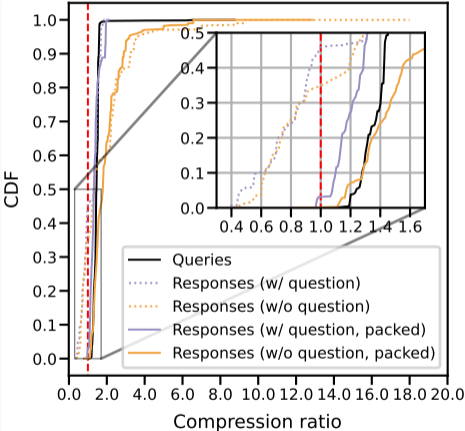
Evaluation

Conclusion

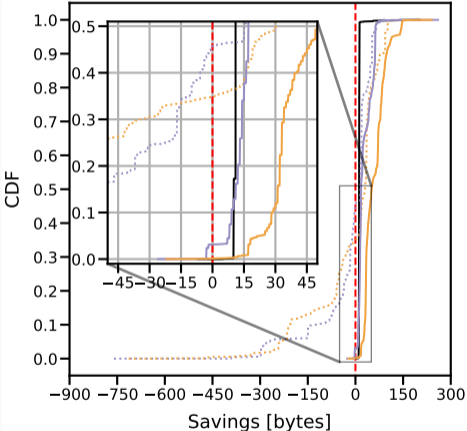
Addendum: Flag Rotation

Last Meeting: Compression Ratios & Byte Savings

Compression ratio

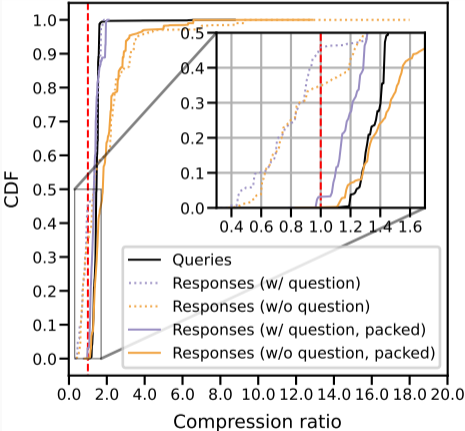


Byte savings

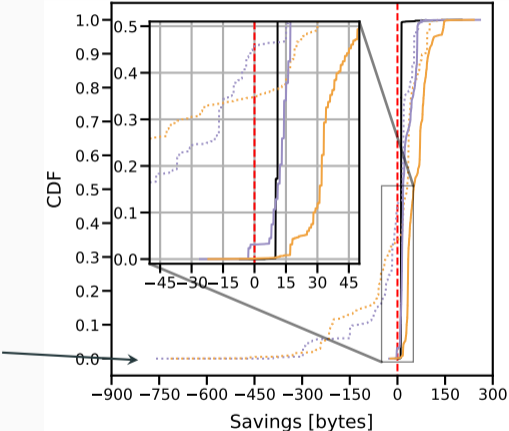


Last Meeting: Compression Ratios & Byte Savings

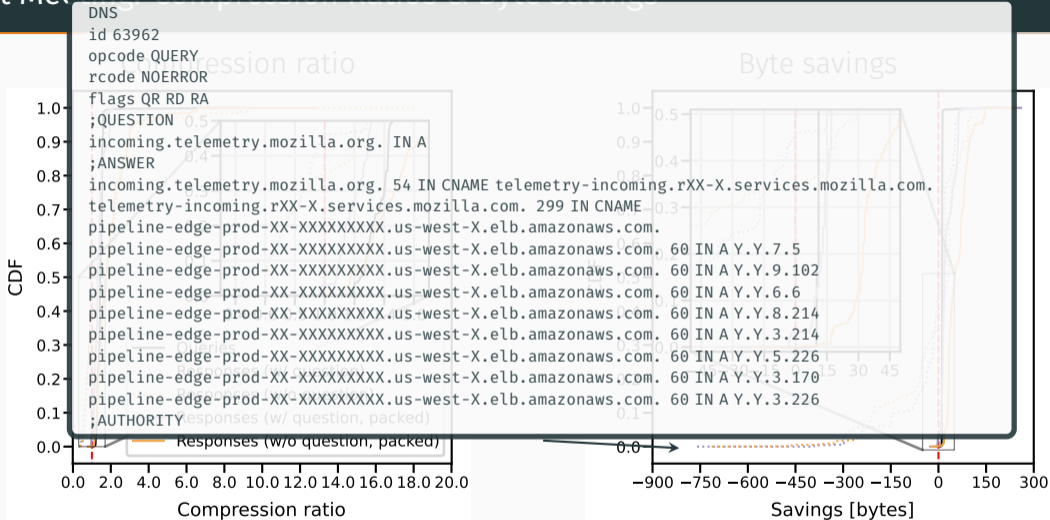
Compression ratio



Byte savings



Last Meeting: Compression Ratios & Byte Savings

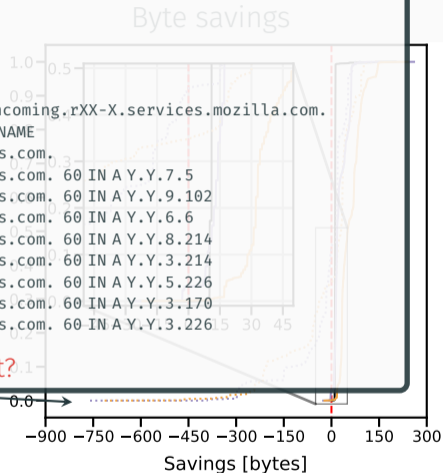
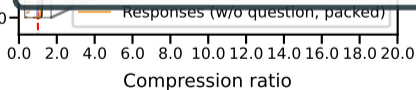


Last Meeting: Compression Ratios & Byte Savings

```
DNS
id 63962
opcode QUERY
rcode NOERROR
flags QR RD RA
;QUESTION
incoming.telemetry.mozilla.org. IN A
;ANSWER
incoming.telemetry.mozilla.org. 54 IN CNAME telemetry-incoming.rXX-X.services.mozilla.com.
telemetry-incoming.rXX-X.services.mozilla.com. 299 IN CNAME
pipeline-edge-prod-XX-XXXXXXXXX.us-west-X.elb.amazonaws.com.
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
```

CDF

⇒ 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:

`comp = tstr / #6.t(uint)`

```
[ ["example", "org", 12, 1],  
  [[3600, "_coap", "_udp", "local"]],  
  [[3600, 2, "ns1", t(0)], [3600, 2, "ns2", t(0)]],  
  [ [ t(2), 3600,  
      h'20010db8000000000000000000000001'],  
    [ t(5), 3600,  
      h'20010db8000000000000000000000002'],  
    [ t(6), 3600,  
      h'20010db800000000000000000000003535']  
  ] ]
```

- 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)

¹<https://mailarchive.ietf.org/arch/msg/cbor/JOHCCB0zC46PrSq-61MMev--mpU/>

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:

`comp = tstr / #6.t(uint)`

```
[ ["example", "org", 12, 1],  
  [[3600, "_coap", "_udp", "local"]],  
  [[3600, 2, "ns1", t(0)], [3600, 2, "ns2", t(0)]],  
  [ [ t(2), 3600,  
      h'20010db8000000000000000000000001'],  
    [ t(5), 3600,  
      h'20010db8000000000000000000000002'],  
    [ t(6), 3600,  
      h'20010db800000000000000000000003535']  
  ] ]
```

- 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)

¹<https://mailarchive.ietf.org/arch/msg/cbor/JOHCCB0zC46PrSq-61MMev--mpU/>

Idea 1: Note on Implementation

Easy to implement (at least in 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`

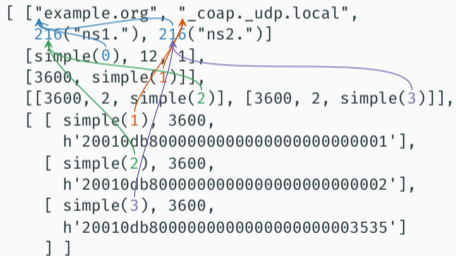
```
[ ["example.org", "_coap._udp.local",  
  216("ns1."), 216("ns2.")]  
 [simple(0), 12, 1],  
 [3600, simple(1)]],  
 [[3600, 2, simple(2)], [3600, 2, simple(3)]],  
 [ [ simple(1), 3600,  
     h'20010db8000000000000000000000001'],  
   [ simple(2), 3600,  
     h'20010db8000000000000000000000002'],  
   [ simple(3), 3600,  
     h'20010db800000000000000000000003535']  
 ] ]
```

- 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`

```
[ ["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'20010db8000000000000000000000001'],  
   [ simple(2), 3600,  
     h'20010db8000000000000000000000002'],  
   [ simple(3), 3600,  
     h'20010db800000000000000000000003535']  
 ] ]
```



- Like `;packed=1` but only take names into account

Evaluation: Data Corpus

YourThings¹

IoTFinder²

MonIoT³

- Collected throughout 2019
- 90 consumer IoT devices from 50 vendors
 - + Phones, Tablets, PCs, ... (in IoTFinder & YourThings)
- 1.20 million queries
- 2.74 million responses
 - 2.07 million w/o mapped query (IoTFinder contains only responses)
 - 0.66 million w/ mapped query

¹O. Alrawi, C. Lever, M. Antonakakis, and F. Monrose. 2019. **SoK: Security Evaluation of Home-Based IoT Deployments**. In *IEEE S&P 2019*. 1362–1380.

²R. Perdisci, T. Papastergiou, O. Alrawi, and M. Antonakakis. 2020. **IoTFinder: Efficient Large-Scale Identification of IoT Devices via Passive DNS Traffic Analysis**. In *IEEE EuroS&P 2020*. 474–489.

³J. Ren, D.J. Dubois, D. Choffnes, A.M. Mandalari, R. Kolcun, and H. Haddadi. 2019. **Information Exposure for Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach**. In *Proc. of the Internet Measurement Conference (IMC)*. ACM.

- 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*

Compression ratio

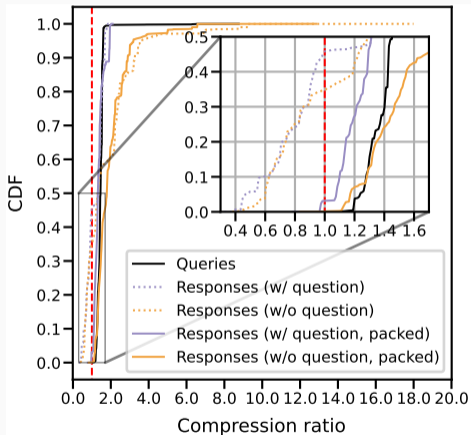
$$\frac{\text{len(wire format)}}{\text{len(CBOR format)}}$$

Byte savings

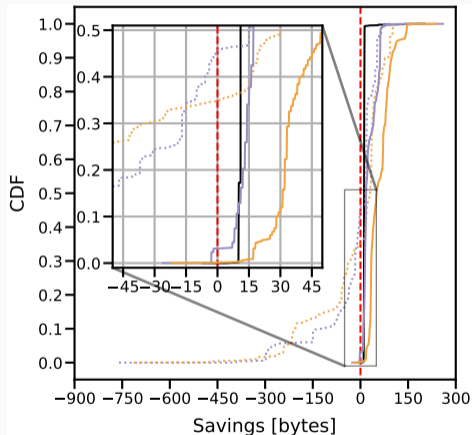
$$\text{len(wire format)} - \text{len(CBOR format)}$$

Original Results (draft-04)

Compression ratio

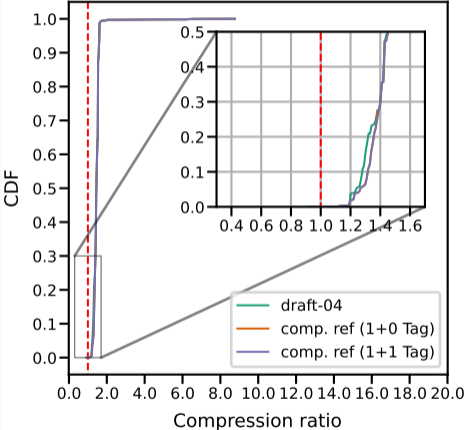


Byte savings

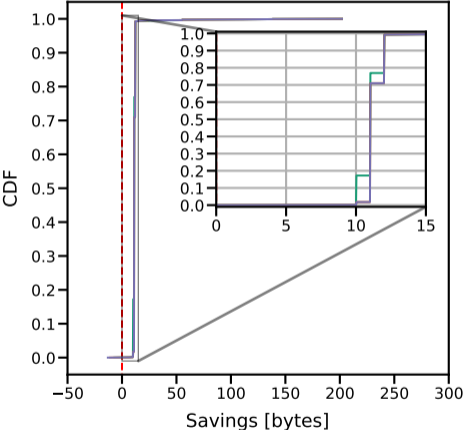


Name Compression: Queries

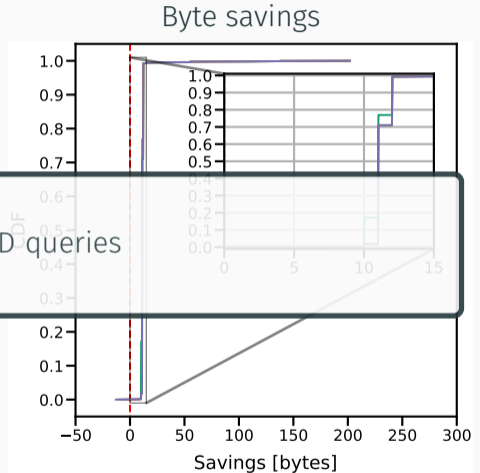
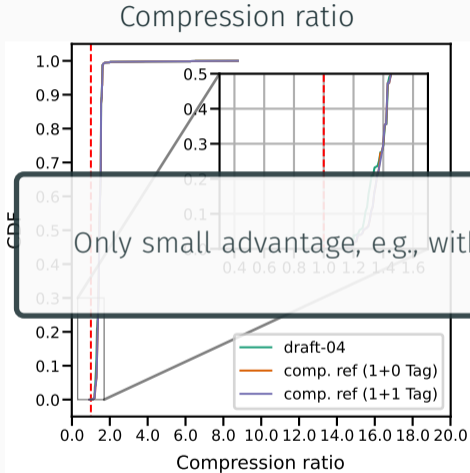
Compression ratio



Byte savings



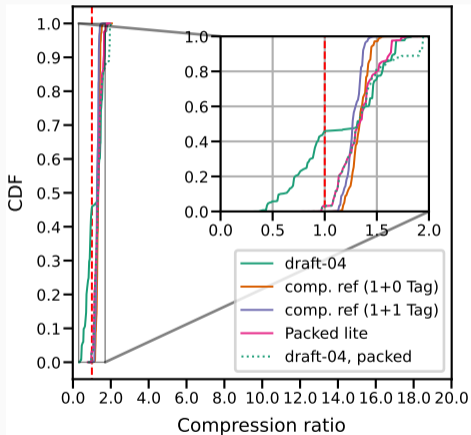
Name Compression: Queries



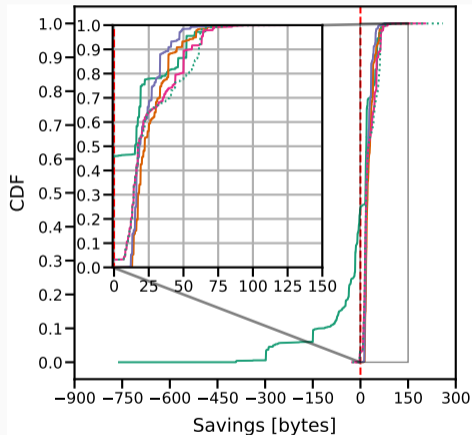
Only small advantage, e.g., with DNS-SD queries

Name Compression: Responses (w/ question)

Compression ratio

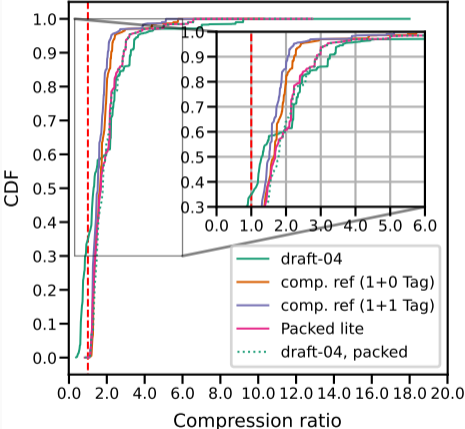


Byte savings

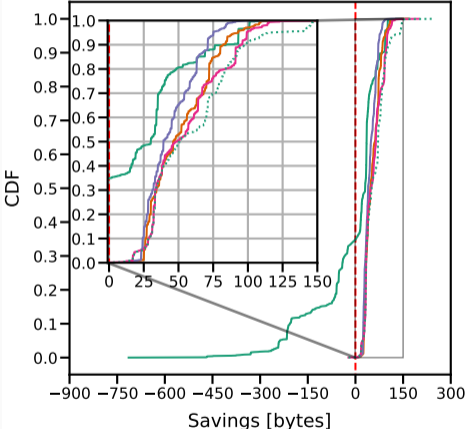


Name Compression: Responses (w/o q.)

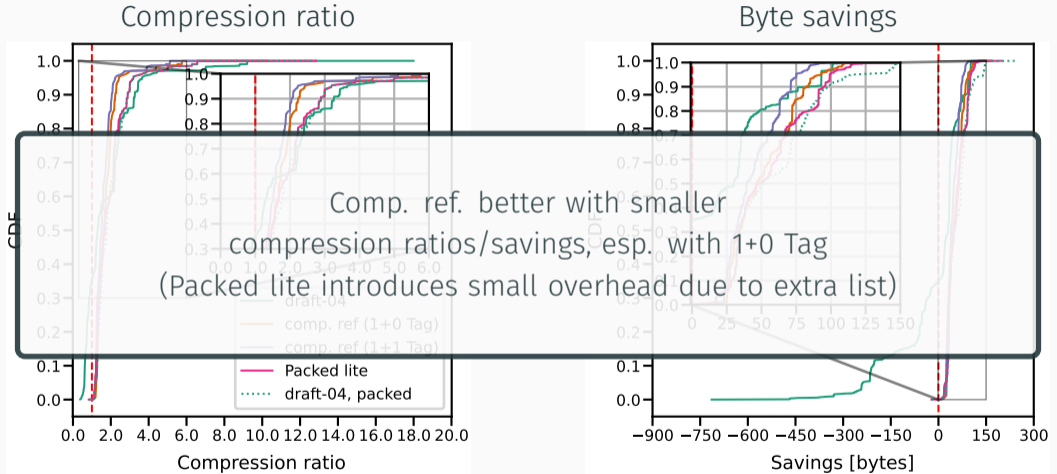
Compression ratio



Byte savings

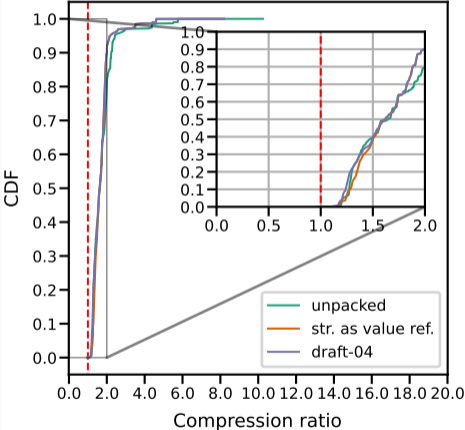


Name Compression: Responses (w/o q.)

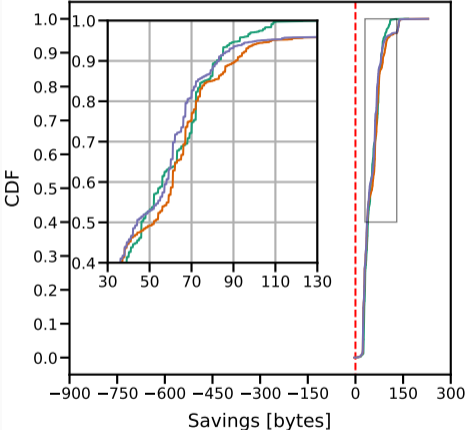


Packed CBOR: Shared Value vs Inverted Reference for Name Components

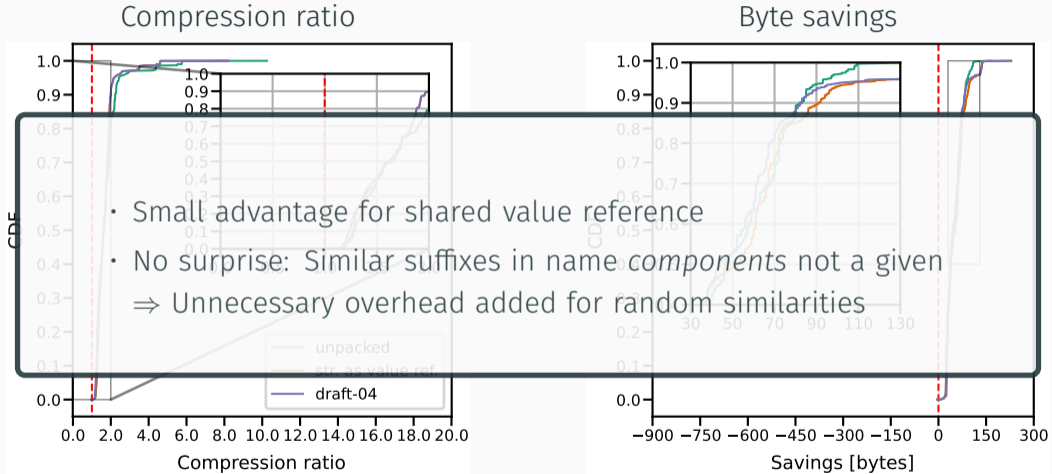
Compression ratio



Byte savings



Packed CBOR: Shared Value vs Inverted Reference for Name Components



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. **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)



In CBOR: **19 80 00**

Rotated CBOR: **01**

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