Research Article

Secure IoT Healthcare Architecture with Deep Learning-Based Access Control System

K. Thilagam,1 A. Beno,2 M. Vanitha Lakshmi,3 C. Bazil Wilfred,4 Santhi M. George,5 M. Karthikeyan6, Vijayakumar Peroumal7, C. Ramesh,8 and Prabakaran Karunakaran9

1Department of ECE, Velammal Engineering College, Chennai, India
2Department of Electronics and Communication Engineering, Dr. Sivanthi Aditanar College of Engineering, Tiruchendur, India
3Department of ECE, Saveetha School of Engineering, SIMATS, Chennai, India
4Department of Mathematics, Karunya Institute of Technology and Sciences, Coimbatore, India
5Department of ECE, RMK Engineering College, Thiruvarur, India
6Dept of Computing Technologies, SRM IST, Kattankulathur, India
7School of Electronics Engineering, Vellore Institute of Technology, Chennai, India
8Department of Mechanical Engineering, M.Kumarasamy College of Engineering, Karur, India
9Department of Mathematics, Mettu university, Ethiopia

Correspondence should be addressed to Prabakaran Karunakaran; prabakaran@meu.edu.et

Received 6 April 2022; Revised 18 April 2022; Accepted 20 April 2022; Published 19 May 2022

Academic Editor: Samson Jerold Samuel Chelladurai

Copyright © 2022 K. Thilagam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The existing healthcare system based on traditional management involves the storage and processing of large quantities of medical data. The incorporation of the Internet of Things (IoT) and its gradual maturation has led to the evolution of IoT-enabled healthcare with extraordinary data processing capability and massive data storage. Due to the advancement in the Industrial Internet of Things (IIoT), the resulting system is aimed at building an intelligent healthcare system that can monitor the medical health of the patient by means of a wearable device that is monitored remotely. The data that is gathered by the wearable IoT module is stored in the cloud server which is subject to privacy leakage and attacks by unauthorized users and attackers. To address this security issue, an IoT-based deep learning-based privacy preservation and data analytics system is proposed in this work. Data is collected from the user, and the sensitive information is segregated and separated. Using a convolutional neural network (CNN), the health-related information is analyzed in the cloud, devoid of users’ privacy information. Thus, a secure access control module is introduced that works based on the user attributes for the IoT-Healthcare system. A relationship between the users’ trust and attributes is discovered using the proposed work. The precision, recall, and F1 score of the proposed CNN classifier are achieved at 95%. With the increase in the size of the training set, higher performance is attained. When data augmentation is added, the system performs better without data augmentation. Further, the accuracy of around 98% is achieved with an increased user count. Experimental analysis indicates the robustness and effectiveness of the proposed system with respect to low privacy leakage and high data integrity.

1. Introduction

In recent years, there has been an unprecedented growth in Wireless Sensor Networks and their applications, in terms of data computation, interoperability, scalability, interfacing, and applications. This advancement in technology along with the innovations in cellular networks, wireless networks, and radio frequency identification (RFID) has provided a strong root for the Internet of Things (IoT). In 1999, Kevin Ashton introduced the term IoT relating it with supply chain management. It shows a better world of objects wherein every object is connected to each other with the help of the internet [1]. Here, each object is represented as an entity that holds a digital identity. These entities are organized, controlled, and managed in a remote fashion. Owing to the tremendous advancement in the IoT [2], smart objects have
become a common device used with a diverse range of intelligent, innovative, and novel applications. Some of these applications include crowdsourcing, crowdsensing, smart agriculture, smart cities, and smart healthcare.

As a result of these advancements and innovative applications, there are numerous challenges faced in effectively enforcing IoT. Some challenges are energy management, quality of service (QoS) [3], interoperability, big data analytics [4], and security. However big data is a more complex field where there exists a relationship between several data streams and IoT objects to generate data. Hence, information is generated using big data analytics to improve decision-making. In recent years, there is a phenomenal development in the field of big data on integrating with IoT to pave the way to an array of opportunities that result in improved services for several applications [5]. Large volumes of information gathered from several resources using IoT are analyzed with the help of big data technologies. Machine learning is one of the technologies that is used to perform this analysis to provide an apt solution for the problem at hand. Based on analysis, it is identified that the incorporation of ML techniques [6] has led to the economic growth of the country at large.

Due to the recent coronavirus, the healthcare system has been imposed with a number of challenges for data storage and processing. In this accord, Internet of Things (IoT) technology is found to be the most effective method to address the future of smart healthcare. Using IoT, it is possible to properly understand data processing, access control, and intelligent identification with the use of artificial intelligence, network technology, and sensor technology to the management of healthcare to communicate and exchange information. This arrangement paves the way to a secure, efficient, real-time, and reinforced healthcare system. However, data tampering and leakage problems [7] have been identified to be the actual issues faced by the IoT healthcare sector. This is indicative of the importance of secure access control in the field of medical data. Access control lists and role-based access control are some of the traditional centralized computing environments that are used to address these security issues to an extent.

In [8], a novel model known as InfGCN is incorporated which integrates Susceptible Infected Recovery (SIR) and Graph Convolutional Networks (GCNs). The drawback with this approach is that it does not take into consideration the structure and working of the social networks about users. Similarly, this methodology requires an access control threshold that is built to preserve data integrity and protect privacy. Thus, taking into consideration these factors, a secure access control architecture using machine learning technologies is incorporated in thin IoT healthcare applications [9]. In particular, the edge access control layer is built with access control servers and trust generation servers that are used to grant user-specific authorities and involve data access requests with the help of machine learning methodologies that will enable support for several healthcare applications based on IoT. Accordingly, the proposed architecture is built with an attribute-based Secure Access Control Mechanism (SACM) [10] that uses federated deep learning for IoT health [11].

The following are the major contributions of this proposed work:

1. Every user is provided with unique authority to indicate their medical data, thereby enabling secure access control. Every edge weight is used with social networks about users, and the edge denotes the probability of connection of a particular pair of users.

2. A federated deep learning (FDL) methodology [12] is adopted to improve access control accuracy. In particular, the deep reinforcement learning method and federated learning framework are integrated to observe the threshold of access control to ensure the integrity of medical data is maintained and the privacy of the patients is preserved.

3. A real-time dataset is used to validate the experiment. Based on the experimental results, the efficiency of low privacy leakage and high data integrity in the IoT-Healthcare is monitored, recorded, and analyzed.

4. The prototype system is designed to ensure data analytics [13] and privacy preservation using deep learning. The nonprivacy data extraction algorithm is used for separating privacy information, and the health-related data is analyzed.

To avoid overfitting [14], a data augmentation methodology is used. The security of the system is enforced with a customized convolutional neural network, and the experimental analysis is carried out with 100 participants. Different scenarios are taken into consideration to determine the robustness and effectiveness of the proposed system [15].

2. Literature Survey

High data generation speed is one of the instigating factors which attracts researchers and industrialists to IoT devices. Moreover, when these devices are interlinked with cloud computing, the scalability and availability of the system will be enhanced. A good example is healthcare services that are improvised using machine learning (ML) algorithms [15] and virtual machines (VM) [16]. This methodology is aimed at utilizing the cloud resources to the maximum capacity. Similarly, several nonparametric models also exist and can be used when we do not have prior information or enough data in certain scenarios. Similar IoT devices also make use of fog computing to delimit the collaboration between the devices that have good data transmission speed and high response time [17–19].

In the field of healthcare, deep learning has played a pivotal role in discovering architectures like Hierarchical Computing Architecture (HiCH) which when integrated with algorithms like convolutional neural network (CNN), and IoT will result in the development of wireless body area network (WBAN) [20] for wearable devices. EM, KNN, C4.5, and C5.0 are machine learning algorithms that are dedicated to determining missing values, decision tree generation, etc., to create an efficient architecture/module using
AI upgrades. To further boost the ML algorithms and their functionality, several meta-algorithms have also been recently developed. Several access control issues in the IoT-Healthcare methodologies exist, and several algorithms have been developed to address these issues. In [21], Yang et al. have used break-glass access control policy and attribute-based access control to tackle the issue of encrypted medical data hacking. Using break-glass mechanism, timely access of data is ensured while attribute-based access control will require certain attributes to be met before providing access to medical data.

Liu et al. in [22] introduced the concept of multiauthority-based access control mechanism that involves multiple authentications from a specific group of people. This works well against collision attack and is lightweight to be incorporated. Similarly, in [23], Roy et al. have enabled healthcare with cloud computing using a fine-grained access control methodology. This technique incorporates an authentication methodology to provide user access control to the requester. Similarly, author Edemacu et al. in [24] have illustrated the possibility of collusion-resistant access control that can be used to share medical data in a secure and safe fashion.

Sun et al. in [25] show that the user attributes and access control policies are converted into vectors of specific lengths, thereby decreasing the access protocol for the data in an encrypted manner. Fan et al. in [26] were able to achieve data sharing and access control with the help of blockchain technology for user certification and nonrepudiation. These words are confined to the access control issue and lead to two specific problems:

(i) How to attain secure and accurate access control using the occupation and trust of the user, without affecting and exposing privacy of the user

(ii) How to user social data of the users to obtain trusts and influences of the users

(iii) Accordingly, the proposed methodology uses a secure access control module that works based on the user attributes for the IoT-Healthcare system

Since the introduction of wearable medical devices in personalized and pervasive healthcare, there is high demand for such IoT-enabled healthcare devices. In [27], the authors have introduced a wearable feedback system that can be used to observe the physical recovery of swimmers. A smart indoor anticollision system using RFID is used in [28] to enable visually challenged people to detect and avoid obstacles. Similarly, facial surface electromyography (sEMG) is used in real-time biomonitoring representing the intensity of pain endured by the patients. In [14], the authors have used an authentication mechanism that used cloud to generate a secret session key, establishing secure communication while in [29], an RFID-based privacy protection technique is used to protect the synchronization and consistency of the authentication information. Authors in [30] have used the noise in the ECG signal to identify the identity of the patient. Finally, in [31], a blockchain-based large-scale privacy protection scheme is introduced by the authors to store the data and protect is privacy in cloud or in the hospital database. The drawback with this methodology is that the privacy information that is mixed with data is still liable to be attacked. This gives rise to the proposal of privacy preservation in the IoT-enabled healthcare system.

3. Proposed Architecture

The proposed system architecture is described in detail in this section. The user, trust generation, and access control servers are chosen as the three entities that play a major role in trust-based access control for maintaining data integrity and preserving user privacy in IoT-based healthcare systems. At the user level, a privacy-isolation zone is designed which can filter out the noise and speech-related sounds and can transmit only the nonspeech body sounds and information. The gait signal is detected at the user end along with other medical information based on the acceleration stream by the privacy-isolation zone. At the cloud end, the security module and data extraction by a nonprivacy module is implemented. The medical data is often manipulated by unauthorized personnel with data tampering and privacy leakage in IoT-based health systems. Ensuring secure access control at all levels is crucial when dealing with important medical data. The trust generation servers estimate the trust of several users and provide information to the access control server to provide access. New users are considered untrusted while the trust generation and access control server-based data are considered semitrusted. The proposed system architecture is provided in Figure 1.

3.1. Privacy Isolation. A privacy-isolation zone is created at the source to receive the nonspeech body sound and data. This further guarantees the security of data transfer and cloud-based storage. A modified deep CNN algorithm is implemented at the cloud end to extract the data along with a security module with access control and trust generation servers. This ensures end to end privacy protection for medical data. The system is tested under privacy leakage and data tampering attack scenarios to estimate the performance. The user’s identity and associated gait information are often mixed with the gesture, movement, and other health-related data that is gathered by the wearable device at the user end. The data and the gait information can be separated by the malicious user if the data is stolen from the cloud, leading to the disclosure of the user’s privacy. Hence, before the upload of data into the cloud, it is essential to analyze and segment it in the privacy-isolation zone. The signal is made 0 outside the boundary and retained within the boundary with the help of a smooth window function. Multiplication is performed between the gathered signal and the signal within the boundary.

Based on a set threshold value, the signal can be extracted within the boundary. Gravity information is also available along with the gait information in the data that is collected. 9.8 m/s² is the fixed downward force gravity value. With the motion of the user, it is challenging to find a fixed threshold as there will be a change on each axis for the gravity projection. A low signal-to-noise ratio (SNR) is achieved due to the interference of the gait signals in the
obtained data making it challenging to distinguish between the data. Also, on the basis of the time domain, the data is aliased. Hence, using a window function directly for separating the aliased signals is not possible. Different frequency characteristics are observed from the different behaviour information. Fourier transform is used for the analysis of the signals in the frequency domain for separating the data. In the frequency domain, it is observed that gravity is a DC component. When compared to gravity, at a relatively high-frequency band of 1.4 to 2.1 Hz, the gait signals occur. The gait information is filtered using a low-pass filter while the gravity information is filtered using a high-pass filter. Wavelet, Elliptic, Chebyshev, and Butterworth filters can be used for realizing the low-pass and high-pass filtering functions. Due to the complex wavelet decomposition process, it is unsuitable to deploy the wavelet filter on the user terminal with constrained resources. When compared to the elliptic filter and Chebyshev filter, the amplitude-frequency characteristics of the Butterworth filter are more stable and it has relatively slow stopband attenuation and the flattest passband frequency response curve. The average SNR value for these filters is compared and found to be 11.9, 11.5, and 12 for Elliptic, Chebyshev, and Butterworth filters, respectively. The accuracy of the extracted signal is higher with a higher SNR value. Hence, Butterworth filter is the optimal choice of implementation.

3.2. Cloud Security Model. At the cloud end, data extraction and storage are performed. The filtered signal is processed, and feature extraction is performed with drift removal, envelope estimation, and feature selection. Further, it is merged with the raw signal and transferred to the security module. At the security module, data augmentation and classification are performed, and then based on the users trying to access the data, the trust generation server and access control server provide data access to the user. The collection of all sample data is limited by cost and time. Hence, to enhance the generalization of the module, more training samples are generated using data augmentation. Data collection is performed at different speeds by transforming the signal’s time-domain position using the time warping process. Further, the data collection at multiple forces undergoes an amplitude distortion process where a random change in the data amplitude is observed. Further, time scaling, permutation, rotation processing, and random noise addition are performed to represent the signal width, time position, different wearing angles of the data collection device, and noise environments, respectively.

The signal variance is increased with the outliers and internal sensor noises. This leads to signal drift. The intensity of the drift increases with the amplification of variance. Through PCA, the largest variance in the orthogonal direction is obtained when the signal is projected. The linear regression fitting method is used for removing the drift of each component. The original and fitted components are estimated, and the square sum of errors between them is calculated for each component. The trend term is then subtracted from the component. The analysis is then focused on the data fluctuation itself.

At the trust generation server, the construction of a social graph takes place. Social similarities between the users based on their social activities are identified. Based on the social similarity of the users, the connection probability is determined at the corresponding edge for the user’s social graph construction. The deep reinforcement learning (DRL) algorithm helps in achieving the trust-based access control based on the social data. The Susceptible Infected Recovered (SIR) and Graph Convolutional Network (GCN) models integrate the social data for trust evaluation.
The effect of each user node is measured using the GCN in conventional models. Here, InfGCN, an influence identification model, is used for analyzing the eigenvector, closeness, betweenness, and degree centrality representing the features of each node. The centrality of a node in eigenvector centrality represents a function of centrality of the adjacent nodes. If a connection is established between this node and other influential nodes, then the node is more influential. The position of the node in the entire structure is considered while representing the closeness centrality. Here, the node is closer to the geometric center position. The influence of the node is high if it is positioned on multiple shortest paths amongst other nodes. This is represented by the betweenness centrality. The influence of a node is high when it consists of more social connections, representing a higher degree termed as degree centrality. Then, based on the trust estimation, the access control server grants access to the trusted users and withholds permissions to the malicious or unauthorized users.

To achieve maximum data integrity and minimize privacy leakage, the learning rate, infection rate, and recovery rate must be adjusted dynamically to the access control threshold. In existing local access control models, the threshold for construction is learnt using the Twin Delayed Deep Deterministic policy gradient (TD3) algorithm. The TD3 algorithm is applied with the federated learning framework to achieve privacy preservation at the user end in the universal access control model. For each participant of federated learning, a trained model is required for the unique machine learning technology termed federated learning. This helps in building a universal model while serving as an access control server, thus overcoming the need for a private dataset for the preservation of the participants’ privacy.

4. Result and Discussion

The recognition performance of the proposed model is estimated based on the accuracy, F-score, recall, and precision parameters. Further, True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values are estimated for the confusion probability matrix. The expressions

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},
\]

\[
\text{F1} = \frac{2TP}{2TP + FP + FN},
\]

\[
\text{Recall} = \frac{TP}{TP + FN},
\]

\[
\text{Precision} = \frac{TP}{TP + FP},
\]

\[
\text{Missed Detection Rate} = \frac{FP}{FP + TN},
\]

\[
\text{False Alarm Rate} = \frac{FP}{FP + TN},
\]

are used for estimating the accuracy, F-score, recall, and precision of the samples. The relationship between the actual results and predicted results is analyzed using the confusion probability matrix. The true label is represented by each row, and the predicted label is represented by each column of the confusion matrix.

Figure 2 represents the impact of the classifier in the proposed model. In order to compare the performance, classifiers like Deep Neural Network (DNN), Long short-term memory (LSTM), Artificial Neural Network (ANN), convolution Neural Network (CNN), Recurring Neural Network (RNN), Support Vector Machine (SVM), and Random Forest (RF) are compared. The testing and training set consists of data gathered from 100 participants performing 10 different gestures and activities over a duration of 60 seconds each. The accuracy performance of CNN is found to be superior to that of the other models.

Timely updating of the model is crucial to ensure the effectiveness of the model performance in the new
environment. Better generalization can be achieved by collecting more training samples. However, it is challenging and tedious to perform frequent data collection. Data sets with different training set sizes are used for training the model in order to estimate a suitable scheme for data collection. Figure 3 compares the precision, recall, and F1 score parameters for multiple training set sizes with 10 samples per class. It is observed that the performance tends to be stable at around 95% when the training set size exceeds 100. The model is thus built using the gathered data. The participants are chosen between a diverse age group and an equal gender ratio to compare the impact of the data sample. 10 selected gestures are performed 10 times each by all 100 participants during data collection. For a different number of samples, the precision, recall, and F1 score values are estimated. It is observed that the performance does not show any significant difference based on gender or age. However, the number of samples improves the performance in a proportional manner.

The models are trained with and without data augmentation, and the performance is analyzed in terms of precision, recall, and F1 score. Overfitting may be avoided, and the uncollected data can be covered using the data augmentation technique even if it is difficult to attain a large number of training samples. When compared without data augmentation, the model achieves around 8% better performance with data augmentation as observed in Figure 4.

Data integrity, degree of privacy leakage, and accuracy of access control are estimated to analyze the access control module performance. Based on the number of malicious users, amount of available data, and number of authentic users, the system performance varies. Missed Detection Rate (MDR) and False Alarm Rate (FAR) parameters are considered for the estimation of the accuracy of the access control module. The amount of private information that is exposed when compared to the overall data available provides the degree of privacy leakage. The amount of data that is not exposed to malicious users or any kind of contamination is termed data integrity. Figure 5 compares the MDR and FAR values of the proposed model for a different number of users. With the increase in the number of users, there is a proportional raise in the performance of accuracy as observed from this graph. The DRL algorithm helps in achieving the trust-based access control based on the social data from the SIR and GCN models. The DRL algorithm is used for achieving the trust-based access control by estimating the user’s trust based on the social data of the user obtained through GCN. The authorized users are provided access to the medical data while malicious users are identified and refused access.

5. Conclusion

For IoT-based healthcare systems, a data analytics and privacy preservation model using deep learning is presented in this paper for separating raw health data and analysis and refusing access to malicious users through a trust-based and secure access control model. The privacy-sensitive and nonprivacy information is isolated using filters. The robustness and effectiveness of the system are evaluated under various circumstances, and the performance is analyzed. This model can support the smart healthcare systems of the future. This system architecture can be expanded for various wearable IoT healthcare devices. The access control model uses social graphs to decide between authorized and malicious users. These graphs, along with CNN, help in providing authorization to specific users in the IoT healthcare environment. The future scope involves overcoming the cost and time constraints of the work by using larger datasets to further generalize the performance of the system. User identity protection policies can be strengthened further with the introduction of a blockchain-based security module. Real-time sample collection and system updating can also be enabled so that the performance of the system gets better with use.
Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

The publication of this research work is only for the academic purpose of Mettu University, Ethiopia.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References


