HOMEFUS: A Privacy and Security-Aware Model for IoT Data Fusion in Smart Connected Homes

Kayode S. Adewole[1,2][a] and Andreas Jacobsson[1,2][b]

[1] *Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden*

[2] *Internet of Things and People Research Center, Malmö University, Malmö, Sweden*

Keywords: Smart Homes, Internet of Things, Data Fusion, Security, Privacy, Federated Learning, Sensors Selection.

Abstract: The benefit associated with the deployment of Internet of Things (IoT) technology is increasing daily. IoT has revolutionized our ways of life, especially when we consider its applications in smart connected homes. Smart devices at home enable the collection of data from multiple sensors for a range of applications and services. Nevertheless, the security and privacy issues associated with aggregating multiple sensors' data in smart connected homes have not yet been sufficiently prioritized. Along this development, this paper proposes HOMEFUS, a privacy and security-aware model that leverages information theoretic correlation analysis and gradient boosting to fuse multiple sensors' data at the edge nodes of smart connected homes. HOMEFUS employs federated learning, edge and cloud computing to reduce privacy leakage of sensitive data. To demonstrate its applicability, we show that the proposed model meets the requirements for efficient data fusion pipelines. The model guides practitioners and researchers on how to setup secure smart connected homes that comply with privacy laws, regulations, and standards.

1 INTRODUCTION

Internet of Things (IoT) has paved ways for connecting different sensors and smart devices to benefit from a range of applications and services including smart health, smart grids, intelligent transportation, smart manufacturing, autonomous driving, and smart agriculture. With over 7 billion connected IoT devices today, the number of smart devices that will be powered by IoT technology is expected to grow to 22 billion by 2025 (Gartner Inc., 2017). The vision of IoT is to seamlessly connect everything in the physical world over the Internet. This technology has transcended into our homes, allowing us to connect smart home devices using sensors, actuators, and controllers that are equipped with wireless connectivity, and cognitive computing technologies (Rahman et al., 2016). The main challenge is how to ensure privacy and security while also complying with laws, regulations and standards.

There are privacy laws and regulations such as the General Data Protection Regulation (GDPR) in Europe, Health Insurance Portability and Accountability Act (HIPAA) in the United States, Japan Personal Information Protection Act, and China's Cybersecurity Law. For instance, Article 23 of the GDPR enforces data controllers to store and analyze only the necessary data required to achieve data collection objectives, and to limit access to sensitive data to authorized entities (Aïvodji et al., 2019). GDPR mandates data controllers to obtain informed consent from data subjects before collecting and analyzing their data, and to provide them with the right to access, review, correct, and delete their data. Despite the existence of laws, regulations and standards, IoT companies continue to release devices into the market some of which are vulnerable (Bugeja et al., 2019; Rahman et al., 2016). Many users of IoT devices lack the knowledge of the amount and the type of data collected by the devices, and to what extent they are being utilized. This makes it difficult for users to make informed decisions regarding their privacy and how to control the data collection and distribution process.

The privacy and security risks are further complicated when multiple sensors are collaborating and when the data collected have to be transmitted to the cloud via the Internet (Mena et al., 2022; Chimamiwa et al., 2021; Adewole and Torra, 2022b; Adewole and Torra, 2022a). For instance, activities recognition in smart homes may involve data collection about

[a] https://orcid.org/0000-0002-0155-7949

[b] https://orcid.org/0000-0002-8512-2976

Adewole, K. and Jacobsson, A.
HOMEFUS: A Privacy and Security-Aware Model for IoT Data Fusion in Smart Connected Homes.
DOI: 10.5220/0012437900003705

In *Proceedings of the 9th International Conference on Internet of Things, Big Data and Security (IoTBDS 2024)*, pages 133-140
ISBN: 978-989-758-699-6; ISSN: 2184-4976

home occupant's behaviors, some of which are personally identifiable and sensitive. Some smart home devices, e.g., cameras, wearables, and health monitoring devices may not only track occupancy or fitness activities, but also infer sensitive information of the users. The complexity and lengthy nature of the policy documents that usually accompany IoT devices from the manufacturers also make it difficult for the users to have a broad understanding of the nature of the data the IoT devices are collecting and how these data are being shared. With this development, aligning with the GDPR regulations on informed consent is becoming very difficult to implement in smart connected homes (Pathmabandu et al., 2023). This requires users to make adequate decision concerning the risk associated with sharing their personal data with the third party. Therefore, there is a need for models that enforces privacy and security for smart connected homes to minimize information disclosure about users. This is the goal of HOMEFUS. More specifically, this paper contributes in the following ways:

1. We propose a privacy and security-aware model that aggregates sensors' data for effective IoT application and service delivery in smart connected homes.
2. We enforce privacy and security for sensors data collection, processing, and analysis both at the edge and in the cloud using federated learning.
3. The model leverages information theoretic correlation analysis and gradient boosting for sensors selection, fusion and modeling, which provide efficient method at the edge nodes.
4. We show that the proposed model meets the requirements for effective data fusion pipeline.

The remainder of this paper is organized as follows. Section 2 presents IoT ecosystems and characteristics. In Section 3, related work in IoT data fusion for smart connected homes is presented. Section 4 presents the requirements for data fusion architecture. In Section 5, we present the proposed model. Section 6 focuses on the evaluation and risk assessment of the model, and finally, Section 7 concludes the paper and discusses future research directions.

2 IoT ECOSYSTEM AND CHARACTERISTICS

The IoT is typically described as a network of heterogeneous devices and services. These devices are characterised as resource-constrained, making implementation of the traditional encryption and data anonymization algorithms almost impossible. The components of an IoT ecosystem have the capability to communicate autonomously over the network with unique identity. A number of architectures have been proposed for the IoT, often with the consideration of three layers: perception, network and application layers (Aïvodji et al., 2019).

In Fig. 1, we present a view of an IoT ecosystem that highlights the different components with a focus on privacy and security. In this representation, we add two layers thereby recognizing the crucial roles of the users, their applications, and the controlled management of their corresponding information. Layer 1 considers "things" in the IoT and it is the layer that accommodates the smart connected devices. These devices are hardware components capable of sensing, actuating, controlling, communicating, and monitoring. Layer 2 represents the communication and network layer where the communication infrastructure and standard protocols are used to establish interoperability among the connected smart devices of layer 1. Typical communication technologies include Zigbee, WiFi, Bluetooth, Cellular (e.g 5G), LoRaWAN, and MQTT. This layer is responsible for transmitting data to the computing and storage facilities, which are handled by layer 3. Cloud infrastructures are mostly leveraged in this layer. Layer 4 represents applications, which accommodates different software applications that render services to the end-users. Services include activities recognition of home occupants, occupancy sensing, energy consumption monitoring, surveillance, logistics, sport, business, transport, and so on.

We identified the roles of users in IoT ecosystem which permeates across the different layers. Three categories of users can be distinguished in connection to smart connected homes: residents, guests, and malicious actors. The residents and guests are considered as legitimate users. In smart connected homes, attackers or malicious actors represent uninvited guests or users who aim to establish unauthorized connection to the home network to infer sensitive information. An attacker's goal is typically to establish unauthorized access within the IoT ecosystem and to launch different attacks or privacy invasions. There are also external attackers who may want to launch complicated attacks such as denial-of-service (DoS) and malware infections to disrupt the services or infer some information from the target network or smart devices on to the Internet. In layer 5, we emphasise the need for security and privacy assessment, which should consider all the layers of the IoT ecosystem. Enforcing privacy and security in this respect will help security analysts and IoT practitioners to conduct thorough risk analysis of the IoT.

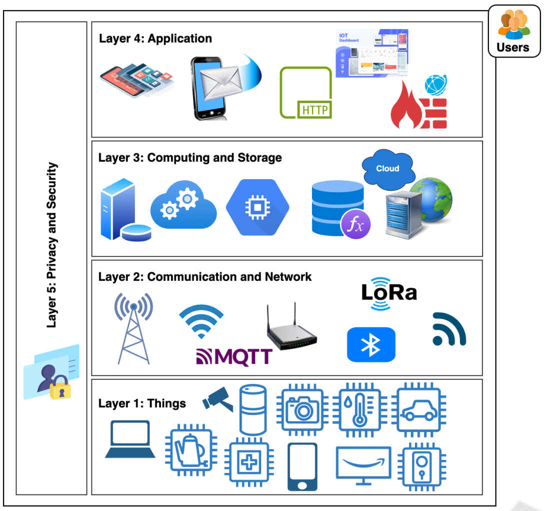
Users
Layer 4: Application
HTTP
Layer 3: Computing and Storage
Cloud
Layer 2: Communication and Network
LoRa
MQTT
Layer 1: Things
Layer 5: Privacy and Security

Figure 1: IoT ecosystem layers.

3 RELATED WORK

Information fusion deals with the study of exploring efficient automatic or semi-automatic methods of transforming data from multiple sources to produce improved representation for better decision making (Ding et al., 2019). In the domain of IoT, data fusion involves aggregation of data from multiple sensors to extract useful information for service improvement. This research domain has been widely studied in the recent years. This section presents related studies in the area of IoT data fusion with a specific focus on smart connected home research. Data fusion in smart connected homes can be categorized as intrusive or non-intrusive approaches based on the types of sensors aggregated. We provide the detail as follows.

3.1 Intrusive Approaches

Intrusive approaches use smart devices such as cameras, wearables, GPS and microphones in smartphones, just to mention a few. This approach has better accuracy for smart connected home applications, particularly for occupant monitoring. For instance, (Monti et al., 2022) investigate the use of multiple cameras for occupant counting. (Kommey, 2022) design an automated ceiling fan regulator for smart connected home based on the fusion of data from web cameras and temperature sensors. (Chaaraoui et al., 2014) proposed a weighted feature fusion approach which relies on the use of camera data. Fusion of passive infrared sensor that detect human presence together with contact, camera and microphones sensors have been studied in (Chahuara et al., 2013).

Although intrusive approaches has the potential to offer better performance for smart connected home applications, they compromise the privacy of the home occupants. Our proposed model addresses the privacy challenges in the existing studies by considering federated learning and hybrid data fusion approach to reduce privacy leakage of sensitive users information and to improve the sensor fusion method.

3.2 Non-Intrusive Approaches

The development of smart home applications that can offer comparable performance with intrusive approaches poses research challenges. To address the privacy concerns in intrusive monitoring, the non-intrusive approaches consider the fusion of environmental sensors, such as carbon dioxide (CO2), total volatile organic compounds (TVOC), air temperature, air humidity, and smart meters. (Sayed et al., 2023) proposed the fusion of temperature, humidity, pressure, light level, motion, sound, and CO2 sensors for occupancy detection in smart connected homes. (Dutta and Roy, 2022) combined different environmental sensors with contextual information to improve home occupancy detection. For a detailed discussion on data fusion as well as the methods used during the fusion pipeline, the reader is refereed to the comprehensive review in (Ding et al., 2019).

Non-intrusive approaches reduces privacy violation of home occupant information, nevertheless, they suffer from shortcomings in terms of efficiency and reliability. Moreover, studies have shown the behavior of home occupants can still be monitored through aggregating multiple sensors data even in a non-intrusive way (Pathmabandu et al., 2023). The question of which sensors should be aggregated to provide better performance while considering the limited capability of smart home devices still remain an open research issues for smart connected home applications. Additionally, risk assessment of fusion methods has not been considered in the existing studies. Our proposed model addresses the challenges in non-intrusive domain by considering hybrid lightweight sensor selection approach to improve efficiency and reliability of the fusion pipeline. Particularly, HOMEFUS leverages federated learning to further preserve privacy of sensitive data in smart connected homes.

4 REQUIREMENTS FOR SMART HOME DATA FUSION

To benefit from IoT applications in smart connected homes and to make accurate predictions and decisions, considering multiple sensors is usually advisable. Given the heterogeneous nature of sensors data from multiple sources, distributing all the data over the network will increase network bandwidth, increase power consumption, decrease the longevity of the battery-driven devices, and thus increase the costs of the system. Therefore, in data fusion, it becomes necessary to extract vital information from diverse arrays of collected sensor data to improve data quality and facilitate decision making. The question of which sensors should be aggregated to improve predictive performance still remain unsolved. Along this development, (Ding et al., 2019) provides the basic requirements for data fusion models in IoT scenarios. We provide brief discussion as follows.

1. Context-Awareness. This involves the ability of data fusion pipeline to accommodate background information in addition to the sensor data to develop intelligent fusion methods. The context may not necessarily directly relate with the sensors, but provide an additional detail that can further improve smart home applications. As an example, (Dutta and Roy, 2022) fuse environmental sensors data with 16 types of context data including AC status, fan status, door status, heat isolation, location type, area type, and so on, to improve occupancy detection.
2. Reliability. This deals with the reliability of the results from the application of the data fusion model. Because fusion results can have serious impact in decision making for smart connected home applications, such as activity recognition, diagnoses, emergency predictions, fall detections, fire outbreak predictions, energy consumption predictions, and surveillance, the output of these applications must be reliable to avoid critical situations. Reliability in data fusion can be assessed using the established metrics in information retrieval. For instance, metrics for classification tasks include accuracy, precision, recall, f-measure, and auc-roc. For clustering tasks, reliability can be assessed using metrics like rand index, Silhouette score, Davies-Bouldin index, Calinski-Harabasz index, and mutual Information. Regression tasks are evaluated using metrics like coefficient of determination, root mean square error, and mean absolute error.
3. Robustness. Data fusion needs to be robust to resist different cyberattacks, such as data injection and other forms of malicious software exploitation. Since existing solutions utilize the Internet protocol to transmit raw sensor data to the cloud, the possibility of data or code injection attacks cannot be underestimated. Robustness can be assessed based on the level of security offered by the fusion method.
4. Efficiency. This attribute deals with the ability of data fusion methods to scale irrespective of the size of the data required for the modeling process. Efficiency can be assessed in terms of training time, testing time, communication cost, and so on.
5. Verifiability. Data fusion results should be verifiable to ascertain if the contributing sensors actually improve the final fusion results. This can be assessed by checking the quality of the data collected from each sensor in relation with the smart connected home application's objective.
6. Security and Privacy: Data fusion architectures must ensure that data transmission to the fusion center is secured and that the data are handled in conformity with privacy regulation. The results of the data fusion should also be protected from unauthorized access or modification, and the availability of the fusion results should not be compromised.
7. Real-Time. Data fusion methods should deliver results to the end-users in real-time. Existing solutions leverage cloud computing capability which offer a near real-time architecture for IoT applications. Thus, data fusion model needs to consider real-time delivery of fusion results for timely decision making.

The above highlighted quality attributes guide the formulation of our proposed model. HOMEFUS is a privacy and security-aware model for data fusion targeted toward smart connected homes. Section 5 provides a detailed discussion of the proposed model.

5 HOMEFUS - THE PROPOSED MODEL

In this section, we discuss the different components of the proposed model, and formally guide through its requirements. HOMEFUS achieves privacy and security, aligns with the requirements for data fusion as previously stated, and includes risk assessment and mitigation strategies that promote the evaluation of privacy and security threats directed to the smart connected home (see Fig. 2). The proposed model considers the sensitivity of the data generated and has the ability to use both intrusive and non-intrusive data sources since the raw data of the home occupants are not transmitted to the cloud. It also advocates secure storage and computation in the cloud to prevent inference attacks on the local federated machine learning model that is transmitted from each smart connected home.

Formally, the proposed approach models a smart connected home ecosystem as a tuple (H,U,C,N,A,F,L_m,G_m,P) where H: smart connected homes, U: users, C: context data, N: connected nodes or devices, A: smart connected home applications, F: mapping function, L_m: local model, G_m: global model, and P: policy. We provide the details as follows.

5.1 Smart Connected Homes (H)

This refers to the set of smart connected homes that have agreed to implement the data fusion pipeline offers by the proposed model. Formally, we identify a set of smart connected homes $H = \{h_1,h_2,...,h_n\}$ where n is the number of homes. To allow for scalability of the model, we assume the value of n is dynamic.

5.2 Users (U)

Users represent smart connected home occupants including family members and invited guests. A set of users in home h_i, $h_i \in H$, is denoted as $U^{h_i} = \{u_1^{h_i},u_2^{h_i},...,u_m^{h_i}\}$, $\forall u_i \in U$. m is also considered to be dynamic, because a user u_i can join or leave the home network at any time.

5.3 Context Data (C)

C^{h_i} represents context data generated in each h_i, $\forall h_i \in H$ at a particular timestamp. Similar to (Dutta and Roy, 2022), we distinguish two type of context data: static (e.g location type, area type) and dynamic (e.g fan status, awake status). Thus, $C^{h_i} = \{C_s^{h_i} \cup C_d^{h_i}\}$.

5.4 Nodes (N)

This represents smart connected devices or nodes. We identify N^{h_i}, the set of devices used by user u_i in home h_i, $\forall u_i \in U$ and $\forall h_i \in H$. These devices have the capability to sense, actuate, process, and transmit data. Usually, $N^{h_i} = \{C_n \cup M_n \cup P_n\}$ where C_n are connected devices with fixed location (e.g., a washing machine, a fridge, etc.); M_n are mobile devices (e.g., a smartphone, a laptop, etc.), and P_n are the processing nodes (e.g., the edges or the cloud). P_n has additional features which include storage, processing, and analytic capability. From a security perspective, they are also equipped with intrusion detection and prevention system (IDPS). Edge nodes are located inside the home h_i while cloud connection is outside the home. Both edge and cloud nodes are connected using a smart home gateway. This gateway is maintained by the middleware layer. From a security perspective, we assume that the gateway uses standard protocols with standard security configurations. A typical example is the use of MQTT, which operates using the publish and subscribe principle and can work with SSL/TLS encryption. When a sending node publish a topic, the receiving node can securely subscribe to it. HOMEFUS leverages homomorphic encryption to secure the local model to be transmitted via the gateway to the cloud.

Each device in C_n and $M_n \in N^{h_i}$ can have one or more sensors $S_j^{h_i}$, $j = 1,2,...,k$ and $i = 1,2,...,n$ attached to it. Our proposed model aims to identify the sensor $S_j^{h_i}$ that should be fused with context data C^{h_i} to improve smart connected home applications and services. To ensure this, we propose filter-based sensor selection using information gain and correlation-based methods. These methods are not computationally expensive compared to the existing state-of-the-art approaches. Additionally, we chose this method to take advantage of edge computing and offer a low-computationally demanding solution.

5.5 Applications (A)

Every activity or behavior of user $u_i^{h_i} \in U$ can signal a specific smart connected home application $a_i \in A$, where A is denoted as a set of smart home applications or services, i.e., $A = \{a_1,a_2,...,a_\ell\}$. Typical applications or services include activity recognition (e.g., cooking, washing, etc.), fire detection, surveillance, energy management, remote monitoring, and so on.

5.6 Mapping Function (F)

The role of the mapping function F is to map sensor $S_j^{h_i}$, $j = 1,2,...,k$ to a specific application or service $a_i \in A$. This mapping is denoted as $F(\{S_1^{h_i},S_2^{h_i},...,S_k^{h_i}\},a_i)$. The mapping function computes the relevance of each sensor to the IoT application or service a_i. F relies on both information gain and correlation-based sensor selection for the mapping.

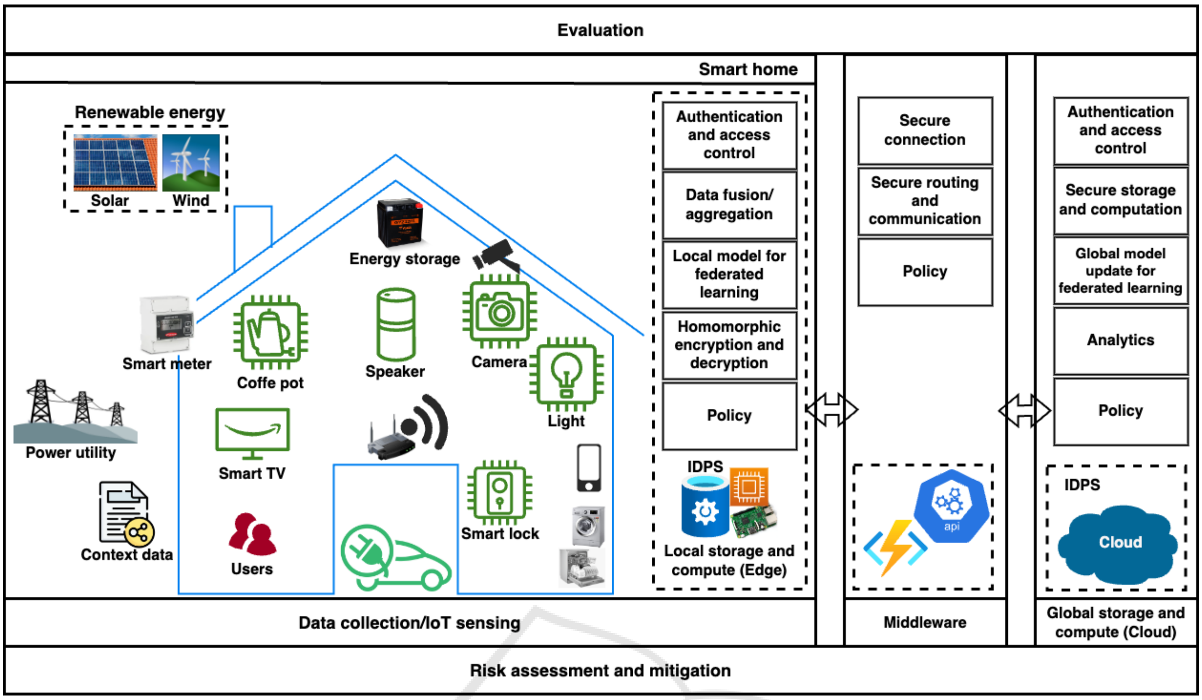
Evaluation
Smart home
Renewable energy
Solar
Wind
Energy storage
Smart meter
Coffe pot
Speaker
Camera
Light
Power utility
Smart TV
Smart lock
Context data
Users
Authentication and access control
Data fusion/ aggregation
Local model for federated learning
Homomorphic encryption and decryption
Policy
IDPS
Local storage and compute (Edge)
Secure connection
Secure routing and communication
Policy
api
Authentication and access control
Secure storage and computation
Global model update for federated learning
Analytics
Policy
IDPS
Cloud
Data collection/IoT sensing
Middleware
Global storage and compute (Cloud)
Risk assessment and mitigation

Figure 2: Proposed model for data fusion in smart connected home.

Formally, given ℓ number of applications, information gain (IG) can be obtained from entropy as follows.

$$Entropy(A) = -\sum_{i=1}^{\ell} p(a_i)log_2(p(a_i)) \quad (1)$$

where $p(a_i)$ is the probability of A extracted according to application a_i. The contribution of each sensor is estimated according to Eqn. 2.

$$F_{IG}(A, S_j^{h_i}) = Entropy(A) - \Sigma_{v \in S_j^{h_i}} \frac{|S_j^{h_i}|}{|A|}.Entropy(S_j^{h_i}) \quad (2)$$

where $F_{IG}(A, S_j^{h_i})$ is a mapping function that computes the usefulness of sensor $S_j^{h_i}$ for applications A based on information theoretic. This quantifies the information gained by this sensor. The higher the value of $F_{IG}(A, S_j^{h_i})$, the more useful the sensor $S_j^{h_i}$ is.

Similarly, a merit score, F_M can be obtained for the sensors using correlation-based sensor selection method as in Eqn. 3.

$$F_M(\{S_1^{h_i}, S_2^{h_i}, ..., S_k^{h_i}\}, A) = \frac{d\overline{r_{AS}}}{\sqrt{d + d(d-1)\overline{r_{SS}}}} \quad (3)$$

where d is the number of sensors in the subset, $\overline{r_{AS}}$ is the mean of application-to-sensor relevance correlation, and $\overline{r_{SS}}$ is the mean of sensor-to-sensor relevance correlation. The subset of sensors with the highest merit is selected as the output of the correlation-based sensor selection approach.

We then define $F_{IGM}^{h_i} = \{F_{IG}(.) \cup F_M(.)\}$, with any duplicate removed, which represent the sensors selected by information gain and correlation sensors selection approaches that are considered in the proposed model. $F_{IGM}^{h_i}$ is then merged with the context data C^{h_i} i.e $F_S^{h_i} = \{F_{IGM}^{h_i} \cup C^{h_i}\}$, where $F_S^{h_i}$ denote the data used to develop the local federated learning model on the edge of each home, $h_i \in H$.

5.7 Federated Learning

The proposed model employs federated learning to enhance privacy of home occupants. In this setting, the raw data of home users U are not transmitted directly to the cloud, but rather the learned model trained from the individual home $hi \in H$ who have subscribed to participate in the ecosystem. This local model is encrypted using homomorphic encryption. Homomorphic encryption allows computation on encrypted data in the cloud. In our proposed context, federated learning allows multiple homes to collaboratively train a machine learning model without necessarily sharing their data with each other. For the machine learning model, we propose gradient boosting methods due to the following reasons: accuracy, train faster especially on large datasets, and capable of handling noisy data (Mwiti, 2023). In practice, a federated XGBoost algorithm can serve this purpose (IBM, 2023).

Formally, let L_m denotes the local model trained at the edge node and G_m represents the global model updated at the cloud node. The learning can be estimated as given in Eqn. 4.

$$G_m = argmin_w \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(F_{sj}^{h_i}, aj) \sim L_m^{h_i}[\ell(w, F_{sj}^{h_i}, aj)] \tag{4}$$

where w is the model parameters, n is the number of homes, $L_m^{h_i}$ is the probability distribution of data at home h_i, $\ell(.)$ is the loss function, and $\mathbb{E}(.)$ is the expected value. The goal of the model is to obtain the global parameter G_m that minimizes the loss function.

5.8 Policy (P)

Policy regulates the different security and privacy aspects of the model. These are the rules that can be implemented to ensure security and privacy. Policy is denoted as $P = \{p_1, p_2, ..., p_r\}$ where $p_i \in P$ may include fusion policy, data transfer policy, data retention policy, encryption policy, key management policy, storage policy, and model update policy. For instance, model update policy may include a rule that specify when a local model should be updated. Considering this policy, a timestamp attribute can be used to automate the update process. The fusion policy will include a rule that ensures only the authenticated sensor nodes are allowed to connect. Other policies, $p_i \in P$, will also have specific rules that guide the model in a collaborative manner.

6 EVALUATION AND RISK ASSESSMENT

The proposed model is formulated taking into consideration the requirements for data fusion. The following privacy and security threats are identified. We briefly highlight how they are addressed in the proposed model.

6.1 Privacy Threats

- Linkage. An attacker wants to reveal information that is not disclosed by home occupants by linking data from different sources. Since raw data of the users are not transmitted, this threat is minimized.
- Localization and Tracking. An attacker wants to record the location of home occupants and track their movements. The proposed model preserves users privacy and the IoT devices are configured to transmit data to the edge nodes, and in doing so; this type of attack is prevented.
- Model Inversion. An attacker wants to determine if a particular home or entity has been used to develop the local or global model. The countermeasure here is that the communication between the edge and the cloud is encrypted, making it difficult to compromise the model.
- Profiling. A hacker wants to collect and correlate data to generate new data about home occupant. The main countermeasure is that raw data transmission is not permitted by the data transfer policy.
- Inventory Attack. An attacker wants to gain unauthorized access to the home network and gather occupant data from an edge node. Since the edge node is equipped with IDPS, this will prevent unauthorized data collection.

6.2 Security Threats

- Confidentiality. A hacker wants to have the knowledge about the data in transit and those at storage to compromise data integrity. The countermeasure here is that the policy governing the model does not permit raw data transfer between the gateway and the cloud. The cloud provides secured storage and computation for the model results, making it difficult for attacker to succeed in this respect. Authentication and access control mechanisms are also provided in the cloud and the edge nodes, which also strengthen confidentiality.
- Data Integrity and Data Poisoning. An attacker wishes to modify the data in transit and to inject fake data to compromise data integrity. The possibility of this attack is limited since authentication and access control mechanisms are provided. In addition, only the model trained is transmitted in encrypted form, which also reinforces the means to ensure data integrity.
- Eavesdropping. This is also known as sniffing or snooping. In this type of attack, the hacker relies on unsecured network communications to access data in transit between nodes or devices. The possibility of this attack is limited since there is a secured connection between the gateway and the cloud for the model transfer.
- Denial of Service. A hacker wants to compromise availability of the data fusion pipeline by flooding the network with superfluous requests. This attack is prevented by the IDPS that is running on the edge and the cloud. In addition, the middleware also ensures that only the secured connections and sessions are routed, and that requests are time-bound.

7 CONCLUSION

In this paper, a privacy and security-aware model for smart connected home applications is proposed. It advocates privacy and security for IoT data fusion in smart pervasive living spaces where a lot of personal data is generated, stored, and distributed. In the model, the requirements for efficient data fusion pipeline are considered, and federated learning to protect home occupants' data and improve predictive analysis are adopted. Edge nodes are considered for local model training and deployment, and a secure connection is established between the edge and the cloud. We show that the proposed model meets the requirements for efficient data fusion and that it can be applied to a variety of smart connected home applications and services. Future work will consider empirical analysis of the performance of the proposed model, considering its different components.

ACKNOWLEDGMENT

This work was partially funded by the Knowledge Foundation (Stiftelsen för kunskaps- och kompetensutveckling – KK-stiftelsen) via the Synergy project Intelligent and Trustworthy IoT Systems (Grant number 20220087).

REFERENCES

Adewole, K. S. and Torra, V. (2022a). Dftmicroagg: a dual-level anonymization algorithm for smart grid data. *International Journal of Information Security*, pages 1–23.

Adewole, K. S. and Torra, V. (2022b). Privacy issues in smart grid data: From energy disaggregation to disclosure risk. In *International Conference on Database and Expert Systems Applications*, pages 71–84. Springer.

Aïvodji, U. M., Gambs, S., and Martin, A. (2019). Iotfla: A secured and privacy-preserving smart home architecture implementing federated learning. In *2019 IEEE security and privacy workshops (SPW)*, pages 175–180. IEEE.

Bugeja, J., Vogel, B., Jacobsson, A., and Varshney, R. (2019). Iotsm: an end-to-end security model for iot ecosystems. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 267–272. IEEE.

Chaaraoui, A. A., Padilla-López, J. R., Ferrández-Pastor, F. J., Nieto-Hidalgo, M., and Flórez-Revuelta, F. (2014). A vision-based system for intelligent monitoring: human behaviour analysis and privacy by context. *Sensors*, 14(5):8895–8925.

Chahuara, P., Portet, F., and Vacher, M. (2013). Making context aware decision from uncertain information in a smart home: A markov logic network approach. In *Ambient Intelligence: 4th International Joint Conference, AmI 2013, Dublin, Ireland, December 3-5, 2013. Proceedings 4*, pages 78–93. Springer.

Chimamiwa, G., Alirezaie, M., Pecora, F., and Loutfi, A. (2021). Multi-sensor dataset of human activities in a smart home environment. *Data in Brief*, 34:106632.

Ding, W., Jing, X., Yan, Z., and Yang, L. T. (2019). A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. *Information Fusion*, 51:129–144.

Dutta, J. and Roy, S. (2022). Occupancysense: Context-based indoor occupancy detection & prediction using catboost model. *Applied Soft Computing*, 119:108536.

Gartner Inc. (2017). Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016. https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016. Accessed: 5th June, 2023.

IBM (2023). Federated learning xgboost tutorial for ui. https://www.ibm.com/docs/en/cloud-paks/cp-data/4.6.x?topic=samples-xgboost-tutorial. Accessed: 5th June, 2023.

Kommey, B. (2022). Automatic ceiling fan control using temperature and room occupancy. *JITCE (Journal of Information Technology and Computer Engineering)*, 6(01):1–7.

Mena, A. R., Ceballos, H. G., and Alvarado-Uribe, J. (2022). Measuring indoor occupancy through environmental sensors: A systematic review on sensor deployment. *Sensors*, 22(10):3770.

Monti, L., Tse, R., Tang, S.-K., Mirri, S., Delnevo, G., Maniezzo, V., and Salomoni, P. (2022). Edge-based transfer learning for classroom occupancy detection in a smart campus context. *Sensors*, 22(10):3692.

Mwiti, D. (2023). Gradient boosted decision trees [guide]: a conceptual explanation. https://neptune.ai/blog/gradient-boosted-decision-trees-guide. Accessed: 5th June, 2023.

Pathmabandu, C., Grundy, J., Chhetri, M. B., and Baig, Z. (2023). Privacy for iot: Informed consent management in smart buildings. *Future Generation Computer Systems*, 145:367–383.

Rahman, A. F. A., Daud, M., and Mohamad, M. Z. (2016). Securing sensor to cloud ecosystem using internet of things (iot) security framework. In *Proceedings of the International Conference on Internet of things and Cloud Computing*, pages 1–5.

Sayed, A. N., Bensaali, F., Himeur, Y., and Houchati, M. (2023). Edge-based real-time occupancy detection system through a non-intrusive sensing system. *Energies*, 16(5):2388.