Workgroup: IAB e-impact Workshop
Published: 31 October 2022

Authors: B. Moran, *Arm Limited*; H. Birkholz, *Fraunhofer SIT*; C. Bormann, *Universität Bremen TZI*

CBOR is Greener than JSON

Abstract

This short position paper illustrates energy use considerations for data formats used in networks. It is intended to support green-focused steering decisions for various Internet stakeholders.

Table of Contents

1. Introduction

For encodings of predominantly non-text content, binary encodings provide performance benefits over textual encodings. ASCII representations of binary or numeric data are substantially larger than binary representations of the same data. Different representation formats are often said to differ in their simplicity of use in the design and debugging of protocols and documents; however, ultimately Internet Standards are for end users [RFC8890]. Implementers often prefer a certain simplicity and focus on the implementation quality of experience, but end users do not care what is perceived to be easier for design and debugging. End users, people who own or operate IoT devices, particularly those that are powered by primary cells, care that:

- The battery is easily replaceable OR the battery will last the full service life of the device.
- The device has a small ecological footprint (this may not be true for all end users, but it is a growing concern and the topic of this IAB workshop).
- The device completes any tasks it is given in a "reasonable" time.

Binary encodings outperform textual representations across each of these metrics. They are simpler to encode and decode, more energy efficient and consume less bandwidth. Therefore, Internet Standards should favor binary encodings over textual representations in any scenario where textual representation does not confer specific benefits. In general, this is only the case where the entire content is predominantly text. Data compression technologies are not discussed in this memo; they can offer additional size benefits, but their use also can increase memory and processing requirements and overall complexity.

2. Comparison of Encodings

To demonstrate systemic energy savings through the use of binary encodings and, thus, the ecological impact of encodings, JSON energy consumption is compared to CBOR energy consumption across a range of examples. This comparison is based on the assumption that, in energy-constrained environments, processing requirements are dominated by the energy expenditure for transmission and reception.

For the comparison illustrated in this document, transmission over LoRa is used [LoRa]. LoRa is a widely deployed IoT WAN networking protocol that is attractive due to its low power consumption, ease of deployment, and simple software stack.

There are numerous related protocols that deserve similar attention, but are not covered in this contribution:

- 6LoWPAN
- Bluetooth Mesh
- NB-IoT
- WiFi
- Ethernet

However, the scaling properties of energy use across all protocols are likely to be similar. The differences generally arise from packet overhead and maximum payload size. As a result, analysis based on a single IoT networking protocol provides adequate reference information for analyzing the impact of encoding on energy consumption and throughput.

3. Impact of Encoding Based on Data-Type

Different data types can have different encoding impacts.

For instance, CBOR encodes integers using the following rules:

Unsigned Integer value	encoding size
0 .. 23	1
24 .. 255	2
256 .. 65535	3
65536 .. 2^{32}-1	5
2^{32} .. 2^{64}-1	9

Table 1: CBOR Encoding Sizes for Unsigned Integers

Based on this, the differences between encodings are shown below. The JSON size column assumes no redundant blank space is sent to bring about some readability by humans. Where UINT(arg) is shown in the CBOR size column, the encoding size above is used based on the value of arg.

Type	JSON Size	CBOR Size
string	strlen+2 + escaping	strlen + UINT(strlen)
octets (hex)	bytesize * 2	bytesize + UINT(bytesize)
octets (b64)	bytesize * 4/3	bytesize + UINT(bytesize)
int8	1 to 3	1 or 2
int16	1 to 5	3
int32	1 to 10	5
int64	1 to 19	9
float32	3 to 16	5
float64	3 to 23	9
Date	12	2 + UINT(days since 1970)

Type	JSON Size	CBOR Size
Array	2 + count-1	UINT(count)
Map	2 + 2*count - 1	UINT(count)

Table 2: Comparing Encoding Sizes

4. Example Data Structures

Example data structures are provided from [RFC8428] (ex_n_), as well as from [RFC9193] and a weather data example (fmi). The examples are encoded in both JSON and CBOR. For information, the examples are also provided in an alternative form that employs half-size float (float16) CBOR; the float16 values are not used for the "reduction" calculations.

A summary of encoding sizes is provided in Table 3.

example	json	cbor	half	red.
ex1	56	47	41	16 %
ex10	141	99	93	29 %
ex11	108	69	63	36 %
ex12	80	46	46	42 %
ex13	152	102	90	32 %
ex2	115	82	70	28 %
ex3	293	195	159	33 %
ex4	278	171	135	38 %
ex5	423	254	188	39 %
ex6	203	131	119	35 %
ex7	160	104	98	35 %
ex8	174	123	117	29 %
ex9	124	78	72	37 %
fmi	4051	2551	1903	37 %
rfc9193-1	125	82	76	34 %

example	json	cbor	half	red.
rfc9193-2	139	91	85	34 %

Table 3: Summary of Encoding Sizes for Example Data

5. LoRa Energy Consumption

The LoRa configuration used in the following analysis is based on using the SX1262 chip as the leaf node and SX1302 + SX1250 as the LoRa concentrator.

The energy consumption calculations in the following tables are based on the following additional premises:

- The leaf node is based on the SX1262, with DC-DC enabled
- Spreading Factor is 7
- Bandwidth is 125 kHz
- Frequency is 868 MHz
- Preamble length is 16
- CRC is enabled
- Coding Rate is 4/5
- Transmit power is +14 dBm with optimal settings (3.3V)
- Receiver is also 3.3 V
- Large payloads are split into packets with no additional framing
- RX Waiting time is not considered
- Calculations are based on the SX1262 datasheet, Section 6.1.4: LoRa(R) Time-on-Air [SX1262].
- No LoRa concentrator RX calculations are provided because a LoRa concentrator is always receiving.
- LoRa concentrator TX calculations are based on the SX1250 LoRa front-end with the SX1302 baseband processor.
- Current consumption in TX and RX mode is based on "typical" numbers provided in the respective datasheets
  - Leaf Node: [SX1262]
  - Concentrator: [SX1250] + [SX1302]

These premises are selected in the context of this contribution to reflect best-case scenarios wherever possible. Regardless, most impacts scale proportionally with encoding size reduction.

Due to network utilization, there are secondary energy consumption impacts that are caused by larger data encodings: as network utilization increases, stronger media access congestion causes more re-transmissions as well as exercises congestion control mechanisms, which in practice are typically more expensive. A need for additional concentrators may arise as a consequence, which can result in further increases in energy consumption.

5.1. LoRa Energy Consumption Values for Data from Examples

Tables 4, #5, and #6 show energy consumption estimates for energy needed to receive and energy needed to transmit on a lead node, as well as energy needed to transmit on a concentrator node. The numbers given are for JSON representation, CBOR representation, and CBOR representation with half-size floating point values; the reduction ("red.") shown is conservatively derived between JSON and CBOR representation without these more compact floating point values. See Appendix A for the method used to arrive at these numbers.

example	json	cbor	half	red.
ex1	1.8	1.5	1.4	13 %
ex10	3.6	2.7	2.5	26 %
ex11	2.9	2.0	1.9	32 %
ex12	2.3	1.5	1.5	34 %
ex13	3.9	2.8	2.5	28 %
ex2	3.1	2.3	2.1	25 %
ex3	7.6	4.8	4.0	37 %
ex4	7.2	4.3	3.5	40 %
ex5	10.4	6.1	4.6	41 %
ex6	5.0	3.4	3.2	32 %
ex7	4.0	2.8	2.7	31 %
ex8	4.3	3.2	3.1	25 %
ex9	3.2	2.2	2.1	31 %
fmi	98.3	62.4	46.4	37 %
rfc9193-1	3.2	2.3	2.1	29 %
rfc9193-2	3.5	2.5	2.4	28 %

Table 4: LoRa leaf-node Receive Energy (mJ)

example	json	cbor	half	red.
ex1	17.2	14.9	13.4	13 %

example	json	cbor	half	red.
ex10	35.5	26.3	24.8	26 %
ex11	28.6	19.5	18.7	32 %
ex12	22.5	14.9	14.9	34 %
ex13	37.7	27.1	24.1	28 %
ex2	30.1	22.5	20.3	25 %
ex3	74.0	46.9	39.3	37 %
ex4	70.2	42.3	33.9	40 %
ex5	102.1	59.8	45.4	41 %
ex6	49.2	33.2	30.9	32 %
ex7	39.3	27.1	26.3	31 %
ex8	42.3	31.7	30.1	25 %
ex9	31.7	21.8	20.3	31 %
fmi	962.1	610.7	454.1	37 %
rfc9193-1	31.7	22.5	21.0	29 %
rfc9193-2	34.7	24.8	23.3	28 %

Table 5: LoRa leaf-node Transmit Energy (mJ)

example	json	cbor	half	red.
ex1	37.7	32.7	29.4	13 %
ex10	77.7	57.7	54.4	26 %
ex11	62.7	42.7	41.1	32 %
ex12	49.4	32.7	32.7	34 %
ex13	82.7	59.4	52.7	28 %
ex2	66.1	49.4	44.4	25 %
ex3	162.1	102.7	86.0	37 %

example	json	cbor	half	red.
ex4	153.8	92.7	74.4	40 %
ex5	223.7	131.0	99.4	41 %
ex6	107.7	72.7	67.7	32 %
ex7	86.0	59.4	57.7	31 %
ex8	92.7	69.4	66.1	25 %
ex9	69.4	47.7	44.4	31 %
fmi	2108.1	1338.0	994.9	37 %
rfc9193-1	69.4	49.4	46.1	29 %
rfc9193-2	76.1	54.4	51.1	28 %

Table 6: LoRa concentrator Transmit Energy (mJ)

5.2. Indirect Impacts of Higher Energy Use.

LoRa nodes are frequently arranged to transmit data periodically, for example, every 10 minutes. This contribution is based on the assumption of using a coin-cell powered LoRa node that is designed to report a message containing either a JSON structure or a CBOR structure each time it wakes up. LoRa has a maximum packet size of 255 bytes. Two detailed examples are illustrated below; they are based on data from examples ex2 and ex3.

We assume an inexpensive CR2032 coin cell battery; 3 V @ 220 mAh, which is 2376 J. A simplifying assumption is that the primary source of energy consumption is the LoRa radio and no other significant energy consumption occurs. This forms the baseline for this contribution. These are optimistic assumptions. All real-world applications will be worse than the figures quoted in the examples.

5.2.1. LoRa Reports with Data from Example ex2

Each report is 115 bytes for JSON or 82 bytes for CBOR. From the LoRa leaf-node Transmit Energy table, this means that each report consumes 30.1 mJ for JSON or 22.5 mJ for CBOR.

	JSON	CBOR
Data Size	115	82
Transmit Energy (mJ)	30.1	22.5
Total Messages	78936	105600

	JSON	CBOR
Total Days	548	733

Table 7: Reports with Data from Example ex2

This means that batteries or devices using CBOR to send data from example ex2 last 33% longer than those using JSON, contributing to less e-waste and battery waste.

5.2.2. LoRa reports with Data from Example ex3

Each report is 293 bytes for JSON or 195 bytes for CBOR. From the LoRa leaf-node Transmit Energy table, this means that each report consumes 74 mJ for JSON or 46.9 mJ for CBOR. This example is interesting because the CBOR payload fits in one LoRa packet, while the JSON payload requires two. This means that the CBOR is both smaller overall and has a lower media access overhead.

	JSON	CBOR
Data Size	293	195
Transmit Energy (mJ)	74	46.9
Total Messages	32108	50660
Total Days	222	351

Table 8: Reports With Data from Example ex3

This means that batteries or devices using CBOR to send example 3 data last 58% longer than those using JSON, contributing to less e-waste and battery waste.

6. Discussion

The findings reported here provide arguments for using concise binary data representations in place of traditional text-based data formats. To strengthen these arguments, the findings documented need to be augmented with data from other parties and additional examples, preferably including real-world measurements of power consumed by transmission, reception, encoding and decoding of messages.

On the other hand, it should be clear from first principles that a more concise message encoding provides advantages in transmission and reception power spent. The examples here provide a rough indication of how significant these advantages can be in messages that may be typical for IoT environments. Larger corpora of such messages need to be collected to obtain a quantitatively stronger statement.

In the end, whether these arguments are considered compelling depends on how compelling the counterarguments are. Many developers consider JSON-based communication to be easier to debug than concise binary communication. In practice, non-trivial amounts of JSON need a tool to look at the data just as a tool is required to look at binary data.

In summary, for systems that operate under severe energy constraints, there are good reasons to select binary encodings by default.

Specific situations that may push the decision into one direction or another include:

- Data that are intrinsically predominantly textual in nature may go well with a textual representation format. XML continues to be the preferred format for marking up text.
- Representation formats that deeply suffer from the limitations of JSON should be replaced by CBOR-based ones; e.g., COSE should always be preferred over JOSE.
- CBOR should be used for conveyance-related structures, while 'human-readability', if needed, should be off-loaded to higher layer tools.
- CBOR should always be used where stable semantics exist that can benefit from CBOR's built-in extensibility.
- Where the data format in question is dominated by other data, e.g., a couple of kilobytes metadata for a video that takes hundreds of megabytes, the argument that binary data reduce representation sizes may not be relevant; other features of CBOR such as its built-in extensibility or its more sophisticated number model may still be relevant arguments.

7. Conclusion

In direct comparison, the typical overhead introduced by a 'human-readable' representation in contrast to a binary representation for message transport has been presented as significant in this short contribution.

A generic approach of establishing binary message transfer supported with tooling for human readability will provide resource savings and therefore constitutes a paramount near term goal. Future evaluation and large scale measurements are required to underpin and establish this as a principal approach.

In summary, Internet Standards should use binary encoding as the primary choice. This is actually an [RFC2119] "SHOULD"; valid exceptions do exist, some of which have been discussed above.

8. Informative References

[LoRa] "What Is LoRa®?", <https://www.semtech.com/lora/what-is-lora>.

[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>.

[RFC8428] Jennings, C., Shelby, Z., Arkko, J., Keranen, A., and C. Bormann, "Sensor Measurement Lists (SenML)", RFC 8428, DOI 10.17487/RFC8428, August 2018, <https://www.rfc-editor.org/info/rfc8428>.

[RFC8890] Nottingham, M., "The Internet is for End Users", RFC 8890, DOI 10.17487/RFC8890, August 2020, <https://www.rfc-editor.org/info/rfc8890>.

[RFC9193] Keränen, A. and C. Bormann, "Sensor Measurement Lists (SenML) Fields for Indicating Data Value Content-Format", RFC 9193, DOI 10.17487/RFC9193, June 2022, <https://www.rfc-editor.org/info/rfc9193>.

[SX1250] "Product Details — SX1250", <https://www.semtech.com/products/wireless-rf/lora-core/sx1250>.

[SX1262] "Product Details — SX1262", <https://www.semtech.com/products/wireless-rf/lora-connect/sx1262>.

[SX1302] "Product Details — SX1302", <https://www.semtech.com/products/wireless-rf/lora-core/sx1302>.

Appendix A. Lora Calculations

LoRa power consumption is calculated with the following equations:

$$Energy = I \times V \times ToA$$

Where:

- I is the current consumed by the transceiver
- V is the voltage applied to the transceiver

$$ToA = \frac{2^{SF}}{BW} \times N_{symbol}$$

Where:

- SF: Spreading Factor (5 to 12)
- BW: Bandwidth (in Hz)
- ToA: the Time on Air in seconds
- Nsymbol: number of symbols

$$N_{symbol} = N_{symbol\_preamble} + 4.25 + 8 + \left\lceil \frac{max\left(8 \times N_{byte\_payload} + N_{bit\_CRC} - 4 \times SF + 8 + N_{symbol\_header}, 0\right)}{4 \times SF} \right\rceil \times (CR + 4)$$

Where:

- N_bit_CRC = 16 if CRC activated, 0 if not
- N_symbol_header = 20 with explicit header, 0 with implicit header

- CR is 1, 2, 3 or 4 for respective coding rates 4/5, 4/6, 4/7 or 4/8

These calculations are derived from the SX1262 datasheet, Section 6.1.4 [SX1262].

Authors' Addresses

Brendan Moran
Arm Limited
United Kingdom
Email: brendan.moran.ietf@gmail.com

Henk Birkholz
Fraunhofer SIT
Germany
Email: henk.birkholz@sit.fraunhofer.de

Carsten Bormann
Universität Bremen TZI
Postfach 330440
D-28359 Bremen
Germany
Phone: +49-421-218-63921
Email: cabo@tzi.org