PFilter: Privacy-Aware and Secure Data Filtering at the Edge for Distributed Edge Analytics

Annanda Rath, Anna Hristoskova, Sarah Klein

PFilter: Privacy-aware and secure data filtering at the edge for distributed edge analytics

Annanda Rath ✉, Anna Hristoskova, and Sarah Klein

Software Engineering & Security Department, Sirris, Brussels, Belgium
{annanda.rath, anna.hristoskova, sarah.klein}@sirris.be
www.sirris.be

Abstract. This paper is presenting a conceptual mechanism for lightweight privacy-aware and secure data access control and filtering. This mechanism can be deployed at an edge node in order to assure that all data coming in and going out of it is properly protected and filtered. Goal is to keep private data locally and limit its exposure to outside entities (e.g., Cloud backend, external application or other edge nodes) while preserving the performance and security requirements for edge analytics. The data filtering at the edge node is done in a way that it is not possible for outside entities to identify end-devices and the data associated with them.

Keywords: Privacy filter · Distributed edge analytics · Security · IoT.

1 Introduction

The Internet of Things (IoT) is growing exponentially with connected devices ranging from smart doors to industrial machines and installations. By 2025, it is expected that there will be 41.6 billion connected IoT devices, generating 79.4 zettabytes (ZB) of data[1]. With such amount of data, the current practice of sending data to central platforms (e.g., Cloud) for analysis becomes unsustainable as it does not scale well and represents a single point of failure. Given that, the new paradigm of Edge Computing serves a variety of purposes in the current IoT landscape. This distributed, local computing paradigm can help addressing several issues from latency, bandwidth, connectivity, security and privacy issues that would otherwise make some IoT cases impossible. Data collection, machine learning and AI applications can be distributed at the edge in a federated architecture. Such a solution is robust because tasks can migrate in case of component failures and is scalable since workload can be shared among many computing devices. However, there are also many challenges [2] in federated architectures. One of which is the security of edge analytics as it is no longer a single point of concern as in a centralised architecture, but distributed over the federated approach. Privacy-sensitive edge analytics is also a challenging security issue [2],

[1] https://www.zdnet.com/article/iot-devices-to-generate-79-4zb-of-data-in-2025-says-idc/

especially, when private data needs to be shared between edge nodes[2] for aggregation, analysis and modelling purposes. Questions arise such as: How to share minimal information without losing its usability? How to control and enforce the use of data in analytics processes, resulting in transparency? This becomes even more challenging in the IoT context since edge nodes tend to have limited computing power and memory where traditional data filtering and minimisation [1] do not fit. Therefore, a lightweight privacy-aware and secure solution is needed. This paper addresses the data security and privacy-preserving issues in the distributed edge computing taking into account the requirements and security constraints in an IoT edge analytics environment. We plan to investigate the following aspects:

1. A lightweight & configurable data filtering mechanism that supports different types of data with different levels of privacy protection requirements.
2. A lightweight access control and a coordinated enforcement mechanism that controls all aspects of data access on the edge node, between edge nodes or between edge nodes and end devices.

This paper is structured as follows. Section 2 highlights the need for privacy-aware and secure data filter and access control mechanism. Section 3 focuses on the description of the industrial cases from different companies within the EUREKA-ITEA3 project MIRAI[3] having similar concerns to what we are envisaging to address. Finally, Section 4 provides concluding remarks.

2 The need of data filtering, access and usage control for data analytics at the edge

The key challenge of edge analytics is the hardware constraint. Edge nodes tend to have limited computing footprint. This is why embedding the functionalities of a fully-fledged data analytics mechanism and heavy access control and data filtering process, similar to the ones implemented in the centralised architecture [3], is challenging. This becomes even more challenging in the context of battery-powered IoT edge nodes, that require to run for a prolonged lifespan with minimal intervention. Therefore, a lightweight solution is needed.

2.1 Need of lightweight data filtering and control on filtering process

Edge computing can reduce the number of sensors and actuators connected directly to the Internet. As a result, reducing number of connections to the Internet. This reduces the potential attack vector of security attacks. Local data processing and filtering by an edge node (gateway) can also reduce the amount of

[2] In this paper, edge node refers to device gateway (or IoT gateway), while end-device refers to IoT devices, such as sensors/actuators.

[3] https://itea3.org/project/mirai.html

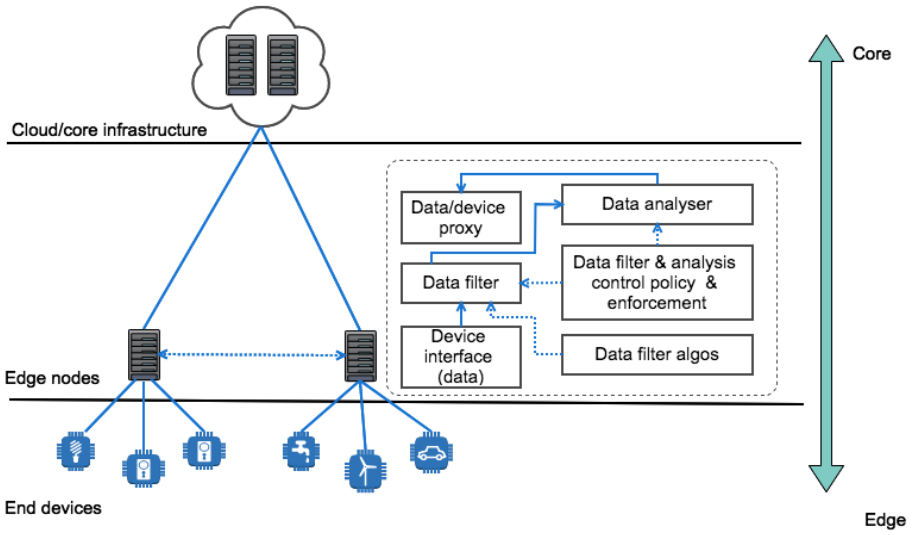

Fig. 1. High level architecture of edge node system components

sensitive and private information that is sent through a network. Thereby, it addresses privacy needs for the application. However, the data filtering mechanism [1][5][6][7][8] that requires heavy processing and computing power is not suitable for resource-constrained edge nodes. There is a need for a lightweight data filtering mechanism that is able to run on devices with very limited computing power and memory without compromising data usability and privacy requirements. Our goal is to investigate a new lightweight data filtering approach that can run on very resource-constrained edge nodes, taking into account the following three factors during its design: (1) minimal processing and less computing, (2) privacy-compliance and zero data leaks, and (3) avoidance of information loss and guaranteed data usability.

In addition, we also envisage to add a security layer on top of the data filtering process at the edge nodes by means of a security control on the data filtering process (see Fig 1, data filter control policy). This data filter control policy module (see Fig 1) regulates what and how the data filtering module should process different types of data with different filtering requirements. It also instructs the data filter module which data filter algorithms should be selected. The data filter module outputs the non privacy-sensitive data, to the data analyser where the AI or machine learning is located (see Fig 1). The data analyser processes and extracts relevant data and outputs to the data proxy (see Fig 1), which makes the data available to the outside entities (see Fig 1). The data proxy acts as a secure gateway for the exchange of data between the edge environment and core network (see Fig 1) and also isolates the edge environment from the core network.

2.2 Need for data access and usage control and its enforcement for distributed edge analytics

In the case of distributed edge analytics, part of the data can be shared between edge nodes. In order to ensure a fine-grained data access control for such data sharing, a reliable data access control and enforcement is required. Moreover, to address privacy concerns in such data-sharing scenarios, a privacy-preserving method needs to be incorporated in the access control and enforcement scheme. This access control mechanism should be lightweight and be able to run on very resource-constrained edge devices while also being able to meet the security and privacy requirements. As of today, most access control systems, such as OAuth, OpenID, are running on high performance devices with complex processing steps. There are many attempts to extend OAuth to be used on resource-constrained devices such as in a IoT system [4] [9]. However, looking into their processing steps that involve heavy encryption algorithms, they are still considered as heavy schemes. Our goal is to look into a new lightweight access control and enforcement mechanism for data sharing in distributed edge analytics for very resource-constrained devices, devices with limited power and memory and that require long battery lifespan. We take three following factors into account when designing it: (1) fine-grained and privacy-aware, (2) low computing power and (3) few processing steps.

3 Industrial cases: sustainability and smart city domain

In this section, we provide a description of actual industrial cases where private data access control and filtering on resource-constrained devices are critical ongoing challenges. These cases are from three companies, Macq, Shayp and 3E[4] within the EUREKA-ITEA3 project MIRAI. They will be used as test cases for our envisaged data filter and access control demonstrator.

3.1 3E case and requirements

3E is a Brussels-based company providing consultancy and software solutions for monitoring and improving the performance of sustainable energy installations and for optimising energy consumption. More precisely, 3E SynaptiQ[5] solution enables to manage solar and wind assets, monitor, report and improve their performance, and organise their maintenance. In order to further improve their platform, 3E works on a use case in the domain of distributed renewable energy systems, linked to the following research area: (1) Leveraging the computational power of edge devices in order to reduce latency and (2) tackle both data sharing obstacles and privacy aspects. These research areas relate to the challenges we are addressing in this paper.

[4] https://mobility.macq.eu, https://www.shayp.com, https://3e.eu
[5] https://3e.eu/our-platform

3.2 Shayp case and requirements

Shayp is a Brussels-based company providing solutions for the monitoring and management of water usage and leakages for different types of buildings, ranging from home to large industrial buildings, such as public and business buildings. Shayp collects water consumption information through a connected water metering device installed in those buildings. Water measurements are collected periodically and are sent to the cloud through Sigfox and NB-IoT. From this data, Shayp extracts water consumption analytics and detects potential leakages present in a building. Shayp is currently pursuing several tracks for the further enhancement of their services, among which is privacy-preservation of the data on device (water metering). Shayp aims at securing and filtering data in transit in order to prevent disclosing privacy-sensitive water consumption data. This research links to one of our research goals in this paper.

3.3 Macq case and requirements

Macq is a Brussels-based company providing solutions around two main areas: industrial automation and smart mobility. While the former represents the historical market targeted by Macq, smart mobility has become the largest activity domain of the company in the last years and drives many of Macq research activities. Under EUREKA-ITEA3, Macq focuses on the smart mobility where traffic is the main use case. Macq aims at improving the safety of vulnerable road users in the area of railway crossings, school streets and at complex intersections. For this, intelligent street cameras are installed. Macq works towards a solution in the domain of road safety, linked to the different research areas, among which are: (1) efficiency of edge processing without the need to send images to the cloud and (2) private data filtering of images captured by street cameras. These two research cases link to what we are addressing in this paper.

4 Conclusion

This research position paper highlights the potential of data security and privacy-preserving challenges at the edge. We also present some early concepts on how privacy and data filtering at edge nodes should be designed and explain why existing mechanisms do not fit. Our next steps are to extensively work on the above-mentioned concepts and validate them through a demonstrator based on the requirements from the three industrial cases (Macq, 3E and Shayp).

References

1. Abigail G., Gilad E., Ron Sh., Micha M., and Ariel F. Data Minimization for GDPR Compliance in Machine Learning Models. IBM Research - Haifa, Haifa University Campus, Haifa, Israel.
2. W. Shi, J. Cao, Q. Zhang and Y. Li and L. Xu. Edge Computing: Vision and Challenges. IEEE IoT Journal, 2016, volume 3, number 5, pp=637-646.

3. Suzan Al., Nusaybah A. and Muhammad M. Survey of Centralized and Decentralized Access Control Models in Cloud Computing. International Journal of Advanced Computer Science and Applications, 2021, volume 12, number 2.
4. A. Karim and M. A. Adnan, "An OpenID Based Authentication Service Mechanisms for Internet of Things," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 687-692.
5. J. Zhang, B. Chen, Y. Zhao, X. Cheng and F. Hu. Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues, in IEEE Access, vol. 6, pp. 18209-18237, 2018.
6. R. Lu, K. Heung, A. H. Lashkari and A. A. Ghorbani, "A Lightweight Privacy-Preserving Data Aggregation Scheme for Fog Computing-Enhanced IoT," in IEEE Access, vol. 5, pp. 3302-3312, 2017, doi: 10.1109/ACCESS.2017.2677520.
7. Tangade, S.; Manvi, S.S. Scalable and privacy-preserving authentication protocol for secure vehicular communications. In Proceedings of the 2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Bangalore, India, 6 November 2016.
8. Chim, T.W.; Yiu, S.M.; Li, V.O.; Hui, L.C.; Zhong, J. PRGA: Privacy-preserving recording gateway-assisted authentication of power usage information for smart grid. IEEE Trans. Dependable Secur. Comput. 2015, 12, 857.
9. Mahmood, K.; Chaudhry, S.A.; Naqvi, H.; Shon, T.; Ahmad, H.F. A lightweight message authentication scheme for Smart Grid communications in power sector. Comput. Electr. Eng. 2016, 52, 11424.