The Ninth International Conference on Big Data, Small Data,
Linked Data and Open Data
ALLDATA2023
April 24, 2023 - April 28, 2023 - Venice, Italy

Software monitoring of an IoT chain communicating over LoRaWAN
Follow-up of planned data collection

Mohamed El kharroubi [1], and Fabrice Mourlin [2]

(1)

mohamed.el-kharroubi@u-pec.fr Algorithmic,
Complexity and Logic Laboratory
UPEC University
Creteil, France

(2)

fabrice.mourlin@u-pec.fr Algorithmic,
Complexity and Logic Laboratory
UPEC University
Creteil, France

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

A short Resume of the Presenters

Prof. Dr. **Fabrice Mourlin** is associate professor at Paris East University (Creteil) (UPEC). He is a researcher in LACL lab (Logic, Algorithm and Complexity lab). His main research topics deal with distributed systems, Big Data, AI models.
He is leader expert at the ministry of higher education and research for computer science domain. He collaborated with researchers from abroad lab like NIST (Maryland), Siemens (New Jersey)

Mohamed El kharroubi is a PhD student preparing a thesis at the University of Paris-Est Créteil (UPEC), whose subject is the resilience of re-inforcement learning algorithms. He is also a project manager specialized in middleware and distributed systems working in private companies.

Outlook

- Introduction
- Related Work
- Our Use Case
- Physical deployment diagram of authentication system
- System Architecture : Component Diagram Structure
- Big Data Workflow
- Analysis Results : Data Visualization
- Analysis Results : Streaming Configuration
- Conclusion and future works

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Introduction

Monitoring an Internet of Things (IoT) considerations :

- Involves necessarily data collection phase.
- IOT Systems usually depend on a complex IT deployment System.
- The more complex is the deployment architecture (usually this is the case), the more complex is the monitoring the IOT.
- The understanding of collected information is a key point, but it can be impacted by several factors :
- Malicious interventions that harms the applications and the monitoring also.

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Introduction

Monitoring an Internet of Things (IoT) System considerations :

- Deployments into local or remote clouds prolong the impact of these biases and harm the observations on the long term.
- Collecting data on a widely distributed system so from many origins causes a lot of data interpretation issues. How can we avoid or limit these issues ?
- The use of log message schema is a step forward to apply aggregation operations on messages of different origins.
- Avoid imposing data schemas remains too strong a constraint to respect when collecting data on a widely distributed system.

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Related Work

Lv et al.[1] work : Designed a monitoring system for an urban environment that relies on a dedicated communication network to complement on-the-fly data collection

Contribution : An observer station is identified to ensure live monitoring with the possibility to react in case of unexpected events.

Limitations : It remains precarious especially in the case of long duration simulation. Events can occur in a cascade mode and it could be difficult to have a suited reaction.

W. T. Hartman et al.[2] work : Designed a System for building, testing, and implementing a low-cost energy monitoring and control system using IoT devices

Contribution : monitoring mixed with business code.

Limitations : need for structured monitoring broken down into monitoring layers.

[1] Z. Lv, B. Hu and H. Lv, "Infrastructure monitoring and operation for smart cities based on IoT system. IEEE Transactions on Industrial Informatics", vol.16 no.3, pp. 1957-1962, 2019.

[2] W. T. Hartman, A. Hansen, E. Vasquez, S. El-Tawab, and K. Altaii, "Energy monitoring and control using Internet of Things (IoT) system. In 2018 Systems and Information Engineering Design Symposium (SIEDS)", pp. 13-18, IEEE, 2018

Related Work

S. Nocheski et al.[3] : Monitoring to maintain the sustainable habitat environment for certain fish species inside fishing ponds through distributed machine-to-machine communication

Contribution : - The system reduces the chain of responsibility by measuring many biological factors. -
- The Wivity modem allows the user to communicate with the IoT system through Wi-Fi, cellular, Long Range Wide Area Network (Lora WAN) or satellite communication, all in one product.

Limitations : There is a need to configure the monitoring by event family.

G. Yang et al.[4] work : Designed a wearable device with a bio sensing face mask to monitor a patient's pain intensity using facial surface electromyogram (sEMG)

Contribution : -The wearable device is integrated with the Internet for remote pain monitoring. It transmits data to the server via an http gateway.
- The collected data are stored in a cloud.

Limitations : Security constraints as monitoring data is exposed remotely on the cloud.

[3] S. Nocheski and A. Naumoski. "Water monitoring iot system for fish farming ponds. Industry 4.0", 2018.

[4] G. Yang, M. Jiang, W. Ouyang, G. Ji, H. Xie, A. M. Rahmani and H. Tenhunen, "IoT-based remote pain monitoring system: From device to cloud platform. IEEE journal of biomedical and health informatics", vol.22 no.6, pp. 1711-1719, 2017.

Related Work

K. Sabanci et al.[5] work : Designed an electronic device to identify people in a store and count them by calculating their path through an enclosed space

Contribution : - Mobile clients running on a nomadic terminal can read in a shared cloud the paths of people to monitor their theme of interest and distinguish between stationary and mobile people .

Limitations : - Shared access introduces security issues that must be managed or prohibited.

J. Wang et al.[6] work : Control of the greenhouse environment. The acquisition and control parameters, network protocols are different for various greenhouse feedback.

Contribution : - An XML-based data encapsulation method (with XML data schema) was designed to enable data interoperability in a distributed greenhouse IoT system

Limitations : - The concept is to be propagated to all data collection phases with JSON, YAML, etc.

[5] K. Sabancı, E. Yigit, D. Üstün, A. Toktaş and Y. Çelik, "Thingspeak based monitoring IoT system for counting people in a library. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP)", pp. 1-6, IEEE, 2018.

[6] J. Wang, M. Chen, J. Zhou and P. Li, "Data communication mechanism for greenhouse environment monitoring and control: An agent-based IoT system. Information Processing in Agriculture", vol.7 no.3, pp. 214-455, 2020.

Related Work

T. M. Bandara et al.[7] work : make a new contribution with a large area data collection where wireless sensors are accurately distributed over an agricultural area

Contribution : - The notion of data schema is introduced in a format specific to the project but it allows the unification of representations.
- The use of schema brings the validation of data before its use.

Limitations : - The use of specific data transmission is an asset for the separation of concepts.

D. Davcev et al.[8] work : Studied Agricultural monitoring using a low-energy cost Long Range Wide Area Network (LoRa WAN) protocol for data communication

Contribution : - As the message body is encrypted, it increases security against network intrusions.

Limitations : - Introduces an additional risk because new properties appear such that the message body is encrypted, which means an additional cost for the processing .

[7] T. M. Bandara, W. Mudiyanselage and M. Raza, "Smart farm and monitoring system for measuring the Environmental condition using wireless sensor network-IOT Technology in farming. In 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA) " , pp. 1-7, IEEE, 2020.

[8]D. Davcev, K. Mitreski, S. Trajkovic, V. Nikolovski and N. Koteli, "IoT agriculture system based on LoRaWAN. In 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS) ", pp. 1-4, IEEE, 2018.

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Our Use Case

- Based on the use of authentication cards that all the people of the university have in order to enter the different premises of our site.
- Card is nominative and gives access to certain premises are accessible and not others.
- The same cards allow also to use classic and 3D printers in different rooms (automatic printing output to the right room printers) .
- The authentication card can be used in two ways: simple contact with a sensor or insertion of the card in a dedicated reader.
- Insertion of the card in a dedicated reader is necessary to use 3D printer or long time printing tasks

⇒ It is very frequent that a user forgets his card when taking the documents out of the printer.

⇒ Many users discover that their card is missing the next time they use it.

What the solution can be ?

- we asked for the design of a LoRa WAN cardholder with adapted monitoring system.

Our Use Case

Why LoRa WAN cardholder ?

- The use of such a protocol is explained by the autonomy of the cardholder in energy.
- It’s a low cost radio protocol that we master in the laboratory,.
- A LoRa WAN antenna on the site and an IP/LoRa WAN gateway per floor.
- It's protocol that ensures that the body of the messages is encrypted and thus ensures better security of the collected data.

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Our Use Case

How the solution should work ? Several phases of monitoring data collection

A Cardholder has a sensor to detect the presence or absence of card.

⇒ When the card is absent, it emits a LoRa WAN message to indicate the absent card and the identifier of the place of emission (Kept in cardholder memory)

⇒ It will be sent again one hour later with an increment until the card is back to his holder.

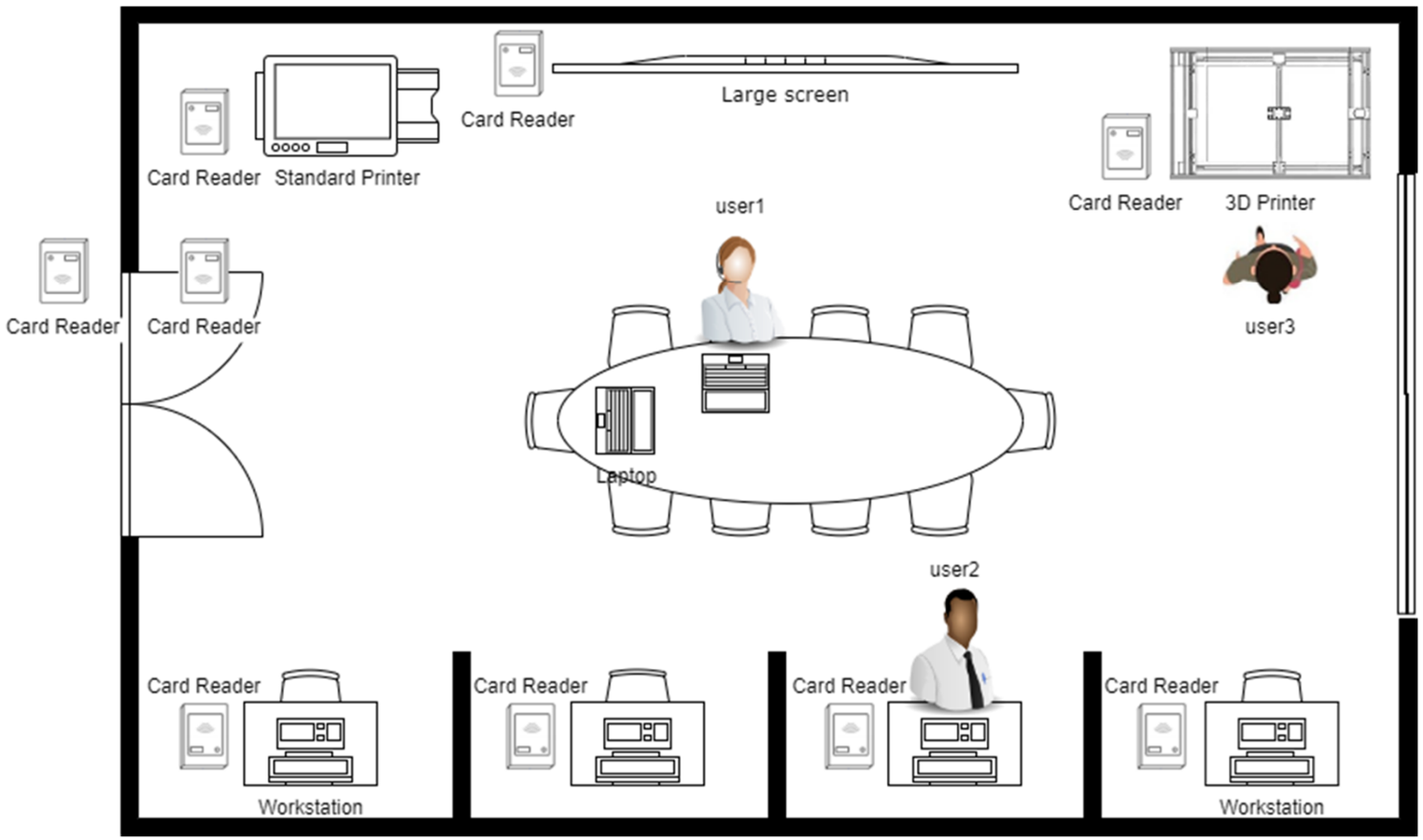
⇒ Messages collected by the network monitoring show the loss of a card.

⇒ A script is performed to send a message to the person concerned.

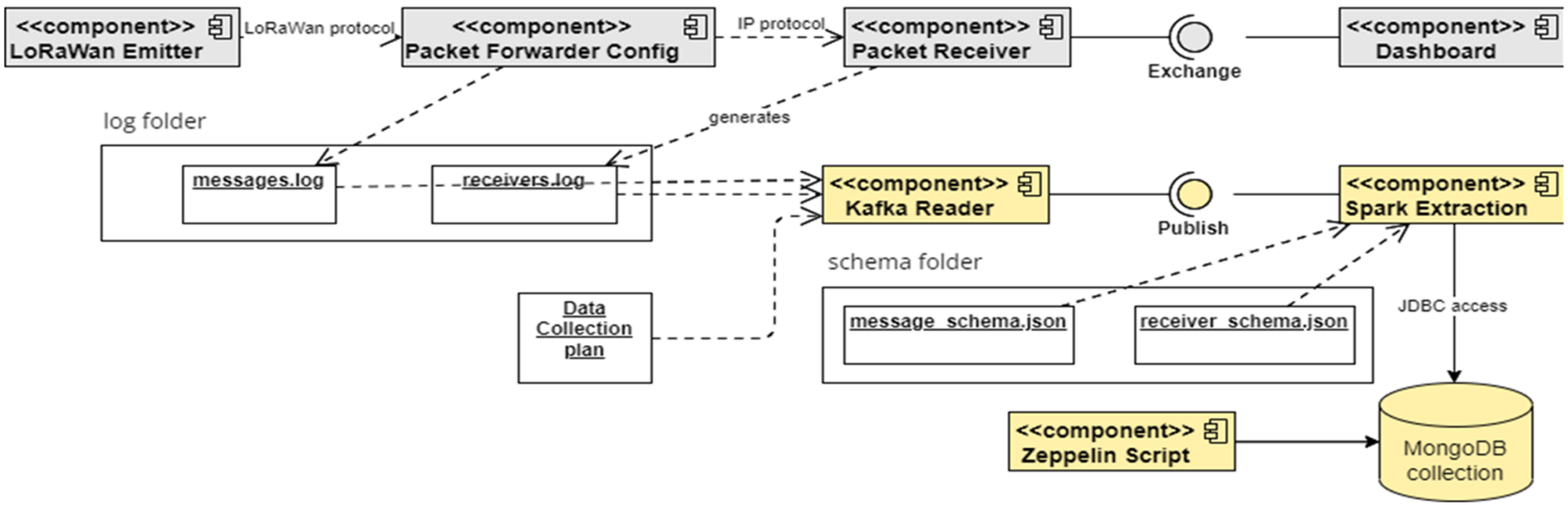
⇒ Big Data batch extracts the useful information to insert it in a NoSQL database.

⇒ A report on the activity of this database is published for the security department.

Physical deployment diagram of authentication system



System Architecture : Component Diagram Structure



Grays Components : - belong to the user authentication application within the laboratory.

Yellow Components : - Belong to the badge monitoring application and includes a data collection phase.
- Followed by information routing using the Kafka framework and a Big Data analysis component based on the Apache Spark framework .

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Big Data Workflow

Based on sampling data, extracting information, and then enriching this data before saving it in a persistence system.

Kafka Reader Component

- The objective of this component is to deliver, analyze and index millions of log messages from professional badge monitoring activity over LoRaWan protocol.
- Kafka Reader can process every event without being overwhelmed and without putting pressure on the log message producers.
- Log message producers send events to Kafka whose content respects a data schema, which divides them evenly into partitions.
- Defined a message schema for each sending source.
- Each partition is an ordered queue of events assigned to our component named Spark Extractor.
- In case of damaged partition, Kafka automatically switches to a replica, which provides a fault tolerance system without us having to implement it.

Big Data Workflow

Based on sampling data, extracting information, and then enriching this data before saving it in a persistence system.

Spark Extractor Component

- Spark component determines the data schema to apply in order to validate the data.
- Using of JSON Path queries allows us to extract the data we want to keep in a document inserted in MongoDB.
- It's protocol that ensures that the body of the messages is encrypted and thus ensures better security of the collected data.
- The role of the data schema is essential because it is used for our indexing process of the log messages before insertion.
- If the data schema evolves over time, its externalization to the Spark Extractor component ensures an adaptation of the treatments performed on the analyzed log messages.

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Big Data Workflow

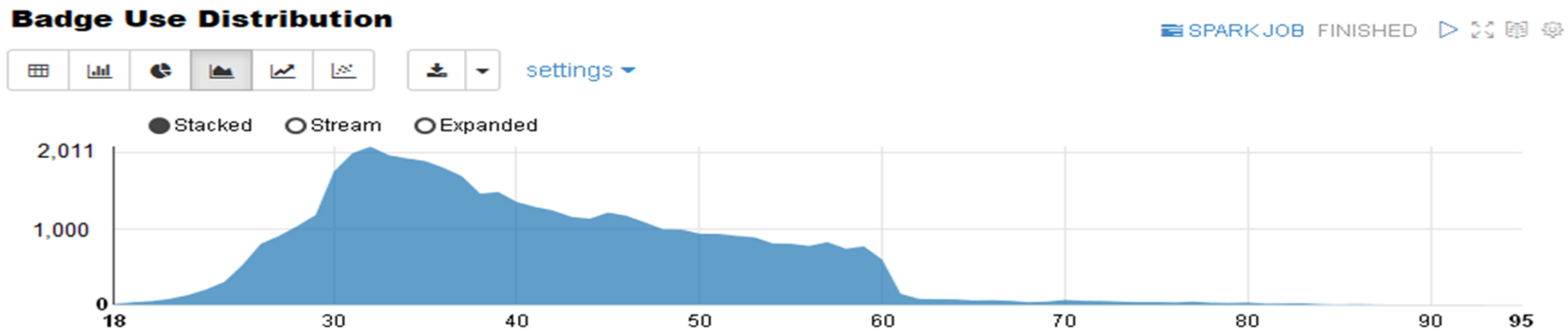
Based on sampling data, extracting information, and then enriching this data before saving it in a persistence system.

Spark Extractor Component

- MongoDB is used because it allows us to combine and store multivariate data.
- From the data frames built in memory with Spark Engine, we save each of them via the mongo-spark-connector.
- data schemas are used to parse the value associated with the message key, enrich the JSON document and add new keys/values not present in the original document.
- Many messages are issued to track the monitoring behavior of business badges.
- The collected information is aggregated to provide richer messages for the observer, even if they don’t have the same origin.
- Aggregated messages offer a follow-up of the badges uses overall laboratory.

UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE

Analysis Results : Data Visualization

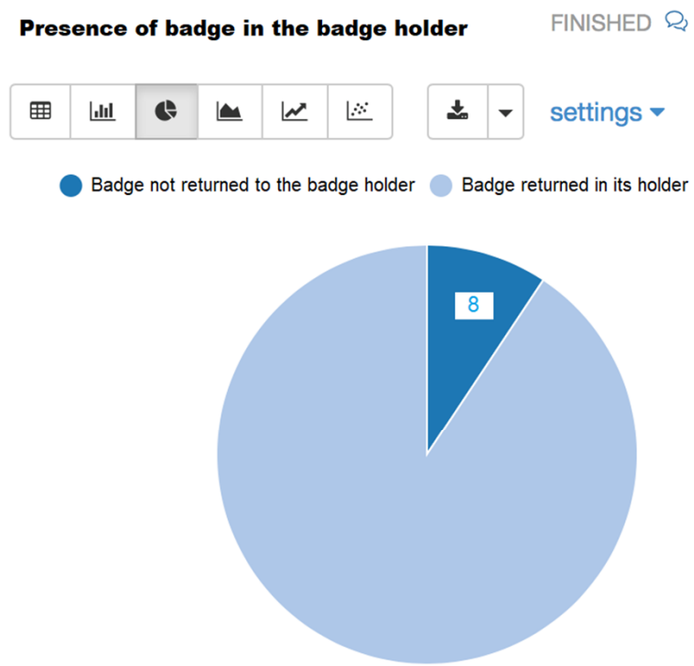


Badge usage over a 4-week period. On the x-axis is the number of operations while on the y-axis is the time of use of these badges in reading

- SQL and scala scripts with the Zeppelin tool to access the MongoDB database.

- The SQL queries are primarily useful to extract information and then use the Zeppelin graphical library to visualize the results

Analysis Results : Data Visualization



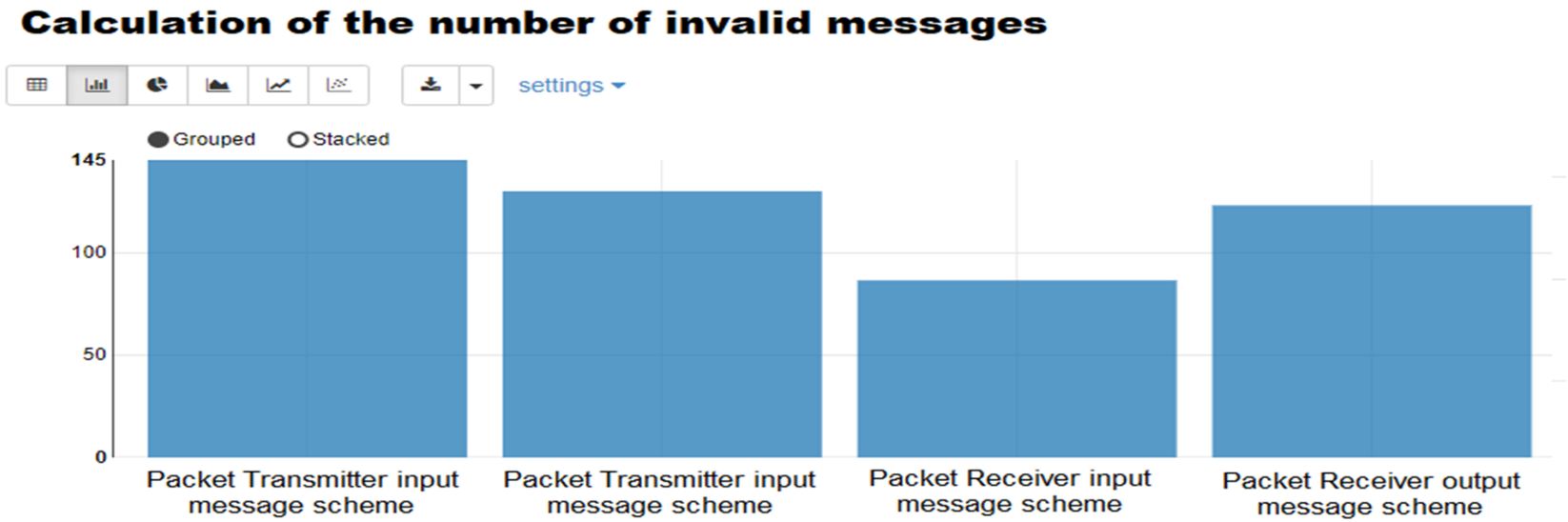
- Once the query is validated, it is easy to include it in a Scala script.

- As the Scala script is compiled means that the query is not interpreted for each data set but its evaluation is reused for all the date sets, for each sampling window.

- Zookeeper tool in conjunction with Spark because it allows more flexibility between data engineers and developers.

Calculation of the number of badges not returned to the badge holder over a period

Analysis Results : Streaming Configuration



1. Use of the streaming parameters

- Our experiments allowed us to externalize in the configuration of the Spark Extractor component several useful coefficients to manage data sampling.

- An adapted configuration allows a better exploitation of the computing resources in particular on clusters where the computing time is paying.

- From results, we can illustrate with these parameters comparisons between several types of configurations .

UPEC
UNIVERSITÉ
PARIS-EST CRÉTEIL
VAL DE MARNE
Connaissance - Action

IARIA

Conclusion and future works

- We presented our work around the planning of data collection and analysis.
- We illustrated our work with the analysis of log messages from LoRa WAN transmission tools.
- We have shown that the use of data schema is an asset to have more accurate collection results.
- We have built an efficient kafka component that's performing data collection. Its configuration allows us to define our data collection roadmap.
- We have built a streaming component that applies vetro pattern : Validation, Enrich, Transform, Route & Operate.
- As future work, we plan to improve our maintenance policy based on message collections and the monitoring we have set.