IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON SMART CACHING, COMMUNICATIONS, COMPUTING AND CYBERSECURITY FOR INFORMATION-CENTRIC INTERNET OF THINGS

Received December 30, 2018, accepted January 27, 2019, date of publication February 27, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2899323

Secure Edge of Things for Smart Healthcare Surveillance Framework

ABDULATIF ALABDULATIF[1], IBRAHIM KHALIL[2], XUN YI[2], AND MOHSEN GUIZANI[3], (Fellow, IEEE)

[1]Department of Computer Science, College of Computer, Qassim University, Buraydah 51431, Saudi Arabia
[2]School of Science, RMIT University, Melbourne, VIC 3000, Australia
[3]Department of Electrical and Computer Engineering, University of Idaho, Moscow, ID 83844, USA

Corresponding author: Abdulatif Alabdulatif (ab.alabdulatif@qu.edu.sa)

ABSTRACT The vast development of the Internet of Things (IoT) and cloud-enabled data processing solutions provide the opportunity to build novel and fascinating smart, connected healthcare systems. Smart healthcare systems analyze the IoT-generated patient data to both enhance the quality of patient care and reduce healthcare costs. A major challenge for these systems is how the Cloud of Things can handle the data generated from billions of connected IoT devices. Edge computing infrastructure offers a promising solution by operating as a middle layer between the IoT devices and cloud computing. The Edge of Things (EoT) can offer small-scale real-time computing and storage capabilities that ensures low latency and optimal utilization of the IoT resources. However, the EoT has privacy-preservation issues, which is a significant concern for the healthcare systems that contain sensitive patient data. This paper introduces a novel EoT computing framework for secure and smart healthcare surveillance services. Fully homomorphic encryption preserves data privacy and is stored and processed within an EoT framework. A distributed approach for clustering-based techniques is developed for the proposed EoT framework with the scalability to aggregate and analyze the large-scale and heterogeneous data in the distributed EoT devices independently before it is sent to the cloud. We demonstrate the proposed framework by evaluating a case study for the patient biosignal data. Our framework rapidly accelerates the analysis response time and performance of the encrypted data processing while preserving a high level of analysis accuracy and data privacy.

INDEX TERMS Smart healthcare, Internet of Things, edge computing, homomorphic encryption.

I. INTRODUCTION

In recent years, there have been remarkable advances in the Internet of Things (IoT). Smart healthcare technology has similarly advanced alongside a rapid growth in the amount of biomedical data and the rise of 'smart' healthcare communities. As a result, the provision of smarter and more cost-effective healthcare services has now become essential. Utilizing technology to improve healthcare services can enhance the quality of patient care and reduce health care costs. Developments in the integration of the Internet of Things (IoT) and cloud computing paradigms, referred to as Cloud-of-Things (CoT), are spurring the development of smart connected healthcare systems that can monitor, aggregate, and diagnose massive amounts of biosignal data and provide convenient analytical tools for smart healthcare communities [1], [2]. However, in such systems IoT devices generate massive amounts of data and the nature of the CoT paradigm means that it relies completely on cloud computing to store and process data. These factors combined pose significant storage and processing issues that can significantly hinder the development of time-critical analytical services, such as those required in healthcare services.

Edge-of-Things (EoT) is a new computing paradigm that represents a middle computing layer between IoT devices and cloud computing, bringing computing power (e.g., IoT gateways) closer to IoT devices [3]. The EoT layer is not only useful for basic transmitted functionality, but can also perform real-time analytic services and smart decision-making within a local smart community domain. Moreover, healthcare data can be sent through the EoT layer to cloud computing for further global data processing. Figure 1 shows the architecture of our secure EoT framework for smart healthcare surveillance. Such a framework allows for the early detection and treatment of diseases; potentially reducing the harm caused by diseases and prolonging many lives. Machine learning techniques,

The associate editor coordinating the review of this manuscript and approving it for publication was Jun Wu.

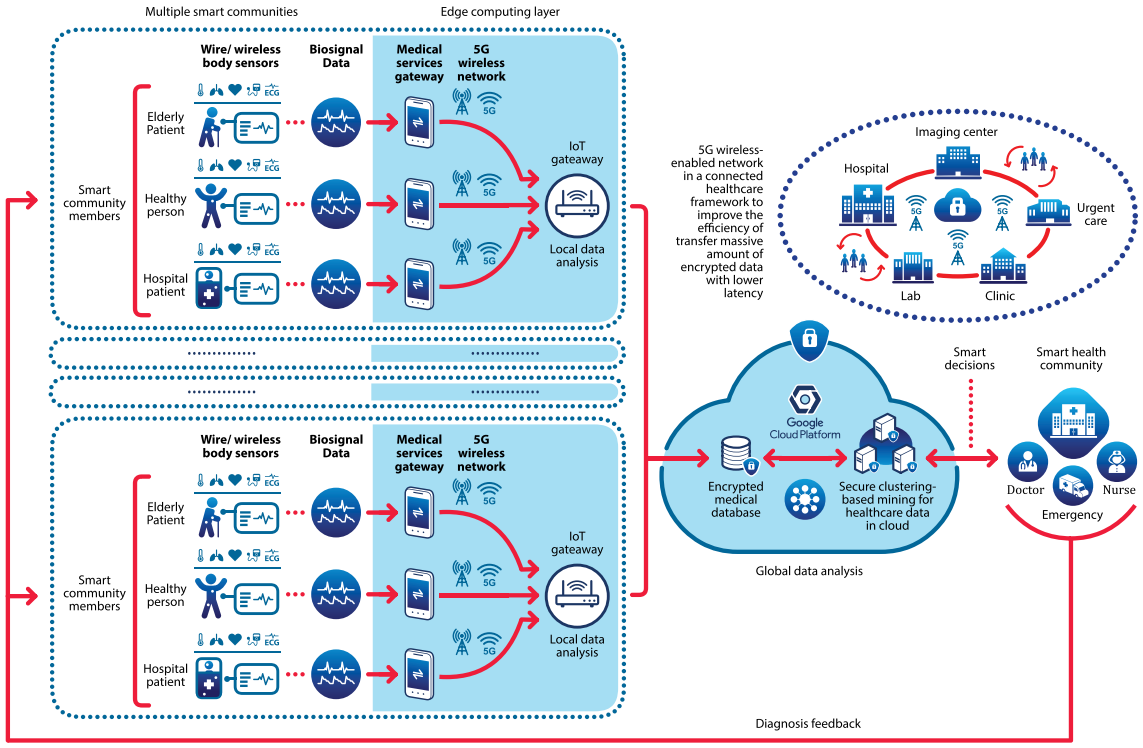

FIGURE 1. The architecture of our secure EoT framework for smart healthcare surveillance. The members' biosignal data of each smart community are sent to the edge IoT gateway upon being encrypted for analysis in a local community and then sent to the cloud computing for further global analysis of multiple smart communities data. The community members' data is analyzed based on a secure machine learning algorithms, and smart clinical decisions are made and sent to the healthcare professionals.

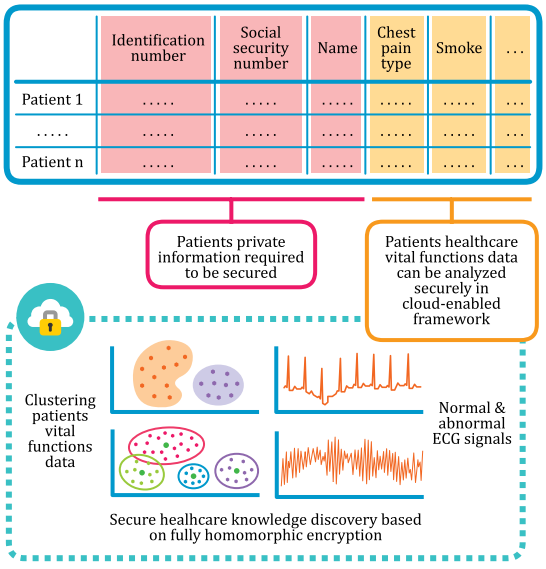

FIGURE 2. An example of the patient information structure and its personal and vital functions data. Personal data is required to be secured. The vital function data is used for analysis purposes in a secure manner.

such as clustering-based algorithms, are widely used for analyzing biosignal data and classifying patients into different groups according to specific health conditions, with the ability to detect abnormal patterns. Figure 2 shows an example of the patient information structures, which contain patients' personal and biosignal data. Analysis of biosignal data can predict several chronic diseases. For example, the classification of different types of chest pain and blood pressure can indicate the presence of heart disease. Clustering-based techniques are an efficient analytic tool that can support clinical decision-making in healthcare communities. The clustering of community members' data allows for the discovery of new health insights and the development of clinal treatment plans.

Numerous healthcare surveillance frameworks have been proposed in the literature, such as [4]–[6]. However, the existing cloud-enabled systems have several limitations that prevent their use in real-world applications. These include issues in latency (particularly problematic in time-critical healthcare applications), communication overhead between healthcare entities and cloud computing, and privacy concerns regarding the aggregated of sensitive data [7]. The EoT paradigm can reduce latency and bandwidth consumption through bringing computing power closer to the data source and delegating some of the time-critical computation tasks to edge devices while moving other computation-intensive tasks to cloud computing resources. With regard to the privacy issue, Personal Health Information (PHI) is the demographic and healthcare information collected by health professionals. PHI is considered to have one of the highest levels of sensitivity and is protected by the laws and regulations of the United States (US). The Health Insurance Portability and Accountability Act (HIPAA) [8] complies with strict regulations to

ensure the security and privacy of PHI and to prevent it from being misused.

In this paper, we develop an innovative and secure EoT framework for smart healthcare surveillance. We use clustering-based techniques to analyze biosignal data (e.g., electrocardiogram ECG data for heart healthcare) that is collected from IoT connected sensors for patients within both single and multiple smart communities. Community members include patients in hospitals, elderly people in Ambient Assisted Living (AAL) environments, and healthy people in smart homes who are concerned about their health. Hospitals, doctors and nurses are caregiver members who provide all types of patient monitoring assessments and maintain ubiquitous communication. In our framework, data privacy is ensued though Fully Homomorphic Encryption (FHE) that has ability to provide end-to-end privacy for members' data. Unlike previous approaches, FHE both protects stored data and performs analytic tasks in an encrypted domain. However, this can tremendously increase the size of the data that is being transferred from data source to the cloud. Our framework uses edge IoT devices to perform part of the analytic tasks, thereby enhancing the performance of encrypted analysis computations and reducing the size of transferred data through EoT framework to the cloud. In the future, the 5G wireless network can be used transfer massive amounts of encrypted data with an even lower latency rate in a cloud-enabled framework. An overview of our proposed framework is presented in Figure 1.

II. MOTIVATION

The main motivation of this paper is to build a secure and effective EoT framework for smart health surveillance that can be used in diversified smart healthcare platforms [9] using the advanced edge computing layer. Edge computing devices (e.g., IoT gateways) are located in a middle computing layer between IoT sensor devices and cloud computing. This edge layer is considered an extension of the CoT paradigm, which can significantly enhance the performance in various application domains, such as smart grid, healthcare and industry. The edge layer offers several advantages in our EoT healthcare surveillance framework, including real-time diagnosis services with very low latency, reducing the amount of data transmitted over networks (by processing part of the data in IoT edge devices), optimizing utilization of computational IoT devices and minimizing the cost of cloud recourses [10]. In the EoT paradigm, sensitive healthcare data is exposed to various external and internal attacks and possible leakage during exchange among different parties. In our framework, we adapt Fully Homomorphic Encryption (FHE) to ensure end-to-end data privacy and while it is stored in EoT databases or the cloud. FHE also performs analytic tasks in an encrypted domain without the need to decrypt data at any stage of processing. In our framework, edge IoT devices play a critical role to improve the processing response time and reduce the amount of encrypted data needed to be sent to cloud computing.

III. CONTRIBUTIONS

The main contributions of this paper are as follows.

- We develop a secure EoT framework for smart health surveillance. The framework can aggregate, monitor and perform real-time analysis of biosignal data. Clustering-based machine learning techniques are used to analyze and detect abnormality changes in biosignal data. The privacy of members' sensitive data is ensured by Fully Homomorphic Encryption (FHE) that can perform analytic services in an encrypted domain.
- We develop an innovative distributed approach for adapting clustering-based techniques in the EoT paradigm. This approach is capable of running independent analytic services in the edge layer that receives data for distributed IoT devices. This approach is applied in clustering-based techniques, including K-means (KMC) clustering and Fuzzy c-means (FCM) clustering techniques. However, it can also be extended to various machine learning techniques.
- We demonstrate a comprehensive evaluation of the proposed framework in terms of analysis performance and accuracy. The experimental outcomes show that our framework can rapidly accelerate performance while achieving a high level of accuracy of the overall analysis process in a secure manner.

IV. RELATED WORK

The integration of Internet of Things (IoT) and cloud computing paradigms (CoT) is a mega trend in next-generation technologies that can impact all aspects of daily living. The interconnection between smart objects and devices within the internet infrastructure can provide convenient solutions for various application domains, such as healthcare, smart grid and industrial control [11], [12]. In our framework, we focus on healthcare IoT applications. Healthcare represents one of more attractive domains [13] because the CoT paradigm has the potential to significantly advance healthcare technology and improve human health by remote health surveillance, early detection of chronic diseases and elderly care. Cloud computing empowers healthcare entities (e.g., patients and healthcare institutions) to move their data from billions of distributed healthcare IoT devices, including body sensors, diagnostic and imaging devices, to the cloud to significantly improve the quality of healthcare services and reduce costs [14].

The edge computing paradigm is introduced as a solution for some of the CoT paradigm limitations, including latency, bandwidth consumption and the large volume of transmitted data between IoT devices and cloud computing [15]. Several edge computing frameworks have been proposed for healthcare applications. Salahuddin *et al.* [2] developed a smart healthcare system that uses edge gateways devices as a bridge in an edge layer between public connected networks and Wireless Sensor Network (WSN). These smart gateways support data-driven decisions and notify caregivers in emergency cases. Yang and Gerla [16] introduced a personal

health monitoring that uses a smart phone as a gateway device that aggregates and sends healthcare data to processing servers through the bluetooth technology. The proposed system uses sensors attached to medical equipments to collect patient data and sends the data to the cloud, providing ubiquitous access. Mohapatra and Rekha [17] introduced a hybrid remote healthcare monitoring system that takes advantage of cloud computing resources for aggregating and uploading patient data through medical sensors, thus providing on-demand access privileges for both patients and caregivers. A ubiquitous healthcare system based on mobile gateways using ZigBee and Bluetooth where the gateways provide notification and analysis services of healthcare data [18]. Similarly, Park and Pak [19] introduce healthcare model that aggregates data through personal health devices by using USB, ZigBee and Bluetooth.

In this paper, we focus on privacy-preserving solutions for storing and processing sensitive healthcare data in the EoT paradigm. EoT is exposed to numerous external and internal malicious attacks and possible abuse by shared parties. Privacy and security mechanisms, including authentication, authorization, anonymisation, access control and cryptography techniques, can be used as defense layers to preserve the privacy of sensitive healthcare data [20]–[22]. During the early stages of healthcare development, anonymisation techniques are used to protect both users' identities and their healthcare data [23], [24]. Although anonymisation techniques are efficient for data processing, they can lead to unavoidable privacy breaches [25]. Access control mechanisms are considered to control access healthcare data among interacting entities, including healthcare institutions, data owners and laboratories and other caregivers. Hyped access control mechanisms, such as a role-based access control (RBAC), support the most comprehensive privacy-preserving healthcare systems, but even they have limitations. For example, RBAC can be used to identify family members' access permission in healthcare systems, but family members may require different access level for healthcare data which makes such mechanisms complicated to be used in healthcare domain [26]. Sun *et al.* [27] proposed an access control mechanism that allowed patients to provide leveled access permissions (e.g., doctors, laboratories, insurance companies) for their healthcare information. Although promising, it is suffered from delegation issues for different levels of access permissions, which incorrectly led to overlap between roles and lack of permissions on others [7].

Compared with the aforementioned techniques, cryptography-based solutions are considered to be the most convenient tools to preserve the privacy of stored and processed data in the EoT paradigm. They can be categorized into two main cryptographic techniques: Secure Multi-party Computation (SMC) and Homomorphic Encryption (HE). The former was introduced by Yao [28] in 1982 and its objective is to build a secure, shared environment among several parties to perform certain functions on encrypted data in a secure manner. Most existing SMC applications are based on semi-honest models, where adversaries follow the protocol but they are also able to obtain information during communication. In their current form, the SMC applications based on malicious adversarial models are not suitable for real-time monitoring and analysis applications. HE has the ability to carry out computations in an encrypted domain and has two main approaches: Partially Homomorphic Encryption (PHE) and Fully Homomorphic Encryption (FHE). The PHE approach is suitable for a limited number of real-world applications due to its limited arithmetic operational capabilities. Furthermore, it carries a large communication overhead to perform computations that cannot be conducted in an encrypted domain. In contrast, the FHE approach has the ability to perform an unlimited number of encrypted computations and the encrypted data can be sent only once to the Cloud Service Provider (CSP). Furthermore, all the clustering-based mining computation tasks are performed in a secure manner, without interacting with data owners or any Trusted Third Parties (TTP). However, early FHE schemes are considered as computationally intensive and remain impractical to use in real-world applications. The recently developed FHE schemes have been improved from a performance efficiency perspective. Recently, the Brakerski-Gentry-Vaikuntanathan (BGV) scheme [29] and its implementation HElib library have been identified as promising candidates for a practical FHE scheme. We take advantage of the FHE approach to develop a privacy-preserving EoT framework for smart healthcare surveillance while eliminating the majority of vulnerabilities in the existing work.

V. SECURE EoT FOR SMART HEALTHCARE SURVEILLANCE FRAMEWORK

In this section, we firstly provide an overview of our framework architecture. We then briefly describe the Fully Homomorphic Encryption (FHE) cryptosystem that is applied to ensure the privacy of users' sensitive data while it is stored and processed at different stages in our framework. Then, we explain how we developed a distributed analysis approach for clustering-based techniques in EoT paradigm.

A. SYSTEM ARCHITECTURE

We have developed a secure abnormality detection framework for smart healthcare communities based on the FHE technique. The architecture of our framework relies on different entities interacting to achieve certain analysis tasks, beginning with data aggregation, storage and finishing with analysis tasks, all in a privacy-preserving manner. The architecture has three main entities, as follows.

1) **Community Members (CM):** This includes healthy people, elderly patients and hospital patients within the smart community. Wired/wireless sensors are used to aggregate biosignals data from CMs which are then sent to the cloud-enabled storage upon encryption.
2) **IoT gateway:** This is a smart IoT entity for local analysis processing within each smart community.

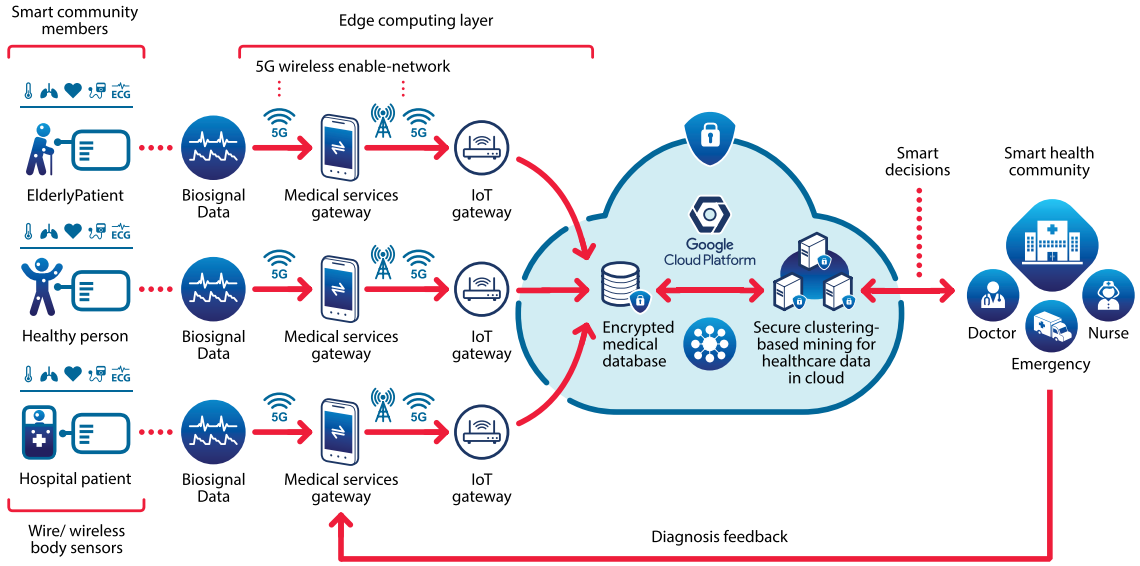

FIGURE 3. System architecture showing different components of proposed privacy-preserving change detection and abnormality prediction model in the cloud.

Collected data from CMs are analyzed for local diagnosis feedback within each individual community. Then each smart gateway sends the encrypted data to the cloud storage for further analysis with other smart communications data.

3) **Cloud-enabled Database (CD):** This is a cloud-based storage for CM's healthcare data from all smart communities in an encrypted form.
4) **Abnormality Detection Model (ADM):** This is the analysis engine in the system where aggregated encrypted data from multiple smart communities is analyzed in its encrypted form.

The entities work together to aggregate, store and analyze biosignal data for abnormality detection purposes. After the encrypted data is sent to the CD entity, the ADM performs encrypted analysis tasks on encrypted data securely and in an independent manner. The CM can receive encrypted feedback results from the CD entity to be decrypted at the CM securely. Figure 3 illustrates the main entities of our framework and the data workflow among them.

B. FULLY HOMOMORPHIC ENCRYPTION (FHE)

FHE is a convenient privacy and security provision mechanism to protect data storage and processing in a cloud-enabled framework for two main reasons: 1) it ensure the end-to-end data privacy of patient data in a public cloud storage; and 2) it has encrypted-based computation capabilities that can perform data analysis tasks without interacting with any TTP. The FHE approach can be deployed in several cloud-based integrated service domains to protect the privacy of data owners, data and secure cloud-based frameworks that are exposed to various malicious activities. Numerous analytic-as-a-service platforms, such as power and smart grids, health and assisted living and industrial monitoring systems, can benefit from such a privacy-preserving approach to ensure data privacy while processing in a public cloud computing platform. The Brakerski-Gentry-Vaikuntanathan (BGV) Cryptosystem is a well-known and practical FHE scheme. This scheme is classified as asymmetric encryption, the security of which is linked to the difficulty of the ring-learning with errors (RLWE) problem [30]. Unlike PHE schemes, the BGV FHE scheme has the ability to perform unlimited arithmetic computations on encrypted data. The basic BGV functions can be shown as follows.

- **BGV.KeyGen()** $\rightarrow$ $(public_{key}, secret_{key})$**:** The outputs of key generation function are two BGV cryptosystem keys. $secret_{key}$ is used to encrypt ciphertexts upon being sent to public cloud computing storage while $public_{key}$ is used to perform arithmetic computations on encrypted data.
- **BGV.Encryption()** $\rightarrow$ $(Ciphertext \;\; c)$**:** The encryption function output is a ciphtertext c that is homomorphically encrypted.
- **BGV.Decryption()** $\rightarrow$ $(Plaintext \;\; m)$**:** The decryption function output is a plaintext m.

The fact that the BGV scheme has homomorphic properties means that the following:

$$m1 \; op \; m2 = Dec(Enc(m1) \; op \; Enc(m2)) \quad \forall \, m1, m2 \in \mathbb{A}_p$$

where op is an unlimited number of arithmetic computations. The BGV scheme is built in order to provide computationally feasible implementations and certain convenient routines on top of the basic arithmetic operations.

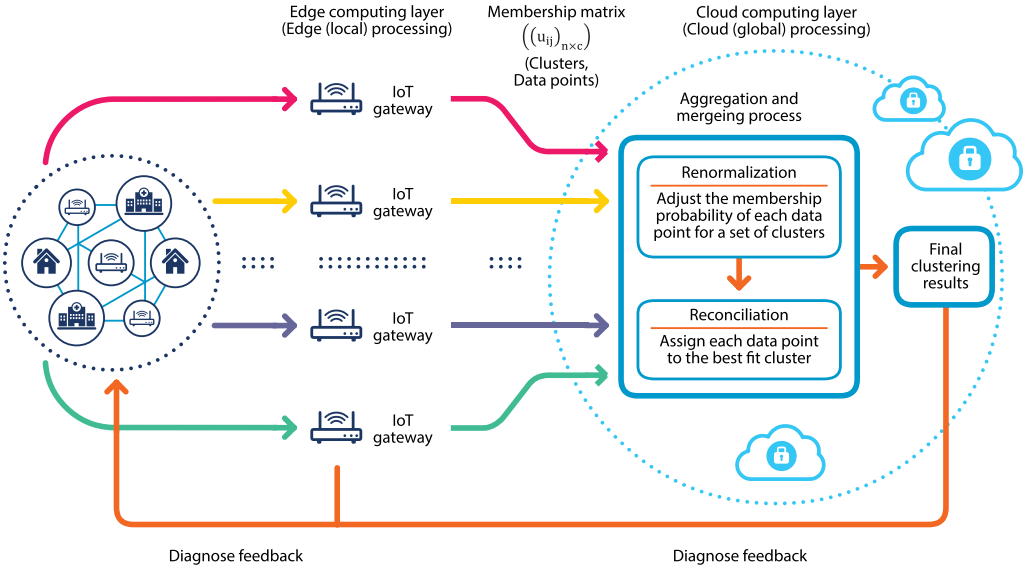

FIGURE 4. System architecture showing different components of proposed privacy-preserving change detection and abnormality prediction model in the cloud.

C. FULLY HOMOMORPHIC ENCRYPTION FOR FLOATING-POINT COMPUTATION

A limitation of the BGV fully homomorphic cryptosystem is the lack of floating-point arithmetic computations and that it can only support integer numbers. We introduced a convenient solution to overcome this limitation through applying the IEEE 754 standard for floating-point arithmetic to convert the representation of floating-point numbers to its integer representation. The IEEE 754 standard representation for a floating point number F occupies 32 bits (can be extended to 64 bits) which is arranged as follows. S is the sign bit (a 1-bit field), E is the exponent field (an 8-bit field) and M is the mantissa field (a 23-bit field).

The sign bit S is 0 for a positive number, and 1 for a negative number. The exponent field E is the actual exponent $+127$, so E should be treated as an unsigned value in the range [0, 255]. The mantissa field represents a number in the range [1.0, 2.0), except that the leading 1 is not encoded in M. Therefore, F can be mathematically expressed as follows:

$$F = (-1)^S \times 2^{E-127} \times (1 + M) \tag{1}$$

All required computations for the analysis process in our framework can be carried on an encrypted domain by using IEEE 754 floating-point representations. A detailed description of floating-point arithmetic computation is shown in our previous work [31].

D. SMART EoT HEALTHCARE SURVEILLANCE MODEL

Our developed EoT healthcare surveillance model relies on a distributed processing approach that takes into consideration the distributed nature of IoT healthcare devices. Two unsupervised clustering-based mining learning techniques, including K-Means Clustering (KMC) and Fuzzy C-Means clustering (FCMC), are used for healthcare surveillance. By using clustering-based mining, we can identify regular (normal) biosignal patterns. Any deviation from these normal patterns are considered as an abnormal (anomaly) pattern. We illustrate how the two clustering-based mining techniques, KMC (hard clustering) and FCMC (soft clustering), can be built in a distributed EoT framework for biosignals abnormality detection in a smart healthcare community.

1) DISTRIBUTED ANALYSIS APPROACH

The distributed approach has two stages: Level-1 "*Edge (local) processing stage*", and Level-2 "*Cloud (global) processing stage*". In the first stage, edge IoT devices (e.g., IoT gateways) analyze the sensed biosignal data within their specific ranges based on selected clustering-based techniques. The analysis result of this stage are reported as feedback for healthcare providers and data owners to identify any abnormality changes. Then, the result is forwarded to cloud computing for further global processing. The second stage merges aggregated results on different IoT edge devices through two steps: *Normalisation step* and *Conciliation step*. Figure 4 shows the workflow of the proposed distributed approach stages.

- **Edge (local) processing stage:** clustering-based techniques, including KMC and FCMC, are performed for a small set of data within each IoT device. We briefly describe KMC and FCMC techniques as follows.
  - **K-Means Clustering (KMC)**: The KMC algorithm [32] is an unsupervised clustering technique that classifies a set of data objects into different disjointed k clusters. KMC is also known as the hard clustering technique, where each data object belongs to a cluster with a shortest distance from the object to the cluster centroid. The square of the Euclidean distance is used to measure the distance

Algorithm 1 K-Means Clustering (KMC) Algorithm

Inputs: Encrypted data objects $(x_1), \ldots, (x_n)$
Outputs: Encrypted set of cluster centroids $(c_1), \ldots, (c_k)$
Initialization: Choose a random set of clusters centroids $(c_1), \ldots, (c_k)$ from a given data objects.
while Not converged **do**
  for all the data objects 1, .., n **do**
    for all the cluster centroids 1, .., k **do**
      calculated the distance between each data object and each cluster centroid.
    end for
    Assign each data object to a cluster centroid with the minimum Euclidean distance.
  end for
  for all the cluster centroids 1, .., k **do**
    calculated the average of each cluster based on the data objects that assign to it.
  end for
end while

Algorithm 2 Fuzzy C-Means Clustering (FCMC) Algorithm

Inputs: Encrypted data objects $(x_1), \ldots, (x_n)$
Outputs: Encrypted set of cluster centroids $(c_1), \ldots, (c_k)$
Initialization: Choose a random set of clusters centroids $(c_1), \ldots, (c_k)$ from a given data objects.
while $||U^{k+1} - U^k|| >= \beta$ **do**
  for all the data objects 1, .., n **do**
    for all the cluster centroids 1, .., k **do**
      Calculated the distance between each data object and each cluster centroid.
      Calculate the membership value.
    end for
  end for
  for all the cluster centroids 1, .., k **do**
    Update each cluster centroid based on the data objects that assign to it.
  end for
end while

between the data objects and cluster centroids. FHE is applied to protect data and KMC analysis tasks in our EoT framework. We illustrate the main KMC steps as follows.

1) Let $(x_1), \ldots, (x_n)$ be the number of encrypted data objects. The algorithm selects $(c_1), \ldots, (c_k)$ number of cluster centroids.
2) Calculate the Euclidean distance (d_{ij}) between each data object (p_i) and each cluster centroid (c_j).
3) Assign each data object (p_i) to each cluster centroid (c_j) based on the shortest Euclidean distance between each data object and the cluster centroids.
4) Recalculate the cluster centroids $(c_1), \ldots, (c_k)$ by obtaining the average of the data objects that are assigned to each cluster centroid (c_j).
5) Repeat steps 2,3 and 4 until the shortest centroids converge. Algorithm 1 illustrates the basic pseudo-code of the FH-KMC approach.

– **Fuzzy C-means clustering (FCMC) algorithm:** The FCMC algorithm [33] is an unsupervised soft clustering technique and is one of most popular fuzzy clustering techniques because it can retain much more information than hard clustering approaches, especially in the healthcare domain. In the FCMC algorithm, each data object belongs to each cluster centroid with a certain degree, which is called a membership value. Similar to the KMC algorithm, we apply FHE to preserve data privacy and to perform FCMC analysis tasks in a secure manner. We illustrate the main FCMC steps as follows.

1) Let $(x_1), \ldots, (x_n)$ be the number of encrypted data objects. The algorithm selects $(c_1), \ldots, (c_k)$ number of cluster centroids.
2) Calculate the Euclidean distance (d_{ij}) between each data object (p_i) and each cluster centroid (c_j).
3) Calculate the fuzzy membership value (μ_{ik}) that indicates the degree to which each data point (p_i) belongs to each cluster centroid (c_j).
4) Update the cluster centroids $(c_1), \ldots, (c_k)$.
5) Updated the fuzzy membership values of the data objects and then the cluster centroids based on steps 3 and 4 until $||U^{k+1} - U^k|| < \beta$, where U is the fuzzy membership matrix $(\mu)_{n*c}$ that contains the membership values between the data points and cluster centroids and β is the termination criterion value that is pre-determined. Algorithm 2 illustrates the basic pseudo-code of the FCMC algorithm.

– **Cloud (global) processing stage:** After the convergence of the edge analysis stage, analysis results of each edge IoT device (weight submatrices) are send to cloud computing where it can be represented as: $U_{N*C} = \sum_{j=1}^{C} \sum_{i=1}^{N} w_{ij}$ from each edge device. We assume that each of the submatrices (U_{N*C}) is normalized but it is not necessarily the case that all submatrices are reconciled. In the cloud processing stage, we unify the weight submatrices into a single global matrix. However, the resulting matrix $(U_{P*Q} = \sum_{j=1}^{Q} \sum_{i=1}^{P} w_{ij})$ needs to be normalized. Since the row sums of each data object of the submatrices (U_{N*C}) is 1.0, then the row sum of each data object in the global matrix (U_{P*Q}) will be a. This is the first issue that must be resolved by the global process stage.

The second issue is that we need to have one merged set of cluster centers C, but we actually have ab sets of C cluster centers, potentially leading to as many as $ab \times C$ cluster centers. We need to reconcile the cluster values such that so there are exactly Q cluster centers. To overcome these issues, we developed a merging process that has two steps, including *normalization step* and *conciliation step*.

* *Definition-1(Normalisation Step): If weight probabilities row sum of a data object* $V(i) = rowsum(j, M(i, j)) \neq 1.0$*, then divide each element of row i by* $V(i)$.
  Although the normalization step is necessary for global matrix (U_{P*Q}), it is not sufficient. To see why this is the case, consider a single data object (x) whose weight probabilities is being computed relative to two sets of clusters: $C1 = \{c_{11}, \ldots, c_{1q}\}$ and $C2 = \{c_{21}, \ldots, c_{2q}\}$. When both computations are done, (x) will have a first set of weight probabilities $\{\mu_{11}, \ldots, \mu_{1q}\}$ for $C1$ and a second set of weight probabilities $\{\mu_{21}, \ldots, \mu_{2q}\}$ for $C2$ where $|C1|$ and $|C2| = q$. In the KMC (univalent) case, a data object (x) can only be assigned to one cluster, so if one of the (c_{1j}) probabilities is 1 and also one of the (c_{2j}) probabilities is 1, then we have an inconsistency. Let us refer to these two clusters as c'_1 and c'_2. Since (x) cannot belong to both c'_1 and c'_2, we set one of these probabilities to 0 and leave the other probability as 1. To do this, we define a conciliation step for both KMC and FCMC algorithms as follows.
* *Definition-2(Conciliation Step - KMC): If a data object appears to belong to more than one cluster, then recalculate the metric distance of that data object to each of the clusters and choose the cluster to which it is closest. Set its weight probability to be* 1*, and all other weight probabilities to be zero, in row x of the membership matrix.*
  For FCMC algorithms, the conciliation step is more complex. In FCMC algorithms, we wish to identify the primary cluster to which a data object belongs. Typically, this cluster will have a probability weight value > 0.5. However, it may be the case that none of the probability weights is greater than 0.5. To address this, we define *conciliation step* for FCMC algorithm, which is used in conjunction with normalization.
* *Definition-3(Conciliation Step - FCMC): Let* μ_1 *be the largest probability weight, and* μ_2 *be the second largest weight of a given data point in the weight matrix* U_{P*Q}*. Suppose that after normalization, it is the case that both* $\mu_1 < 0.5$ *and* $\mu_1 < 0.5$*. Then, we adjust the weight of* μ_1 *to be* 0.5*, and adjust the weight of* μ_2 *to be* $\mu_2 + (0.5 - \mu_1)$.

Given these steps, we can describe the overview of the distributed analysis approach as follows. First, aggregated data Ui is processed in j edge IoT device independently based on selected clustering-based techniques where $1 \leq j \leq m$. Each subset weight matrix is defined as $Ui = \sum_{j=1}^{C} \sum_{i=1}^{N} \mu_{ij}$ and $\{Ui | 1 \leq i \leq m\}$. The global weight matrix $U_{P*Q} = \sum_{j=1}^{Q} \sum_{i=1}^{P} \mu_{ij}$ consists of merged set of overlapping submatrices $(U1, \ldots, Um)$. Second, given that there are m distributed edge IoT devices, the data is processed in distributed fashion among IoT devices and each of them runs independently to completion. Then, the cloud (global) processing is performed through the normalization and conciliation steps to obtain final analysis result.

VI. SECURITY ANALYSIS

This section discusses the privacy related considerations in our proposed EoT framework and also the security of the underlying cryptography mechanism. The main objective of the proposed EoT framework is to perform analysis tasks for healthcare surveillance on biosignal encrypted data without revealing users' sensitive information. The principal algorithms of concern are machine learning algorithms, such as clustering and anomaly detection.

The EoT healthcare surveillance framework is built based on a client /server architecture in which only the client represents a secure side of the architecture and has the ability to access to the unencrypted (plaintext) data. The threat model assumes the client generates a cryptography key-pair (private/public keys) and the public key is distributed to both edge IoT devices and the server side, which is in this case a Cloud Service Provider (CSP). The data is first tagged and then encrypted from its source by using the private key. This implies that the plaintext data is never transmitted to edge IoT devices or cloud servers and both perform analysis tasks only on an encrypted (ciphertext) data. This is possible because the applied machine learning algorithms (KMC and FCMC) can work based on Fully Homomorphic Encryption (FHE).

In our proposed framework based on FHE, a set of primitive operations, denoted by $\hat{\circledast}$ where $\hat{\circledast} = \{\hat{\oplus}, \hat{\ominus}, \hat{\otimes}, \hat{\oslash}, \hat{\ominus}, \hat{\ominus}\}$, can be performed on the encrypted data to get the same result as if it performed on the unencrypted data. Mathematically this implies that an FHE cryptography has the following property: $A \hat{\circledast} B = Enc(A) \hat{\circledast} Enc(B)$. The standard implementation of FHE only operates on the integer numbers, but we extended it to the floating-point numbers by applying IEEE 754 standard (32-bit - single precision floating point numbers). The IEEE 754 representation is developed without any modification on the standard FHE functionalities. Therefore, under the assumption that the underlying implementation of FHE operations is secure, then the analysis tasks of the proposed EoT framework are implemented in a secure manner. We assume, however, that any other data path, compute edge IoT devices or the cloud servers are completely insecure, and that an attacker can intercept and expose the

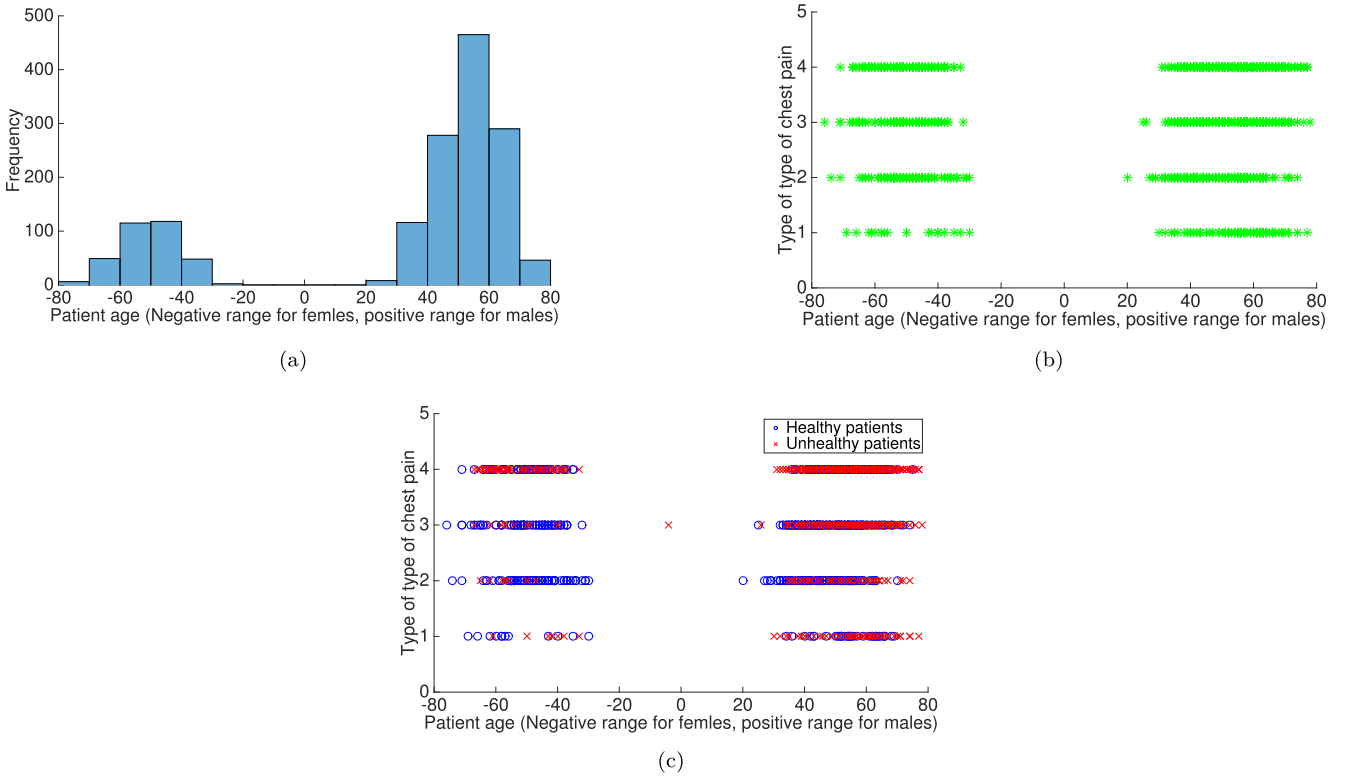

FIGURE 5. The distribution of male and female patients is shown in Figure 5a (top, left) and the distribution of male and female patients according to the type of chest pain is shown in Figure 5b (top, right). The abnormality detection for the presence of heart disease is shown in Figure 5c (bottom), where the red color refers to the male and female patients who had detected heart disease while the blue color refers to healthy patients. (a) The distribution of male and female patients. (b) The distribution of male and female patients according to the type of chest pain. (c) The abnormality detection for the presence of heart disease.

contents of a data communication path, or the memory of edge IoT devices, as well.

VII. EXPERIMENTAL EVALUATION

We evaluated the proposed secure EoT smart healthcare surveillance framework using the Google Cloud Platform (GCP). We used a real heart disease dataset from the University of California Irvine's (UCI) Machine Learning Repository [34] and synthetic datasets that varied in size and distribution to provide a comprehensive demonstration of the proposed framework. The real heart disease dataset contained information from 303 patients, with 75 features. We used the following fields: age; gender; chest pain indicator; and record label (0 = healthy, any other value = unhealthy). The synthetic datasets were produced based on the model in [35]. These contained two dimensional data for clustering purposes with varying size and distribution, which can simulate realistic scenarios. A data point is represented by a vector of two values $[x, y]$.

The main objective of this evaluation was to demonstrate the performance of the proposed EoT healthcare surveillance framework in terms of accuracy and execution time using the KMC algorithm, which can also be extended to FCMC algorithm. The experiment scenario assumes that data is analyzed locally in subsets based on distributed edge IoT devices and that analysis results are sent from all edge devices to cloud computing for further global analysis. For the real heart disease dataset, Figure 5a shows the distribution of patients in the dataset by gender and Figure 5b shows the distribution of patients based on the type of chest pain. We evaluate the existence of heart disease based on the KMC algorithm in Figure 5c.

Figure 6 shows the performance of the proposed framework based on KMC algorithms, with varying sizes of patient data to demonstrate the scalability of the distributed analytics approach. We used a Virtual Machine (VM) with a two-core Intel i5-2500 CPU running at 2.8 GHz with 16 GB and 64 bit linux operating system for local processing with similar machine type for global processing in our EoT framework. In the real heart disease dataset, performance of the developed model achieved linear speedup when the size of the data is increases. The distributed approach can significantly reduce the performance overhead in both local and global processing stages based on the number of VMs. For example, a single edge IoT device takes about 123 seconds to analyze data of 40 patients in the local processing stage but can be reduced to about 68 seconds if two edge IoT devices are used to process the same amount (size) of data.

The synthetic datasets analysis showed similar results, where an increase in computing resources can rapidly improved performance, as shown in Figure 8. The accuracy of distributed-based analysis for synthetic datasets was slightly

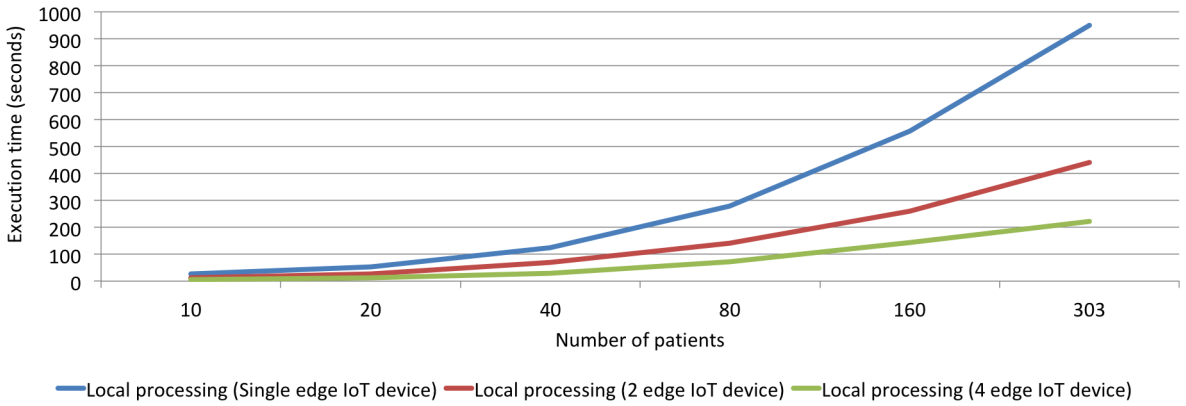

FIGURE 6. The performance of secure abnormality detection model in local analysis with 1,2 and 4 edge IoT devices with a varying size of data of the heart disease dataset.

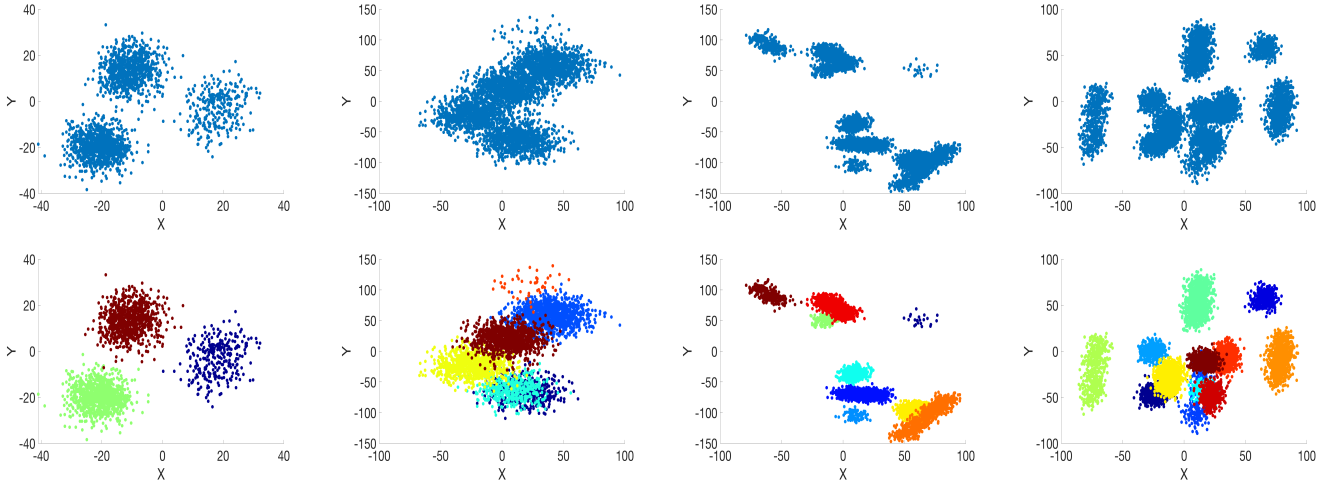

FIGURE 7. 2D synthetic datasets with varying size and distribution. The first row shows raw datasets with 2000, 4000, 6000 and 8000 data points respectively. The second row shows the datasets after performing analysis based on KMC algorithm.

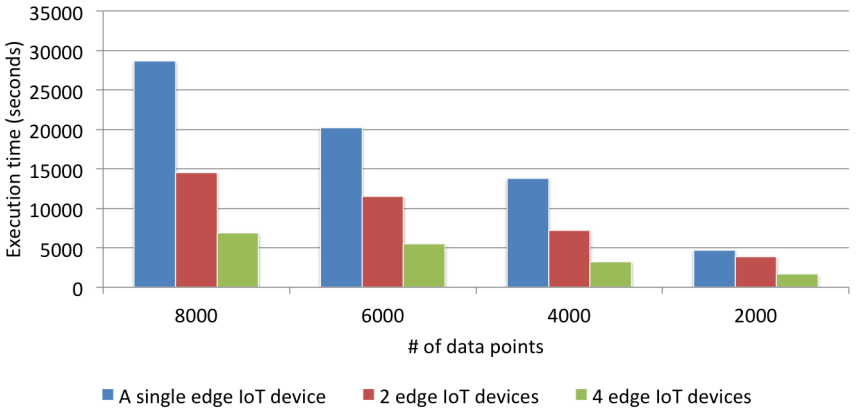

FIGURE 8. The performance of data analysis in local analysis with varying number of edge IoT devices and size of datasets.

low compared to centralized-based analysis (see Table 1). We used data labels to identify four quantities, including True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These quantities compute the accuracy level of synthetic datasets as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

TABLE 1. The analysis accuracy for synthetic datasets with a varying number of data points.

# of edge IoT devices	# of data points			
	2000	4000	6000	8000
1	0.9979	0.9950	0.9884	0.9832
2	0.9804	0.9850	0.9773	0.9602
4	0.8909	0.8870	0.8421	0.8019

We found that the accuracy varies between 0.29% to 8.90% based on the size of dataset. Moreover, the accuracy level is affected by the number of VMs between 1% to 18.13%. The performance evaluation shows significant results where the execution time improves between 49.38% to 76.06% for 8000 data points, and between 43.02% to 72.83% for 6000 data points, and between 47.42% to 76.42% for 4000 data points and between 17.41% to 57.20% for 2000 data points when varying the number of VMs from 1, 2 and 4 respectively, as shown in Figure 8.

VIII. CONCLUSION

The smart healthcare technology revolution is constantly developing better services for healthcare in smart communities. One of the greatest challenges facing smart healthcare systems is the protection of sensitive data. In this paper, we introduced a privacy-preserving EoT framework for smart healthcare surveillance. Clustering-based techniques are applied to analyze biosignal data in a secure manner and developed a distributed analysis approach, appropriate to the nature of the proposed EoT framework. We adapt Fully Homomorphic Encryption (FHE) because of its ability to store and analyze data in an encrypted form. In our framework, FHE ensured the privacy of outsourced biosignal data from its source and while it is processed in both edge IoT devices and cloud computing and encrypted analysis results can be retrieved by data owners and decrypted in a secure side. The 5G wireless network promises to address the challenges associated with the transmission of massive amounts of encrypted data between different entities in an EoT paradigm. Future research will improve the proposed framework in terms of further reducing homomorphic computational overheads and enhancing its capabilities to be extended for more advanced data mining models.

REFERENCES

[1] D. He, R. Ye, S. Chan, M. Guizani, and Y. Xu, "Privacy in the Internet of things for smart healthcare," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 38–44, Apr. 2018.

[2] M. A. Salahuddin, A. Al-Fuqaha, M. Guizani, K. Shuaib, and F. Sallabi, "Softwarization of Internet of things infrastructure for secure and smart healthcare," *Computer*, vol. 50, no. 7, pp. 74–79, Jul. 2017.

[3] N. Verba, K. M. Chao, A. James, D. Goldsmith, X. Fei, and S. D. Stan, "Platform as a service gateway for the Fog of Things," *Adv. Eng. Inform.*, vol. 33, pp. 243–257, Aug. 2017.

[4] Y. Oh, J. Han, and W. Woo, "A context management architecture for large-scale smart environments," *IEEE Commun. Mag.*, vol. 48, no. 3, pp. 118–126, Mar. 2010.

[5] Y. Zhu, N. M. Nayak, and A. K. Roy-Chowdhury, "Context-aware activity recognition and anomaly detection in video," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 91–101, Feb. 2013.

[6] A. R. M. Forkan, I. Khalil, A. Ibaida, and Z. Tari, "BDCaM: Big data for context-aware monitoring—A personalized knowledge discovery framework for assisted healthcare," *IEEE Trans. Cloud Comput.*, vol. 51, no. 4, pp. 682–641, Oct./Dec. 2015.

[7] M. A. Sahi *et al.*, "Privacy preservation in e-healthcare environments: State of the art and future directions," *IEEE Access*, vol. 6, pp. 464–478, 2017.

[8] Cosmetics and Fragrance Marketing and Management Services. (1996). *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. [Online]. Available: http://www.cms.hhs.gov/hipaa/

[9] A.-M. Rahmani *et al.*, "Smart e-Health Gateway: Bringing intelligence to Internet-of-Things based ubiquitous healthcare systems," in *Proc. 12th Annu. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2015, pp. 826–834.

[10] N. Bessis and C. Dobre, *Big Data Internet Things: A Roadmap for Smart Environments*, vol. 546. New York, NY, USA: Springer, 2014.

[11] J. Höller, D. Boyle, S. Karnouskos, S. Avesand, C. Mulligan, and V. Tsiatsis, *From Machine-to-Machine to Internet Things*. Amsterdam, The Netherlands: Elsevier, 2014.

[12] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the internet of things," *IEEE Internet Comput.*, vol. 14, no. 1, pp. 44–51, Jan./Feb. 2010.

[13] G. Yang *et al.*, "A health-IoT platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box," *IEEE Trans Ind. Informat.*, vol. 10, no. 4, pp. 2180–2191, Nov. 2014.

[14] C. Doukas and I. Maglogiannis, "Bringing iot and cloud computing towards pervasive healthcare," in *Proc. 6th Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, Jul. 2012, pp. 922–926.

[15] O. Osanaiye, S. Chen, Z. Yan, R. Lu, K. Choo, and M. Dlodlo, "From cloud to fog computing: A review and a conceptual live VM migration framework," *IEEE Access*, vol. 5, pp. 8284–8300, 2017.

[16] S. Yang and M. Gerla, "Personal gateway in mobile health monitoring," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2011, pp. 636–641.

[17] S. Mohapatra and K. S. Rekha, "Sensor-cloud: A hybrid framework for remote patient monitoring," *Int. J. Comput. Appl.*, vol. 55, no. 2, pp. 7–11, Jan. 2012.

[18] T. H. Laine, C. Lee, and H. Suk, "Mobile gateway for ubiquitous health care system using ZigBee and bluetooth," in *Proc. 8th Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, Jul. 2014, pp. 139–145.

[19] K. Park and J. Pak, "An integrated gateway for various PHDs in U-healthcare environments," *BioMed Res. Int.*, vol. 2012, May 2012, Art. no. 954603.

[20] J. L. Fernández-Alemán, I. C. Señor, P. Á. O. Lozoya, and A. Toval, "Security and privacy in electronic health records: A systematic literature review," *J. Biomed. Informat.*, vol. 46, no. 3, pp. 541–562, Jun. 2013.

[21] C. Camara, P. Peris-Lopez, and J. E. Tapiador, "Security and privacy issues in implantable medical devices: A comprehensive survey," *J. Biomed. Inform.*, vol. 55, pp. 272–289, Jun. 2015.

[22] A. Abbas and S. U. Khan, "A review on the state-of-the-art privacy-preserving approaches in the e-health clouds," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1431–1441, Apr. 2014. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/6714376/

[23] B. Riedl, T. Neubauer, G. Goluch, O. Boehm, G. Reinauer, and A. Krumboeck, "A secure architecture for the pseudonymization of medical data," in *Proc. 2nd Int. Conf. Availability, Rel. Secur.*, Apr. 2007, pp. 318–324.

[24] A. Pfitzmann and M. Hansen, "Anonymity, unlinkability, unobservability, pseudonymity, and identity management-a consolidated proposal for terminology," Citeseer, Tech. Rep. v0.30, 2005. [Online]. Available: https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.30.pdf

[25] J. H. Abawajy, M. I. H. Ninggal, and T. Herawan, "Privacy preserving social Network data publication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1974–1997, 3rd Quart., 2016.

[26] H. A. J. Narayanan and M. H. Günecş, "Ensuring access control in cloud provisioned healthcare systems," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2011, pp. 247–251.

[27] J. Sun, Y. Fang, and X. Zhu, "Privacy and emergency response in e-healthcare leveraging wireless body sensor Networks," *IEEE Wireless Commun.*, vol. 17, no. 1, pp. 66–73, Feb. 2010.

[28] A. C. Yao, "Protocols for secure computations (extended abstract)," in *Proc. Proc. 23rd Annu. Symp. Found. Comput. Sci.*. Nov. 1982, pp. 160–164.
[29] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(Leveled) fully homomorphic encryption without bootstrapping," in *Proc. 3rd Innov. Theor. Comput. Sci. Conf.*. New York, NY, USA, Jul. 2012, pp. 309–325. doi: 10.1145/2090236.2090262.
[30] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," *J. ACM*, vol. 60, no. 6, p. 43, 2013. doi: 10.1145/2535925.
[31] A. Alabdulatif, I. Khalil, H. Kumarage, A. Y. Zomaya, and X. Yi, "Privacy-preserving anomaly detection in the cloud for quality assured decision-making in smart cities," *J. Parallel Distrib. Comput.*, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731517303428. doi: 10.1016/j.jpdc.2017.12.011.
[32] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1. 1967, pp. 281–297.
[33] F. Höppner, *Fuzzy Cluster Analysis: Methods for Classification, Data Anals Image Recognite*. Hoboken, NJ, USA: Wiley, 1999.
[34] M. Lichman. (2018). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml
[35] N. Fachada, M. A. T. Figueiredo, V. V. Lopes, R. C. Martins, and A. C. Rosa, "Spectrometric differentiation of yeast strains using minimum volume increase and minimum direction change clustering criteria," *Pattern Recognit. Lett.*, vol. 45, pp. 55–61, Aug. 2014.

ABDULATIF ALABDULATIF received the B.Sc. degree in computer science from Qassim University, Saudi Arabia, in 2008, and the M.Sc. and Ph.D. degrees in computer science from RMIT University, Australia, in 2013 and 2018, respectively. He is currently an Assistant Professor with the College of Computer, Qassim University. His research interests include applied cryptography, cloud computing, data mining, and remote healthcare.

IBRAHIM KHALIL received the Ph.D. degree from the University of Bern, Switzerland, in 2003. He has several years of experience in Silicon Valley companies. He was with EPFL, the University of Bern, and Osaka University, Japan. He is currently an Associate Professor with the School of Computer Science and IT, RMIT University, Melbourne, Australia. His research interests include scalable computing in distributed systems, e-health, wireless and body sensor networks, biomedical signal processing, remote health care, network and data security, and secure data analytics and privacy.

XUN YI is currently a Professor with the School of Science, RMIT University, Australia. He has published more than 150 research papers in international journals, such as the IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATION, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON CIRCUIT AND SYSTEMS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGIES, IEEE COMMUNICATIONS LETTERS, and *IET Electronics Letters*, and conference proceedings. Recently, he has led a few of the Australia Research Council Discovery Projects. His research interests include applied cryptography, computer and network security, mobile and wireless communication security, and privacy-preserving data mining. He has ever undertaken program committee members for more than 30 international conferences. Since 2014, he has been an Associate Editor for the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING.

MOHSEN GUIZANI (S'85–M'89–SM'99–F'09) received the B.S. (Hons.) and M.S. degrees in electrical engineering and the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He was the Associate Vice President of Graduate Studies at Qatar University and the Chair of the Computer Science Department, Western Michigan University, and the Computer Science Department, University of West Florida. He also served in academic positions at the University of Missouri–Kansas City, the University of Colorado Boulder, and Syracuse University. He is currently a Professor and the ECE Department Chair with the University of Idaho, USA. He is the author of nine books and more than 500 publications in refereed journals and conferences. He has guest edited a number of special issues in the IEEE journals and magazines. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He has served as a member, the Chair, and the General Chair of a number of international conferences. He is a Senior Member of ACM. He received three teaching awards and four research awards throughput his career. He received the 2017 IEEE Communications Society Recognition Award for his contribution to outstanding research in wireless communications. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the TAOS Technical Committee. He serves on the editorial boards of several international technical journals. He has served as the IEEE Computer Society Distinguished Speaker, from 2003 to 2005. He is currently the Editor-in-Chief of the *IEEE Network Magazine*. He is also the Founder and the Editor-in-Chief of *Wireless Communications and Mobile Computing* journal (Wiley).