



HAL
open science

Modèles basés sur CBOR pour les données de séries chronologiques de l'Internet des objets : un pas vers la normalisation

Sebastian Molina, Ivan Martinez, Nicolas Montavont, Laurent Toutain

► To cite this version:

Sebastian Molina, Ivan Martinez, Nicolas Montavont, Laurent Toutain. Modèles basés sur CBOR pour les données de séries chronologiques de l'Internet des objets : un pas vers la normalisation. CoRes 2023 - 8èmes Rencontres Francophones sur la Conception de protocoles, l'évaluation de performances et l'expérimentation de Réseaux de communication, May 2023, Cargese, France. hal-04089503

HAL Id: hal-04089503

<https://hal.science/hal-04089503>

Submitted on 4 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèles basés sur CBOR pour les données de séries chronologiques de l'Internet des objets : un pas vers la normalisation

Sebastian Molina A¹, Ivan Martinez², Nicolas Montavont³, et Laurent Toutain³

IMT Atlantique, IRISA, France

¹*juan.molina-araque@imt-atlantique.fr* ²*ivan@marinomtz.fr* ³*firstname.lastname@imt-atlantique.fr*

Actuellement, il y a une expansion rapide de la technologie IoT (Internet of Things) et un déploiement généralisé des appareils IoT. De ce fait, assurer l'interopérabilité avec les systèmes d'information est devenu un enjeu critique dans le déploiement accéléré de ces dispositifs. Malgré la prévalence de la représentation en série chronologique des informations IoT, un format standardisé n'a pas émergé. En effet, la majorité des études de la littérature se concentrent sur le traitement, la prédiction ou la compression de séries temporelles. Dans cet article, un nouveau format de représentation de séries chronologiques pour les appareils IoT est proposé sur la base de Concise Binary Object Representation (CBOR). Le format introduit des deltas, des balises et différents modèles comme Measurements-based TS (MTS) ou Variable-based TS (VTS) pour permettre une représentation compacte des données et une intégration efficace du système d'information. De plus, la représentation proposée est capable de regrouper de manière compacte les données de séries chronologiques collectées par les appareils IoT et est capable de réduire la quantité de données transmises jusqu'à 76% et 96% par rapport à JavaScript Object Notation (JSON), ce qui entraîne des extensions potentielles de la durée de vie de la batterie, et la durée de vie utile des appareils IoT.

Mots-clefs : Internet des objets IoT, Series Temporelles TS, Interopérabilité, CBOR, JSON.

1 Introduction

The Internet of Things (IoT) market has experienced significant growth in recent years and is expected to continue growing, forecasting 50 billion devices at the end of the decade [Ins, DE]. As the number of devices grows, integrating them into information systems becomes more complex, making it important to find new and innovative ways to manage the data generated by these devices. In most cases, this data is presented as Time Series (TS), and the majority of the studies in this field focus on prediction compression or processing [ea20, ea19, ea18]. Nevertheless, a standardized representation has not been established, to the best of our knowledge.

In this paper, we propose a new versatile format for representing TS based on CBOR [BH18]. It uses CBOR[BH20] to represent measurements which use the difference of subsequent values (deltas) in the TS, a tree format to identify where the values are in the TS, and metadata (e.g time stamp, sensor identifiers, and sensor precision), which give context to the data sent by our proposal. Additionally, we introduce Variable-based TS (VTS) and Measurements-based TS (MTS) templates to group the representation of TS by variables or measurements, respectively. Then, the performance evaluation of the proposed format shows that it can significantly reduce the amount of data sent by IoT devices by between 96% and 76% compared to JSON.

Furthermore, this paper concludes that the format proposed is well-suited for the representation of IoT data and has the potential to improve the efficiency of IoT systems. Deeper explanation details, a better understanding of the technical background, and other binary formats are available in the long version of this paper [ea22].

2 The Proposed Architecture

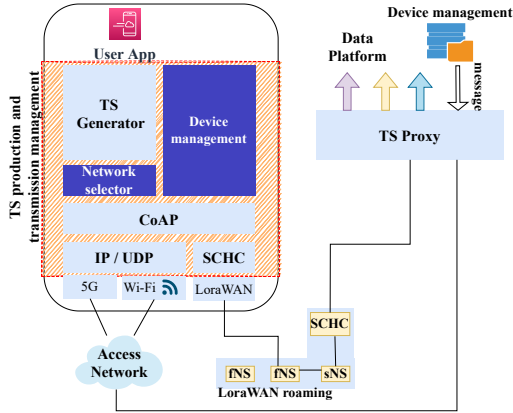


FIGURE 1 – The proposed IoT architecture.

Fig. 1 illustrates our vision of the IoT architecture on IoT networks. This architecture can be explained in three parts :

1. The user application will produce all the data retrieved from the measurements.
2. A middleware below the user application that will be in charge of (i) regrouping the TS data generated by the sensor in a compact manner with a common representation, and (ii) select the access network (Wi-Fi, 5G, LoRaWAN) to send the data depending on the network requirements.
3. The proxy, where all the data sent by the middleware will converge. This proxy will reconstruct the TS data format proposed in this paper to any data format needed by the data platform or cloud application.

3 TS representation with CBOR Templates

[BH18] defines a CBOR template as a CBOR data item containing one or more variables. These variables are represented as CBOR data item containing a specific identifier called a tag. Furthermore, this CBOR variable definition allows reducing the number of bytes that are sent by IoT devices by only sending the corresponding values and a further transformation into any data format needed by a cloud application.

[BH18] proposed a template, from now on called Static-based Time-Series (STS) template. To explain, consider that we want to represent the JavaScript Object Notation (JSON) structure presented in Listing 1, which has information related to a temperature and humidity sensor transmitting three measurements at the same time.

In this case, the corresponding STS template would be the one depicted in Listing 2. In this template, the variables are identified by the CBOR tag 42(). Nevertheless, we will refer to the CBOR variable as a tag number, Tag Number (TAGN). Finally, a CBOR array is required to undergo substitution and reconstruct the information, which would be : [32, 20, 31, 21, 30, 22]. However, STS is a static template, which means, if this template receives a different CBOR data item with a different amount of measurements (e.g., an array with only 2 measurements [32, 20]), then the template is expecting 6 variables and would not be able to make the substitution.

```
[{'n': 'temperature', 'u': 'Cel', 'v': 32},
 {'n': 'humidity', 'u': '%RH', 'v': 20},
 {'n': 'temperature', 'u': 'Cel', 'v': 31},
 {'n': 'humidity', 'u': '%RH', 'v': 21},
 {'n': 'temperature', 'u': 'Cel', 'v': 30},
 {'n': 'humidity', 'u': '%RH', 'v': 22}]
```

Listing 1 – A JSON structure that will be employed in the rest of the paper.

```
[{'n': 'temperature', 'u': 'Cel', 'v': 42(0)},
 {'n': 'humidity', 'u': '%RH', 'v': 42(1)},
 {'n': 'temperature', 'u': 'Cel', 'v': 42(2)},
 {'n': 'humidity', 'u': '%RH', 'v': 42(3)},
 {'n': 'temperature', 'u': 'Cel', 'v': 42(4)},
 {'n': 'humidity', 'u': '%RH', 'v': 42(5)}]
```

Listing 2 – Example of Listing 1 as a template with CBOR variables.

3.1 Tree Formatting and Delta Between Measurements

In this paper, we propose a tree representation inside the TAGN, indicating where the values of the variables will be found in the CBOR data item. This allows the introduction of non-static templates in a further section. Let us now consider $[[0, 1], [3, [4, 5]]]$, this array contains nested arrays in four different levels. Therefore, if it is necessary to select any value according to the new representation of the TAGN proposed, the following values can be selected :

1. The value **1** can be selected with the tag representation : **TAGN [0,1]**. It refers to the first array of data (TAGN [0,1]), in this case ([0,1]), and the second position of that array (TAGN [0,1])

2. It is possible to select the values (4 and 5) adding the value **'true'** to the TAGN representation : **TAGN [1, 1, true]**. It refers to the (TAGN [1,1,0]) and (TAGN [1,1,1])

Consequently, we introduce the use of deltas in CBOR templates. Here, we propose that the devices will transmit only the difference of the subsequent measurements, by doing that we reduce the amount of data sent by IoT devices. Then, the CBOR data item for the STS template required to reproduce Listing 1 is : [32, 20, 31, 21, 30, 22]. Thus, the difference present in each value for temperature is -1, and for humidity is +1. Therefore, the result array is : [32, 20, -1, 1, -1, 1]. Only 8 bytes are used against 10 from the previous array when using CBOR representation. Finally, in the rest of the paper, all examples will be expressed in deltas as described in this section.

3.2 Variable-based TS (VTS) and Measurements-based TS (MTS)

The STS representation helps to reduce the amount of data sent by IoT devices. However, it is limited by knowing the exact number of data items or variables to be replaced.

In this section, using the 'true' value in the TAGN, will allow differentiating when it is not possible to know how many values are going to be sent inside a CBOR data item. Thus, it is possible to represent TS data items with a template and an array of information as follows :

```
[{ 'n': 'temperature', 'u': 'Cel', 'v': TAGN [0, true] },  
{ 'n': 'humidity', 'u': '%RH', 'v': TAGN [1, true] }]
```

Listing 3 – VTS template.

Thus, continuing with the structure of the template proposed, the CBOR data item needed to represent the JSON structure presented in Listing 1 must be : [[32, -1, -1], [20, 1, 1]]

Consequently, this template allows us to send more or fewer values if needed thanks to the 'true' value. For example, it allows sending two values for humidity and two for temperature with the following CBOR data item [[32, -1], [20, 1]]. Thus, the CBOR data item will be accepted by the VTS template and this would be able to reconstruct the information. Another proposed approach is Measurements-based TS (MTS) template, as a way to order TS by measurements. Listing 4 depicts the MTS template to represent the structure in 1. Thus, the following CBOR data item is required to represent the JSON data values : [[32, 20], [-1, 1], [-1, 1]]. A detailed explanation about MTS is present in [ea22].

```
[{ 'n': 'temperature', 'u': 'Cel', 'v': TAGN [true, 0] },  
{ 'n': 'humidity', 'u': '%RH', 'v': TAGN [true, 1] }]
```

Listing 4 – MTS template.

3.3 Metadata

In addition to the actual measurement of the variable, some use-cases may need additional information. Thus, to represent this information we propose metadata. This metadata is represented by a map type ({}) in CBOR, i.e [{METADATA}, [32,20], [-1,1], [-1,1]]. Furthermore, with the metadata it is possible to represent the following values : (i) Base time which indicates, the time when the measurements were taken, it is represented with 'bt' e.g 'bt' :15964834, then (ii) the difference in time which indicates the time, in seconds, between the subsequent value in the CBOR data item, it is represented with 'dt' e.g 'dt' :30, and finally (iii) the base precision which indicates the precision of the measurements, it is represented with 'bp', e.g 'bp' :3.

4 Performance Evaluation

To evaluate the performance of our proposals, we first compare the JSON example presented in Listing 1 when transformed into CBOR, STS, MTS, and VTS, finally (ii) we take a real-world data set from [Jua22] and perform the same transformation.

1. **JSON representation for Listing 1.** : As shown in Figure. 2a, when Listing 1 is represented using CBOR requires 139 bytes. However, when represented with MTS or VTS is represented with 11 and 10 bytes, a reduction of 92% and 93% respectively. The STS template reduces the 88% but at the expense of a non-flexible template. All measurements and representations are present in [Jua22].

2. **Real-world data** : In this case, the real-world data is composed of the first 100 measurements of an electrophoresis painting plant with 7 sensors sensing every 10 s. This yields 700 measurements. Figure. 2b depicts a reduction of 77% (from 9294 to 2141 bytes) when using VTS template, and a reduction of 76% (from 9294 to 2228 bytes) when using MTS template. All measurements and representations are present in [Jua22].

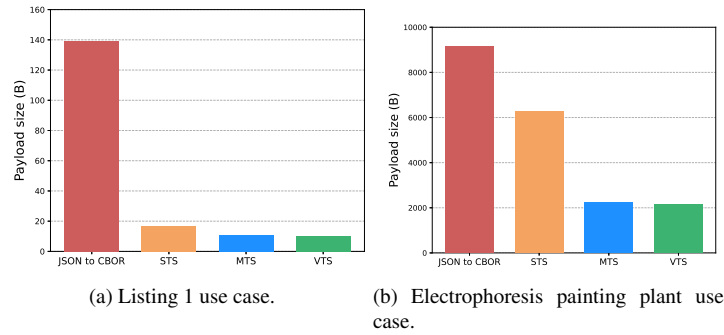


FIGURE 2 – Payload size per CBOR template.

5 Conclusions and Future Work

The VTS and MTS templates, as proposed, have shown a compact representation of data collected by IoT devices. Furthermore, their implementation results in a substantial reduction, ranging from 76% to 96%, of the total data transmission as compared to JSON.

The outcome of this proposal has practical applications : (i) improved battery life by reducing the size of the data transmitted, (ii) with a standard format to represent TS, it is possible to reduce the dependency between cloud applications and IoT devices by enabling transformation to any desired data representation.

As for future work, we plan to review the impact of the metadata on the performance of our templates using real deployments, and complex data structures.

Références

- [BH18] C. Bormann and P. Hoffman. Concise Binary Object Representation (CBOR) Tag for CBOR Templates. DRAFT 02, IETF, July 2018.
- [BH20] C. Bormann and P. Hoffman. Concise Binary Object Representation (CBOR). RFC 8949, IETF, December 2020.
- [DE] Cisco Dave Evans. The Internet of Things. How the Next Evolution of the Internet Is Changing Everything (white paper). Accessed : January 2022.
- [ea18] Davis Blalock et al. Sprintz : Time Series Compression for the Internet of Things. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 2(93) :1–23, 2018.
- [ea19] Kenda Klemen et al. Streaming Data Fusion for the Internet of Things. *MDPI AG Sensors*, 19(8) :1424–1450, 2019.
- [ea20] Andrew A. Cook et al. Anomaly Detection for IoT Time-Series Data : A Survey. *IEEE Internet of Things Journal*, 7(7) :6481–6494, 2020.
- [ea22] Molina A. Sebastian et al. Toward a standard time series representation for IoT based on CBOR templates. In *Global Information Infrastructure and Networking Symposium (GIIS)*, Sept 2022.
- [Ins] Fortune Business Insights. Internet of Things (IoT) Market Size, Share & COVID-19 Impact Analysis, By Component (Platform, Solution & Services), By End-Use Industry (BFSI, Retail et al), and Regional Forecast, 2021-2028. Accessed : March 2022.
- [Jua22] Molina A. Juan. Time series data set for cbor templates. <https://github.com/sebasmol96/Time-Series-data-set-for-CBOR-templates.git>, 2022.