Securing IoE Environments with Semantic Data Stream Analysis and Behavioral Fingerprinting

Securing IoE Environments with Semantic Data Stream Analysis and Behavioral Fingerprinting

Marco Arazzi
DIII, Università di Pavia
via A. Ferrata 5, 27100, Pavia, Italy
marco.arazzi01@universitadipavia.it

Monica Marconi Sciarroni
DII, Università Politecnica delle Marche
via Brecce Bianche, 60131 Ancona, Italy
monica.marconi@staff.univpm.it

Serena Nicolazzo
DISIT, Università del Piemonte Orientale
V.le T. Michel, 11, 15121 Alessandria, Italy
serena.nicolazzo@uniupo.it

Antonino Nocera
DIII, Università di Pavia
via A. Ferrata 5, 27100, Pavia, Italy
antonino.nocera@unipv.it

Emanuele Storti
DII, Università Politecnica delle Marche
via Brecce Bianche, 60131 Ancona, Italy
e.storti@univpm.it

***Abstract*—In the landscape of Industry 5.0, Internet of Everything (IoE) networks are emerging as crucial components for connecting diverse industrial sensors and devices, expanding beyond traditional IoT boundaries to integrate people, processes, and data. However, this increased connectivity raises significant security concerns, as the growing complexity of IoE environments introduces new attack vectors and privacy risks. Additionally, the integration of heterogeneous devices and data sources presents both technical and semantic interoperability challenges, requiring robust mechanisms for meaningful data interpretation and secure exchange. This paper, developed within the HOMEY project, presents an architecture for gathering and monitoring semantic data streams in IoE environments, addressing both interoperability and security challenges. Our approach leverages Knowledge Graphs to represent sensor metadata, locations, access rights, and operational contexts, enabling dynamic stream monitoring and data querying. An approach based on Federated Learning allows distributed behavioral fingerprinting of IoE devices, which is exploited on top of the platform to perform anomaly detection from real-time data streams. The approach enhances reliable, privacy-preserving anomaly detection, contributing to the security and resilience of next-generation industrial IoE ecosystems.**

***Index Terms*—Stream management, Big Data Applications, Anomaly detection, Knowledge Graphs, Industry 5.0, Internet of Everything, IoT**

DOI: 10.1109/ETFA65518.2025.11205645 .

This work has been supported by the PRIN 2022 Project "HOMEY: a Human-centric IoE-based Framework for Supporting the Transition Towards Industry 5.0" funded by the European Union - Next Generation EU, Mission 4 Component 1 (code: 2022NX7WKE, CUP: F53D23004340006).

I. Introduction

Industry 5.0 represents the next phase in industrial evolution, building on the advancements of Industry 4.0. It emphasizes the integration of human intelligence with machine automation to create more efficient, sustainable, and human-centric manufacturing processes. In this context, the novel paradigm of the Internet of Everything (IoE, hereafter) is an evolution of the Internet of Things (IoT) that integrates not only connected devices but also people, data, and processes into a unified intelligent network [1]. The ability to efficiently monitor data streams from various sensors, machines, and devices deployed in industrial environments is central to this transformation. However, the dynamic nature and scale of data in this interconnected environment present peculiar challenges [2] such as scalability, real-time processing, data quality, anomaly detection, and secure data access, which necessitate innovative approaches for real-time analysis and decision-making. In particular, IoE generates vast amounts of real-time streaming data from heterogeneous sources and diverse applications (e.g., smart homes, healthcare, industrial IoT), thus leading to interoperability issues. Additionally, much of the data generated by the IoE is sensitive (e.g., production data, location of people and devices, cameras) making it a prime target for data breaches. Moreover, also the sheer volume of interconnected devices increases potential entry points for cyberattacks [3], further complicating security efforts.

To overcome these issues, in this discussion paper, we introduce a novel and complete framework for *(i)* collecting and managing data streams in an IoE ecosystem and *(ii)* monitoring the behavior of entities interacting in the environment to detect possible anomalies. As for the first objective, the ability to integrate heterogeneous data from multiple sources is essential for achieving full integration and supporting advanced tasks in this industrial context. A crucial step towards this goal is achieving technical and semantic interoperability. Technical interoperability focuses on network protocols, interfaces, and operating standards, while semantic interoperability aims to

define a higher abstraction layer that supports meaningful data exchange and interaction. This is typically achieved through shared reference data models, such as ontologies or vocabularies, which provide mutually understood meanings for entities in the IoE. In this work, we discuss an architecture for semantic interoperability in an IoE scenario, based on a Knowledge Graph, on top of which we developed a Semantic Stream Platform for gathering and monitoring data streams. As for the second objective, data streams gathered by the platform are monitored in order to identify all interactions among IoE entities. These interactions, along with network parameters included in the exchanges, can be exploited to draw a normal behavior of an entity, also known as Behavioral Fingerprint (BF) [4], [5]. Borrowing some ideas from recent approaches [6] we leverage BF to build behavioral models representing the expected conduct of target entities in the network at the starting phase of our framework, and monitoring activity to detect possible variations during the normal network functioning. Our approach utilizes Federated Learning (FL), an advanced Machine Learning strategy that enriches the fingerprinting model by incorporating insights from multiple interactions across different nodes. Additionally, it effectively addresses security and privacy challenges associated with data exchange among the participating entities.

This article, developed in the context of the HOMEY project, consolidates previous work [4], [6]–[8] by providing an overall view of the platform. Its main contributions are the following:

- we present a complete architecture for semantic data stream gathering and monitoring in IoE;
- we employ Knowledge Graphs to model sensor metadata, locations, access rights, and operational contexts to facilitate dynamic data querying and dynamic stream monitoring;
- we leverage the distributed Behavioral Fingerprinting approach to profile IoE devices and detect anomalies in the network.

The rest of this work is structured as follows. Section II surveys relevant literature on IoE platforms and behavioral fingerprinting for anomaly detection. The general methodology is introduced in Section III. In Section IV we present the knowledge model of the platform, while the architecture for data gathering and monitoring is introduced in Section V altogether with a discussion on the employed anomaly detection approach. Finally, Section VI concludes the article and discusses future work.

II. Related Work

This section is related to the description of the existing approaches related to *(i)* management of the semantic data stream in IoE; *(ii)* Behavioral Fingerprinting for Anomaly Detection.

A. Semantic Data Management in IoE

IoE infrastructures require adaptive and interoperable architectures with advanced data storage and management models to handle complex and different systems. Research has focused on hierarchical architectures with layers for specific functions like data collection, processing, and storage. Various approaches include three-level storage models for Industry 4.0 [9], edge-fog architectures for Smart Cities [10], and AI-enhanced systems for data analysis [11].

Knowledge Graphs (KGs) are increasingly important for IoT data management and analysis [12]. In industrial applications, KGs are often adopted as a solution for addressing semantic inconsistencies in heterogeneous datasets [13]. Some approaches [14] use semantic gateways between sensors and cloud services to add semantic annotations to raw data. Ontological models can improve MQTT-based communications by addressing its limitations through semantic enrichment. Approaches like "Semantic Subscription" [15] exploits ontology-based search to find optimal topic matches, while others [16] automate industrial data subscription for new sensors, or allow semantic data analytics with a process-aware approach [17]. Most existing solutions focus on traditional IoT scenarios with sensors as main data publishers. However, Industry 5.0 introduces more complex ecosystems that integrate smart objects, people, and processes with heterogeneous protocols and schemas.

B. Behavioral Fingerprint for Anomaly Detection

Recent countermeasures to IoE threats are increasingly involving Machine Learning and Deep Learning techniques [18]. In this context, a novel trend is to model peculiar characteristics of target entities to detect compromised devices within a network. The set of these features possessed by an IoE actor and composing its behavior when it interacts with other entities in the environment represents its *fingerprint*. Traditional device fingerprinting encompasses soft identities, including device name, type, manufacturer details, serial number, network address, and other attributes extracted from various networking data sources. For instance, the authors of [19] identified 19 features that can be used to estimate the security level of an object directly from the data-link header of 802.11 messages. The authors of [20] instead comprise in the fingerprint the physical aspects of devices, like inter-arrival times of different packets.

Behavioral Fingerprint (BF) represents an evolution of such a method including some features that cannot be easily cloned by a malicious adversary [4], [5], [21], [22]. In particular, this type of technique exploits application-level information to extract features related to the interaction among the entities of the system and, hence, their networking behavior. In particular, in [21] the authors leverage a number of features extracted from the network traffic of the device to train an ML model to detect similar device types. Celdrán et al. [22] examine a detection framework that applies device BF and ML to detect anomalies and classify different threats (e.g., botnets, rootkits, backdoors, and ransomware). The framework presented in [5] describes an enhanced BF model consisting of a fully decentralized scenario, where it is possible to exploit the

features derived from both the analysis of packet payloads and message content.

III. Methodology

In this work, we refer to an Industry 5.0 scenario in which human agents coexist and seamlessly collaborate with smart objects possibly provided with autonomous reasoning. For the technology layer producing data, we refer to the notion of the Internet of Everything (IoE) network, which encompasses all data-producing systems in the environment, including both simple or complex devices, wearables, processes, and software within the information system.

From the data perspective, metadata regarding static and dynamic aspects of the IoE network are stored as structured information in the Enterprise Knowledge Graph, capable of providing an interconnected view over disparate heterogeneous resources. These include technical specification of resources, their deployment in the industrial environment, as well as the organization of the enterprise in physical (i.e. its topology), and logical terms (i.e., the organizational charts and the roles with corresponding rights). Furthermore, contextual information on people and smart objects, such as their location and current activities, are represented as well.

On top of this knowledge layer, the general approach adopted for the design of the platform is based on two main principles. On the one hand, every data producer, being a simple IoT device, a smart object, or a module in the information system, is considered a stream generator. As such, we assume it produces data in a certain format, with a certain schema, and with a given (deterministic or non-deterministic) frequency. On the other hand, the platform is built as a micro-service architecture where each component is a stateless stream processor, which consumes one or more streams in input, performs stream manipulation, and generates one or more output streams following a publisher/subscribed approach. All the metadata needed to characterize a stream producer and a stream processor are stored in a Technical Knowledge Graph.

This approach grants high flexibility and modularity to the platform by decoupling each component in the stream processing pipeline. As a result, the platform is capable of collecting, processing, monitoring, and storing multiple heterogeneous big data streams, supporting effective and efficient anomaly detection for highly heterogeneous devices. The platform also supports a number of additional downstream applications with precise access control policies, among which are real-time analytics, predictive maintenance, contextual assistance for employees in daily tasks, co-bots, or autonomous applications performing actions when certain events occur.

IV. Knowledge Model

The data model of the platform is built upon an Enterprise KG, which serves as a unified framework to structure and integrate knowledge within the IoE ecosystem. This approach ensures a comprehensive representation of metadata related to systems (e.g., sensors or actuators), smart objects, and employees, detailing their placement within the organizational environment, as well as the associated rights for accessing systems, preferences, tasks to achieve, and responsibilities. The KG leverages the terminology defined in SemIoE [7], an OWL2 lightweight ontology specifically designed to offer a structured and standardized framework for defining IoE entities and their relationships. By doing so, it enhances semantic interoperability and improves the understanding of IoE environments. SemIoE incorporates and extends various external modules to represent specialized aspects. Among them, the W3C Semantic Sensor Network (SSN) ontology [14] plays a key role in defining the technical characteristics of sensors and actuators. The key SemIoE classes involved in data gathering and monitoring can be summarized as follows:

- *Agent*: represents an employee or a smart object, located in a specific *Site* (e.g., a room within a facility) and involved in an *Activity* within a business process (e.g., machinery maintenance).
- *SmartObject*: consists of one or more systems (e.g., CO2 sensors, ventilation actuators), each characterized by a set of technical properties and operating conditions, following the SSN ontology.
- *System*: a device producing data, for which both the data format and protocol are specified. For example, an IoT CO2 sensor may generate JSON-formatted messages with attributes like timestamp or CO2 level.
- *Role* and *Right*: agents hold Roles (e.g., maintenance technician) associated with Rights (e.g., read/write permissions) on specific systems, whole smart objects or on particular sites.
- *Preference*: agents can define preferences for environmental parameters. For instance, employees may set comfort preferences for temperature and humidity, while ventilation systems may specify optimal CO2 thresholds.

We refer the interested reader to http://w3id.org/semioe for the ontology specifications. To support data collection, monitoring, and access control, a Technical KG is also defined to include operational metadata for devices generating data streams, such as the topic to which the data stream is published, the target storage system (e.g., a time-series database, relational database, or file storage) and its schema. Additionally, the Technical KG documents the transformations applied to data streams, thereby supporting the stream processing steps managed by the Stream Management Platform discussed in the next section.

V. Proposed Architecture

In this section, we describe our proposed architecture, visible in Figure 1, for semantic data stream gathering and monitoring within the IoE, supporting anomaly detection. Our framework consists of two main parts, namely:

- **Stream Management Platform**: is capable of gathering data streams, performing their manipulation, and managing their monitoring and storage in real-time leveraging the Enterprise and Technical KGs.

- **Anomaly Detection Platform**: is employed to monitor the entities of the system through the computation of their Behavioral Fingerprinting.

A. Stream Management Platform

The framework architecture leverages a modular structure to support seamless expansion and integration. It combines technical interoperability through standard protocols with semantic interoperability via the Knowledge Graphs, enabling device integration and real-time monitoring. The system treats all devices as data stream producers, ensuring uniform data handling and easy integration of new sources. The platform architecture consists of the following components:

- *Data collection*: sensors, smart objects, and IT subsystems generate data streams. According to the source type, data may flow to a broker (e.g., MQTT) directly or through a gateway collecting all streams from a certain enterprise location (e.g., a specific lab), using various protocols (e.g., BACnet/IP, MQTT) and formats (XML, JSON). . Brokers typically maintain distinct queues of messages, namely *topics*, according to a pre-defined syntax, e.g., *room1/CO2_sensor2/CO2_lev* for the 'CO2_lev' value produced by a 'CO2' sensor located in 'room1'.
- *Stream preprocessing*: it transforms data streams before publishing them to the stream management system, applying transformations such as merging, splitting, filtering, decompression, decryption, and normalization, based on the metadata stored in the KGs.
- *Stream management*: a Kafka cluster serves as a unique collector for heterogeneous data streams, enabling large-scale, real-time processing. It organized incoming data into topics and partitions, allowing parallel consumption by clients and ensuring an efficient data flow.
- *Stream postprocessing*: this component applies rules for data filtering and transformation before storing the data in the database, guided by device metadata stored in the Technical KG, including stream aggregation or stream compression to reduce the cardinality of the data to store.
- *Persistent data storage*: DBMSs to which postprocessed data streams are routed for persistent storage.
- *Stream monitoring*: this service enables real-time data consumption for a wide set of applications, while enforcing security and regulatory compliance by enabling access to specific streams only to authorized clients.
- *Semantic interface*: it enables the interaction with the KGs to provide metadata on data sources, data schemas, sensor locations and technical metadata. It supports advanced functionalities such as dynamic stream processing and context-aware monitoring, enhancing the adaptability of the system.

The monitoring system offers two main functions to client applications to perform the selection of the required streams and consume them in real-time: (1) a traditional, purely syntactical approach and (2) a more advanced, semantics-based approach leveraging a flexible topic subscription system. The basic approach operates through direct topic subscriptions following standardized naming patterns, such as room/sensor_id/measure_type. This method ensures simplicity and efficiency, but it depends entirely on the predefined naming conventions at enterprise level. This can make it difficult to adapt to changes as any modifications, such as renaming values, adding new dimensions, or restructuring hierarchical elements (e.g., room/floor/building), may complicate topic management, potentially leading to inconsistencies and errors in data handling. To overcome these limitations, the semantic-based approach relies on the KG to decouple topic names from the actual meaning of the data streams. Instead of relying exclusively on fixed topic names, this approach enables dynamic queries using SPARQL to interact with the KGs. This allows the clients, whether human agents or automated systems, to express the monitoring request using the KG terminology, without the hassle of coping with specific syntactical constraints. In this way, it is possible to request a stream in multiple ways:

- by *system*: the service looks up the corresponding topic name in the KG and performs the subscription to the Kafka topic;
- by *smart object*: the service identifies all systems embedded in the smart objects and performs the subscription to their topics;
- by *environment*: the service retrieves from the KG all systems located in the provided environment and subscribes to all their topics;
- by *agent location*: the service identifies the location of the client, and retrieves the sensors that are co-located.

In all such cases, before delivering the retrieved streams to the client, the service preliminarily filters out those streams that are not accessible to the client because of its access rights, as stored in the KG. Authorization is essential for enforcing role-based access control, ensuring that service returns results tailored to the agent's current role and location within the organization. For example, a maintenance technician responsible for monitoring temperature levels on the second floor of a building, has read access to the environment labeled "Floor_2". This access logically extends permissions to all sensors deployed in rooms on that floor through logical inference within the KG.

B. Anomaly Detection Platform

This subsection presents the architecture used for the computation of the Behavioral Fingerprinting (BF) approach. To do this we employ a Federated Learning (FL) based solution. FL is a decentralized and collaborative Machine Learning paradigm that enables model training across multiple devices while ensuring data privacy, as it eliminates the need to share raw datasets [23]. Leveraging FL, we can build a comprehensive behavioral fingerprint for an entity of the network by aggregating insights from multiple interactions across that entity's network with its peers. This enables a broader and more accurate behavioral analysis, incorporating diverse perspectives from various connected devices. Our scenario can be modeled as a directed graph as follows:

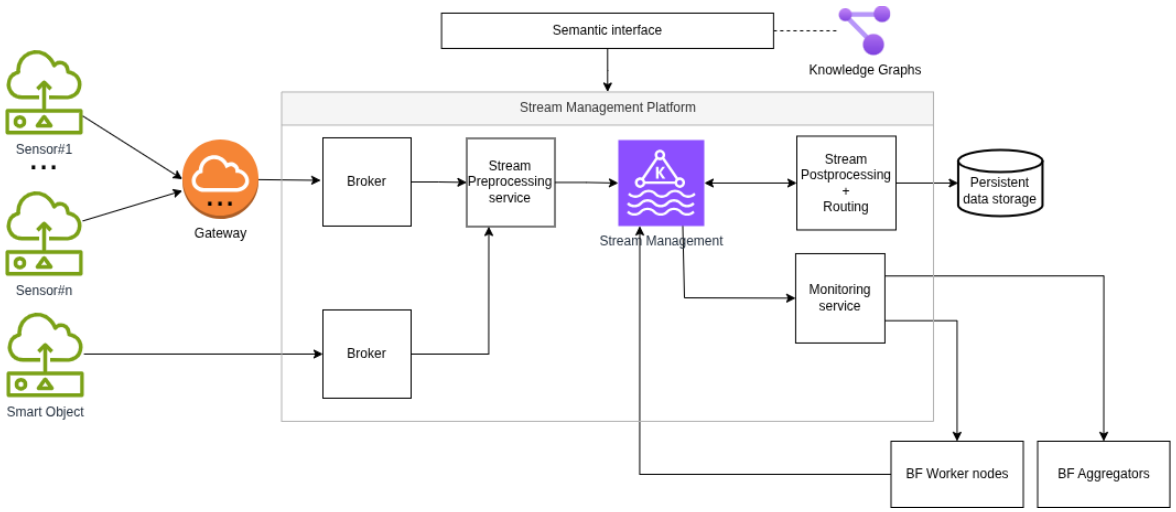

Fig. 1. Overview of the framework architecture. Edges represent information flow.

$$\mathcal{G} = \langle N, E \rangle$$

where N represents the set of nodes (or entities) and E denotes the edges connecting them, formed when nodes exchange messages. The direction of each edge signifies the initiating node in the communication. The neighbors of a node n_i are defined as $\Gamma_{n_i} = \{n_j \in N : (n_i, n_j) \in E\}$, capturing the interactions over the network. Our strategy, visible in Figure 2 for the service monitoring phase consists of two key steps: *(i) training phase* where distributed behavioral fingerprints are computed using FL, and *(ii) inference phase* in which real-time anomaly detection is performed to identify suspicious activities while preserving privacy.

The FL-based BF model involves several key actors, each interacting through the Monitoring Service of the Stream Management Platform to facilitate data collection and model training:

1) *Worker nodes*. These nodes retrieve data produced by IoT devices, namely Target nodes, through the Monitoring service. Once trained, they send model gradients back by republishing on Kafka as a new stream for the Aggregators.
2) *Aggregators*. These nodes collect gradient information from multiple workers and refine them into an updated global model; then, they send it back to the workers, ensuring continuous improvement in BF accuracy.
3) *Target nodes*. These IoT devices generate the data used for BF and are the nodes to be monitored. They send data to the Stream Management Platform but do not perform any computational tasks related to model training.

During the initial *safe* starting phase, all devices publish streaming data on the Stream Management Platform, enabling model training in a controlled environment free from external attacks. This phase ensures that the BF models establish a reliable baseline for normal behavior before deployment. The core of our approach relies on a Deep Learning model designed to classify device behaviors based on extracted features from network traffic. Each worker node extracts key behavioral features, including:

1) *Network parameters*. This information includes source port type, TCP flags, encapsulated protocol types, inter-arrival times of packets, and packet length.
2) *Payload-based features*. This information is gathered from the message payloads to identify meaningful patterns and detect potential cyber-physical attacks.
3) *Sequence-based encoding*. This information maps packet exchange sequences into symbolic representations for improved learning efficiency.

The chosen model architecture is a Gated Recurrent Unit (GRU) neural network composed of two recurrent layers, with 512 and 256 neurons, respectively, and a fully connected layer with 128 neurons. In accordance with the description in [6], the training process can be orchestrated while maintaining privacy by employing a private blockchain that acts as a public ledger. This ledger facilitates the sharing of secrets which helps in identifying the designated aggregator among the participating entities, along with encrypted updates. This approach could similarly be adapted into our solution in a semi-centralized manner by delegating the role of aggregator to the stream postprocessing service. This service can construct various global BF models utilizing the IoE network and distribute them via the Stream Management System.

As a final remark on the training phase, it is worth underlining that the anomaly detection component depicted in Figure 2 is a simplified version adopted here to describe the general idea underlying our solution. However, in privacy-sensitive scenarios, the exchange of information related to gradients and model updates between workers and aggregators must be safeguarded either by adopting ad-hoc encryption solutions or by leveraging dedicated secure channels [6], [24].

Once training is completed, the fully operational phase

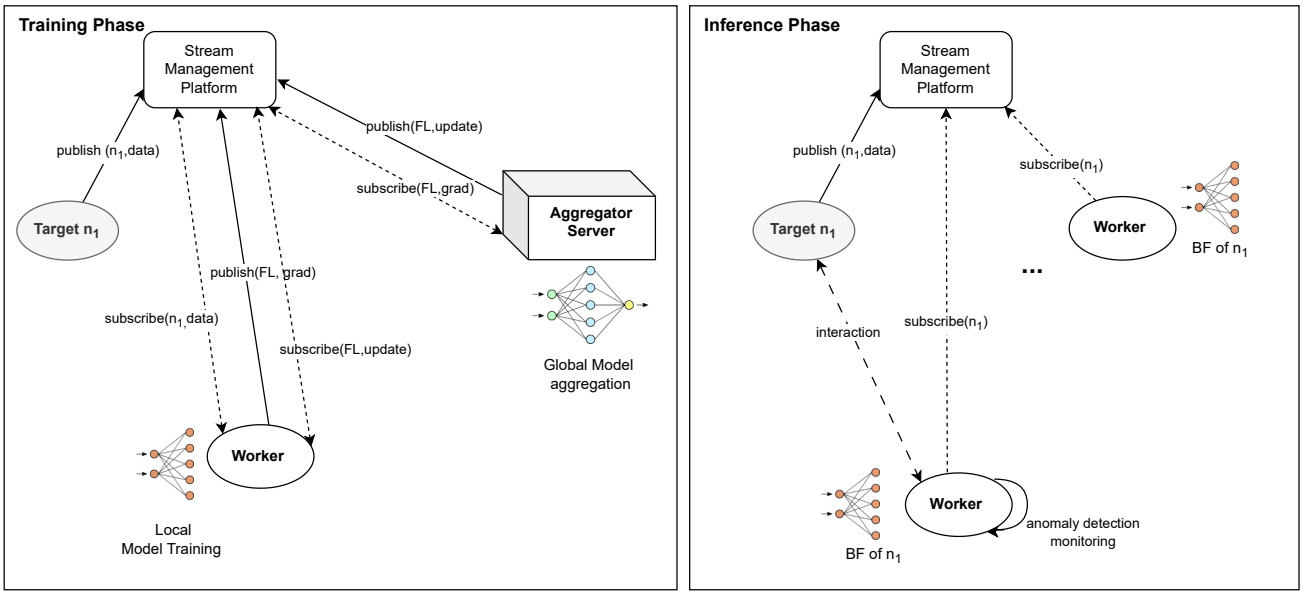

Fig. 2. Anomaly Detection component of our architecture.

(or inference phase) begins. Here, the trained FL models are deployed to detect behavioral anomalies in target nodes. Data from the Monitoring Service continuously feeds into the system, allowing real-time anomaly detection. The trained BF model leverages FL to combine insights from multiple worker nodes, creating a comprehensive behavioral profile of each device as in [4], [6]. By continuously analyzing data gathered through the Monitoring Service, the system detects deviations from normal behavior. If the number of unexpected behaviors surpasses a predefined threshold, the system flags the anomaly, which may indicate a cyberattack, hardware malfunction, or environmental shift as done in [5], [24]. If a node exhibits recurrent anomalies it is isolated from the network until intervention occurs. This ensures system integrity and enhances the security of the IoT ecosystem.

In summary, our approach efficiently gathers data through the Monitoring Service, ensuring privacy-preserving FL training. The BF model enables robust anomaly detection, contributing to a secure and intelligent IoT ecosystem as provided by the results presented in the papers that propose the original ideas [5], [6], [24].

VI. Conclusion

The novel paradigm of the Internet of Everything (IoE) integrates people, processes, data, and things, creating an interconnected and smart environment that presents unique security challenges due to its scale, complexity, and heterogeneity. Additionally, the generation of vast amounts of real-time streaming data from heterogeneous sources leads also to interoperability challenges. To tackle the above issues, in this discussion paper, we describe a novel framework developed within the HOMEY project [25]. Our proposal consists of two main components. The former deals with dynamic data stream collection and management and leverages the concept of Enterprise Knowledge Graph, which acts as a unified framework to structure and integrate knowledge within the IoE ecosystem. The latter focuses on monitoring of entities that produce data through Behavioral Fingerprinting. Entities' behaviors in the system are analyzed and a Federated Learning approach is trained to recognize possible anomalies. By integrating semantic data stream analysis and Behavioral Fingerprint, we enhance real-time threat detection, improving anomaly detection accuracy, and ensuring better privacy protection in IoE environments. This proactive security approach is crucial for protecting critical infrastructure, securing user data, and ensuring the safe expansion of IoE ecosystems in the current Industry 5.0 scenario.

This work can be considered as a starting point for future research. In particular, we plan to enhance the mechanism for retrieval of historical data through an Ontology-Based Data Access approach, and extend the platform with additional downstream applications useful in an Industry 5.0 scenario. Such functionalities will be provided under precise access control policies, and include real-time analytics, predictive maintenance, co-bots or autonomous applications performing actions when certain events occur. Furthermore, more interactive support will be provided to users by contextual virtual assistants for employees' daily tasks, based on immersive applications with Augmented Reality features and chatbots supported by Large Language Models. We plan to further develop our anomaly detection system by exploring a collaborative and distributed approach to behavioral fingerprinting of network objects. Additionally, we aim to optimize workload orchestration using secure delegation, ensuring better workload distribution across network nodes while minimizing power consumption, meeting Service Level Agreement (SLA)

requirements, and enhancing node reliability.

REFERENCES

[1] L. DeNardis, *The Internet in everything*. Yale University Press, 2020.

[2] Y. Liu, H.-N. Dai, Q. Wang, M. K. Shukla, and M. Imran, "Unmanned aerial vehicle for internet of everything: Opportunities and challenges," *Computer communications*, vol. 155, pp. 66–83, 2020.

[3] M. Sajid, A. Harris, and S. Habib, "Internet of everything: Applications, and security challenges," in *2021 International conference on innovative computing (ICIC)*. IEEE, 2021, pp. 1–9.

[4] M. Ferretti, S. Nicolazzo, and A. Nocera, "H2o: secure interactions in iot via behavioral fingerprinting," *Future Internet*, vol. 13, no. 5, p. 117, 2021.

[5] A. Aramini, M. Arazzi, T. Facchinetti, L. S. Ngankem, and A. Nocera, "An enhanced behavioral fingerprinting approach for the internet of things," in *2022 IEEE 18th International Conference on Factory Communication Systems (WFCS)*. IEEE, 2022, pp. 1–8.

[6] M. Arazzi, S. Nicolazzo, and A. Nocera, "A fully privacy-preserving solution for anomaly detection in iot using federated learning and homomorphic encryption," *Information Systems Frontiers*, pp. 1–24, 2023.

[7] M. Arazzi, A. Nocera, and E. Storti, "The semioe ontology: A semantic model solution for an ioe-based industry," *IEEE Internet of Things Journal*, 2024.

[8] M. M. Sciarroni, M. Esposito, P. Pierleoni, and E. Storti, "Monitoring data streams in industry 5.0: a knowledge graph approach," in *2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*. IEEE, 2024, pp. 566–571.

[9] K. Villalobos, V. Ramírez-Durán, B. Diez, J. Blanco, A. Goñi, and A. Illarramendi, "A three level hierarchical architecture for an efficient storage of industry 4.0 data," *Computers in Industry*, vol. 121, p. 103257, 2020.

[10] S. Anand and M. V. Ramesh, "Multi-layer architecture and routing for internet of everything (ioe) in smart cities," in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2021, pp. 411–416.

[11] M. Bolanowski, A. Paszkiewicz, T. Żabiński, G. Piecuch, M. Salach, and K. Tomecki, "System architecture for diagnostics and supervision of industrial equipment and processes in an ioe device environment," *Electronics*, vol. 12, no. 24, 2023.

[12] X. Li, M. Lyu, Z. Wang, C.-H. Chen, and P. Zheng, "Exploiting knowledge graphs in industrial products and services: A survey of key aspects, challenges, and future perspectives," *Computers in Industry*, vol. 129, p. 103449, 2021.

[13] C. Weber, H. Abu-Rasheed, and M. Fathi, "Adding context to industry 4.0 analytics: A new document driven knowledge graph construction and contextualization approach," in *2022 IEEE International Conference on Electro Information Technology (eIT)*, 2022, pp. 550–555.

[14] M. Compton *et al.*, "The ssn ontology of the w3c semantic sensor network incubator group," *Journal of Web Semantics*, vol. 17, pp. 25–32, 2012.

[15] T. C. Piller and A. Khelil, "Semsub: Semantic subscriptions for the mqtt protocol," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, 2020, pp. 1–6.

[16] Y. Wu and B. Yang, "Subscription freedom: Automatic industrial data subscription based on recommendation system," in *2022 China Automation Congress (CAC)*, 2022, pp. 3614–3619.

[17] C. Diamantini, A. Mircoli, D. Potena, and E. Storti, "Process-aware iiot knowledge graph: A semantic model for industrial iot integration and analytics," *Future Generation Computer Systems*, vol. 139, pp. 224–238, 2023.

[18] A. Iftikhar and K. N. Qureshi, "Future privacy and trust challenges for ioe networks," in *Cybersecurity Vigilance and Security Engineering of Internet of Everything*. Springer, 2023, pp. 193–218.

[19] P. Oser, F. Kargl, and S. Lüders, "Identifying devices of the internet of things using machine learning on clock characteristics," in *International conference on security, privacy and anonymity in computation, communication and storage*. Springer, 2018, pp. 417–427.

[20] S. V. Radhakrishnan, A. S. Uluagac, and R. Beyah, "Gtid: A technique for physical device and device type fingerprinting," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 519–532, 2014.

[21] B. Bezawada, M. Bachani, J. Peterson, H. Shirazi, I. Ray, and I. Ray, "Behavioral fingerprinting of iot devices," in *Proceedings of the 2018 workshop on attacks and solutions in hardware security*, 2018, pp. 41–50.

[22] A. H. Celdrán, P. M. S. Sánchez, M. A. Castillo, G. Bovet, G. M. Pérez, and B. Stiller, "Intelligent and behavioral-based detection of malware in iot spectrum sensors," *International Journal of Information Security*, pp. 1–21, 2022.

[23] J. Konečnỳ, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *arXiv preprint arXiv:1511.03575*, 2015.

[24] M. Arazzi, S. Nicolazzo, and A. Nocera, "A novel iot trust model leveraging fully distributed behavioral fingerprinting and secure delegation," *Pervasive and Mobile Computing*, vol. 99, p. 101889, 2024.

[25] HOMEY-PRIN22, "Homey Project," 2025. [Online]. Available: https://homey-prin22.unipv.it