A homomorphic encryption-based system for securely managing personal health metrics data

R. Bocu
C. Costache

Hardware and software solutions for the collection of personal health information continue to evolve. The reliable gathering of personal health information, previously usually possible only in dedicated medical settings, has recently become possible through wearable specialized medical devices. Among other drawbacks, these devices usually do not store the data locally and offer, at best, limited basic data processing features and few advanced processing capabilities for the collected personal health data. In this paper, we describe an integrated personal health information system that allows secure storage and processing of medical data in the cloud by using a comprehensive homomorphic encryption model to preserve data privacy. The system collects the user data through a client application module, typically installed on the user's smartphone or smartwatch, and securely transports the data to the cloud backend powered by IBM Bluemix. The data are stored by the IBM Cloudant infrastructure, while the homomorphic processing of the encrypted data is performed using the Apache Spark service, which is also made available by the IBM Bluemix platform. The event-based handlers are triggered by the IBM OpenWhisk programming service. The initial prototype has been tested using a real-world use case, which is described.

Introduction

The pervasiveness of personal mobile devices fundamentally changes the way personal health information is collected. These devices usually feature a comprehensive set of sensors, which include various biomedical sensing components. This indicates that the individual subjects' physiological parameters may be conveniently monitored, while the personal health information is collected and is used for medical and other related purposes. This process generates large amounts of personal health information. Because of the limited storage and computational capabilities of these devices, the local processing of the collected data is not suitable. Therefore, the data must be stored and processed on external systems. Thus, the sensitive nature of the medical data requires safe and anonymous handling. Here, the safety encompasses two perspectives. First, the communication channel that connects the user-side mobile device to the processing and storage backend should transport the data in a secure fashion. Second, the backend should process the collected data without having access to the individual's identity or personal information. The system we describe in this paper uses a comprehensive homomorphic encryption model to preserve the data privacy. The homomorphic encryption allows for computations to be performed on encrypted data without exposing the raw data. The results of the computations are also encrypted and completely preserve the bit values of the plaintext data [1], and when they are decrypted, they match the results of the same operations performed on plaintext.

This kind of privacy-preserving data processing is useful in the context of the system that is described in this paper. We also suggest that this is a necessary approach when working with personal health and physiological information in the field of mobile sensing and eHealth applications. In the next several paragraphs, we summarize the related

Digital Object Identifier: 10.1147/JRD.2017.2755524

work done in the domain of secure personal information processing systems.

In [2–5], several encryption schemes are considered, to ensure data privacy, but all data aggregation operations are performed at the client side. The *Sum aggregate* and *Min aggregate* operations are achieved through an additive homomorphic encryption scheme in [2]. These operations perform calculations on a set of values and output a single value. Furthermore, it is relevant to note that they are versatile and efficient operations, which are used in other contexts that involve the aggregated processing of multiple data sets, such as the relational databases queries that consider multiple table columns. In a similar manner, in the current approaches, all computations are conducted on the client devices. The homomorphic encryption-related operations are usually resource intensive from a computational perspective. Thus, the current approaches are largely infeasible in the context of the use case that is described in this paper. The papers [3] and [4] describe approaches that are based on the offloading of data processing to the cloud. In [5], the authors propose a privacy-preserving sum aggregation, which also places a considerable computational load on the client devices. The homomorphic encryption scheme, which is proposed in [6], is only capable of performing the sum operation, which is insufficient for our needs.

The concept of verifiable computation was introduced by Gennaro et al. [7]. This enables one or several mobile client devices with constrained resources to delegate the computation of a function to one or more workers. At the same time, it is possible for the client to verify the correctness of the computed results. Benabbas et al. [8] describe the first verifiable computation model considering a plaintext input. Fiore and Gennaro [9] introduce a homomorphic data aggregation scheme in connection to eHealth systems. Nevertheless, this model cannot ensure the correctness of the computation's results, which is essential for SafeBioMetrics, the system described in this paper. Additionally, Fiore and Gennaro [9] propose a publicly verifiable computation scheme considering large polynomials and matrices, whereas Papamanthou et al. [10] introduce a verifiable delegated processing scheme, considering set structures and operations like set union, set intersection, and set difference. These schemes are applicable only to plaintext input data.

Guo et al. [11] design a verifiable computation scheme over encrypted input data in the context of mHealth (mobile health) systems. In [12], the concept of an accumulation tree is introduced in order to verify the results of geographical proximity tests. Furthermore, Fiore et al. [13] report on the efficient results that they achieved concerning the verifiable computation over encrypted input data. In the current approaches, all computations are conducted at the client side, which would be infeasible for the resource-constrained mobile devices that the SafeBioMetrics system considers.

Although the advantages of the cloud-based data storage and processing are obvious, some drawbacks are immediately discernible [14, 15]. The assurance of the security of private data is one of the most important aspects, which may have challenges for cloud services providers. These providers usually implement multiple layers of fairly sophisticated security mechanisms, but the plaintext data can still be accessed by an unauthorized entity considering various intrusion techniques. Thus, it is natural to encrypt the data before transferring it to the cloud and fetch it back through keyword-based search over encrypted data. The existing similar approaches significantly increase the computation overhead on the client devices. This is particularly valid in the case of the personal mobile devices that collect the personal data for the SafeBioMetrics system. In general, the existing approaches [16–19] do not ensure the security of the data that is transferred to and collected by the cloud. The handling of personal health information (PHI) is subject to strict ethical guidelines and legal regulations. Thus, the SafeBioMetrics system includes unique architectural features that address all these requirements, which are presented in the following sections. Breiter and Behrendt [20] present the characteristics of the services and their lifecycle inside a cloud-based processing environment.

Since Gentry introduced the concept of homomorphic encryption in 2009 [1], a substantial amount of work has been devoted to the optimization of this ciphered computation model, which has proved to be computationally resource intensive, and thus applicable to only certain use cases that involve the presence of a powerful hardware infrastructure. Additionally, the execution time of the homomorphic encryption routines was high in the past. Thus, the homomorphic encryption's algorithmic apparatus has been improved during several successive iterations. In [21–23], the authors describe similar models, which improve the efficiency of the homomorphic encryption routines. It is also useful to note the scientific contributions that are reported in [24–29], because they add certain features to the initial set of algorithms. The algorithms that are presented in [30] are relevant for the implementation of the system's backend. Nevertheless, the comprehensive tests that we performed during the initial stages of the work, which is reported in this paper, proved that even these optimized versions of the homomorphic encryption model are not suitable for the resource-constrained mobile devices, which collect the personal health information. The following paragraph suggests the most relevant existing contributions regarding the cloud-based safe data processing.

The existing similar approaches may often be unsuitable, as they are, for the construction of an efficient system like

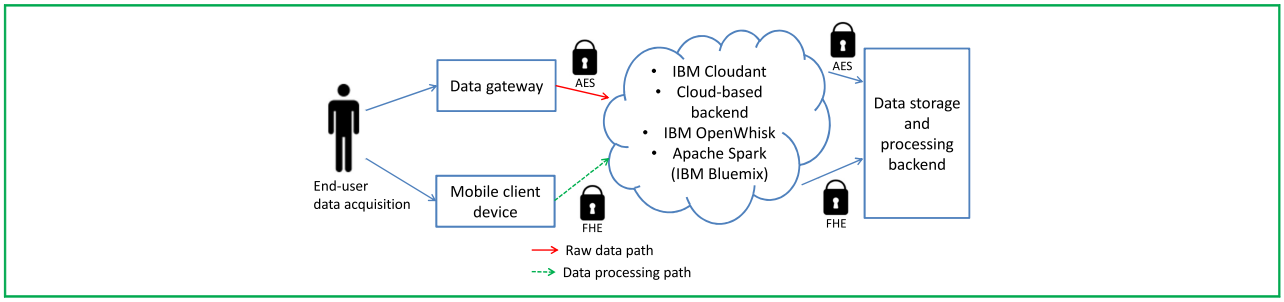

Figure 1

The SafeBioMetrics system architecture.

SafeBioMetrics, while considering all the four major perspectives: the biomedical data collection at the user's end, its transfer to the storage and processing backend, the proper and secure storage of this data, and its privacy-preserving processing. The SafeBioMetrics system is one of the few personal health information collection frameworks that combine the clear separation between the long-term data storage and data processing paths with the possibility to easily attach a variety of medical sensors and data collection devices at the client side. Additionally, the backend component is able to make use of cloud storage and processing services that ensure the system's scalability in the future. Therefore, the following sections present the system considering all the features that differentiate it from most existing contributions in the field.

The rest of the paper is structured considering three main sections. The next section describes the system considering its architectural components. The following section presents an assessment regarding the system's validity and appropriateness for the intended scope considering a real-world usage scenario. The last section of the paper is dedicated to the presentation of the subsequent system optimization plans. It also concludes the paper.

Remarks regarding the system architecture

The standard encryption schemes—for example, AES (Advanced Encryption Standard)—do not allow for arithmetic operations to occur over the encrypted data. In this case, the only allowed operation is the decryption through the usage of the secret decryption key. Thus, the standard encryption schemes define an environment that securely stores the data, but it is not able to compute it.

The fully homomorphic encryption (FHE) schemes offer the possibility to perform computation operations over the encrypted data, without considering the actual plain text significance of the computed data. The SafeBioMetrics system is based on the utilization of the fully homomorphic encryption schemes, which ensure that the personal health information (PHI) is safely collected and analyzed. The personal data is processed by the backend in its encrypted form. Thus, the level of privacy is "optimal," and the overall performance, as it is perceived by the end user, is not significantly affected.

The system architecture is presented in **Figure 1**. The data privacy is considered during the four main stages that define the data transmission pipeline. The first stage is represented by the data collection through each individual's wearable or portable device. Then, the second stage involves the data being safely transmitted to the data storage and processing backend. The third stage is defined by the actual storage of the patient data, while the last stage implies the safe data processing using the fully homomorphic encryption-based approach. The data storage and processing backend is deployed inside the IBM Bluemix [31] infrastructure. Thus, the collected data is efficiently stored using an IBM Cloudant-based [32] storage module. Furthermore, the necessary fully homomorphic encryption computations are performed using the Apache Spark platform, which is also included within the IBM Bluemix infrastructure. The processing events are intercepted and the proper actions triggered using the IBM OpenWhisk programming service [33]. The following sections offer more details on this storage and processing infrastructure.

The data transmission pipeline is intimately related to the cloud-based infrastructure considering the last two stages: the data storage and the safe data processing. The data processing results are sent back to the client devices on request. It is essential to note that the results' transmission is conducted with the data in a fully homomorphic encrypted format. This system model essentially changes the way most of the existent similar systems approach the personal health information collection and processing. Thus, the existing systems only encrypt the data at the storage backend, using a standard encryption scheme such as AES. Furthermore, the communication channels between the client devices and the server-side backend are secured, at best, with a standard encryption scheme. Therefore, it is impossible to ensure the long-term useful storage of the data, as no data processing can be safely performed while

protecting the personal health information. In other words, the SafeBioMetrics system proposes personal health monitoring options and data processing capabilities that have not been available with existing systems that rely on standard data encryption schemes.

The homomorphic encryption routines usually increase the amount of the processed data by a few orders of magnitude compared to the plain text original data. The architecture of the SafeBioMetrics system includes a data pipeline that is formed of two distinct data buses. In Figure 1, the top data bus is intended for storage purposes, while the bottom data bus is intended for the data transfers, which support the fully homomorphic encryption operations. Thus, the data processing path is used during the fully homomorphic encryption computations, while the raw data path is used during normal patient data retrieval. Consequently, this data transmission topology ensures that health status reports can be efficiently obtained by requesting only the necessary personal health data through the storage data bus. Then, only the relevant data chunk is processed and secured using the fully homomorphic encryption mechanism, while the results are safely transmitted to the requesting mobile client device.

Considerations regarding the optimization model

The fully homomorphic encryption model that SafeBioMetrics considers is extensively presented in [25]. It is called the Brakerski-Gentry-Vaikuntanathan (BGV) fully homomorphic encryption (FHE) scheme. We have thoroughly evaluated and tested the existing FHE schemes considering simulated test settings. We have found that most of the FHE schemes are prohibitively resource intensive, even in the case of capable hardware, especially because of the involved noise elimination (recrypt) operations, which are conducted after each multiplication operations [1, 25]. We have found the BGV to be the only feasible solution in the context of the SafeBioMetrics system. This is because the BGV scheme defines a leveled FHE scheme, which disregards the noise elimination operation. This approach considers a better noise management algorithm, which is called modulus-switching. This optimization is completely explained in [21]. It allows for cascaded homomorphic multiplications $(\mathrm{X_h})$ to be performed, while avoiding the risk to encounter decryption errors. This type of processing problem would be catastrophic for a system like SafeBioMetrics, as inaccurate personal health assumptions could be inferred, or even the mapping between the personal health information (PHI) and the respective individual could be rendered impossible. The following paragraphs describe the four types of FHE operations that are supported by the SafeBioMetrics system. In essence, the system introduces a parameter L (the *Level*), which must be determined before starting any computation instruction. The level L is calibrated considering the depth of the multiplication operations to be performed in the given computational context.

The first type of FHE operation that SafeBioMetrics supports is *homomorphic addition* $(+_\mathrm{h})$. This operation takes as operands two ciphertexts that correspond to a slot-wise XOR operation of the related plain text elements. The second type of FHE operation that SafeBioMetrics supports is the *homomorphic multiplication* $(\mathrm{X_h})$. This operation takes as operands two ciphertexts that correspond to a slot-wise AND operation of the related plain text elements. The multiplication increments by 1 the level L of the operation; thus, the depth of the multiplication operations determines the calibrated value of the level L. The third type of operation is *rotate* $(\lll_\mathrm{h}, \ggg_\mathrm{h})$, which essentially provides the possibility to rotate the data elements' slots. The concept of *slots* refers to the storage bits that determine the data elements processed by the *rotate* operation. The fourth operation, which is *select* $(\mathrm{sel_{mask}})$, has the role to correct the potentially altered slots (bits) of the data elements after the *rotate* operation. Therefore, the *select* operation has the role to preserve the data consistency during the fully homomorphic encryption process.

The optimized fully homomorphic encryption scheme

The SafeBioMetrics system relies on the efficient usage of the data storage and processing backend, which must safely process the personal health information. The efficient incorporation of the fully homomorphic encryption primitives into the SafeBioMetrics system relies on the utilization of the communication data path, which is illustrated in **Figure 2**. This figure suggests that each bit of the plaintext data is properly packed into the respective plaintext message. The ciphertext is generated through a fully homomorphic encryption model considering the steps contained by the top data path. The ability to process the encrypted data is the essential feature of this safe computational model. Thus, the bottom data processing path, which is represented in Figure 2, suggests that the input data is translated into a binary format, which is efficiently understood by the central processing unit. This is achieved using the computation $(f_c(.))$ and aggregation $(f_a(.))$ functions that are represented as the first elements of the bottom data processing path. Following this, the binary data is optimally processed using a parallel single instruction, multiple data (SIMD) model. The data processing is performed considering all the four types of operations that have already been introduced.

Current scope of the system

The system architecture is sufficiently flexible to accommodate any use case scenario that requires the client

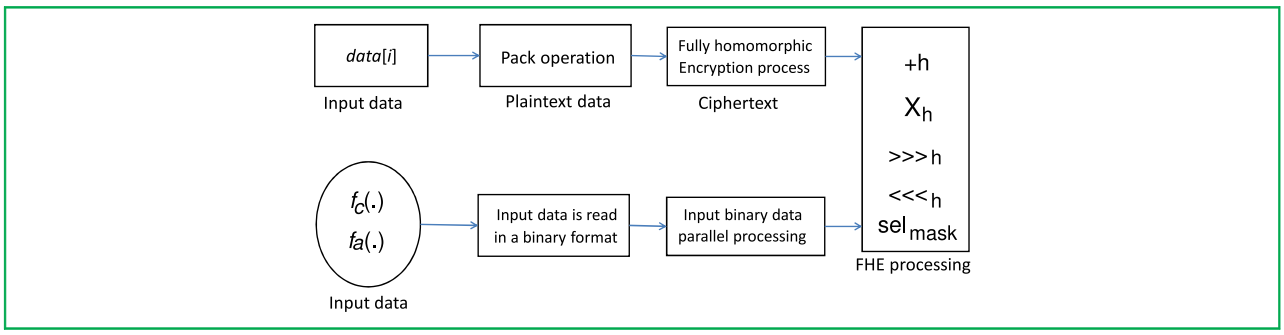

Figure 2

The SafeBioMetrics system basic dataflow.

data collection through a certain mobile device, and its safe transportation, storage and processing at the backend. The system can be configured to accommodate various existing and future mobile devices that perform the health data collection at the user's end. In order to demonstrate the validity and appropriateness for the intended scope of the SafeBioMetrics system, we focus on the collection of the cardiac rhythm data in the current version of the system. The data storage and processing backend has the ability to store the cardiac rhythm data and provide it on request to the entitled end user using the raw data path that is presented in Figure 1. Furthermore, the system has the ability to detect the delayed repolarization of the heart syndrome (DRHS) [27], and consequently report this condition. The following section describes the most relevant implementation details that are connected to this specific use case scenario.

Implementation details considering the specific use case

The practical computational performance of the SafeBioMetrics system, which relies on the usage of the optimized fully homomorphic encryption scheme, depends on two factors. First, the value of the level L, which determines the operation of the FHE scheme, is an important performance factor. Second, the system performance depends on the number of multiplication and rotation operations, which are computationally expensive. The multiplication operation is also relevant considering that it influences the calibration of the level L. The SafeBioMetrics system incorporates a series of improvements that pertain to the reduction of the level L, and also induces the minimization of the number of FHE operations. The system is designed to compute the necessary level L. This operation considers a number of N_{CT} ciphertexts, which have the role to encrypt an array, with n bits, that stores cardiac rhythm data.

Detection of the average heart rate

The computation of the average heart rate is based on the storage of the encrypted values in N_{CT} ciphertexts. The implementation optimization that is included in the SafeBioMetrics system considers two main types of improvements. The first one refers to the reduction of the computationally expensive multiplication operations. The second one pertains to the reduction of the computation operations depth, so that the level L is calibrated at the optimal level.

The addition operation is optimized considering two mechanisms. In the context of the SafeBioMetrics system, these two mechanisms are called the *additive compression* and the *prefixed parallel addition*. The additive compression transforms three data inputs (H, M, and F), each of them composed of n bits, into two outputs. These are represented by the A_{R} (addition result), and L_{OVER} (leftover). The $A_{\mathrm{R}} = H\Delta M\Delta F$, and $L_{\mathrm{OVER}} = [(H \times M)\nabla (H \times F)\nabla(M \times F)] \ll 1$. Here, Δ represents an additive single instruction multiple data (SIMD) operation, while the *nabla* operand (∇) also denotes a SIMD operation, which is performed on all n bits of the input data in a parallel fashion. The prefixed parallel addition has been designed and implemented considering the algorithmic model presented in [28].

The computation of the average heart rate considers the N_{CT} ciphertexts, which encrypt the input messages that are represented on n bits. Thus, the first step of this data flow involves the usage of the additive compression in order to transform N_{CT} ciphertexts into two ciphertexts. Following this, the resulting two ciphertexts are added through the prefixed parallel addition operation. We will report the system evaluation results, which prove that this approach is efficient in the context of the SafeBioMetrics system.

Detection of the delayed repolarization of the heart

The algorithmic model that implements the detection of this abnormal cardiac condition is based on the usage of the

scientific apparatus that is described in [29]. The main equation that is presented in [29] is optimized. Thus

$$\frac{T_{\mathrm{QT}}}{\sqrt{T_{\mathrm{RR}}}} > 475 \text{ ms} \Rightarrow {T_{\mathrm{QT}}}^2 > T_{\mathrm{RR}} \times 225,625, \quad (1)$$

$$\Rightarrow T_{\mathrm{QTH}} > T_{\mathrm{RRH}}. \quad (2)$$

Here, the expressions ${T_{\mathrm{QT}}}^2 = T_{\mathrm{QTH}}$ and $T_{\mathrm{RR}} \times 225,625 = T_{\mathrm{RRH}}$ are computed using the frontend, client-side devices that are represented in Figure 1. The T_{QT} and T_{RR} represent the time intervals that are measured and recorded during any electrocardiogram test. Essentially, T_{QT} represents the time taken for ventricular depolarization and repolarization, while T_{RR} measures the variability in the timing of the heartbeats. The subscript H denotes the homomorphic nature of the comparison, which detects the existence of the DRHS condition. We have applied an extensive set of calibration tests in order to design the optimized version of (1). In this context, the equation is optimized regarding the accuracy of the detection and the efficient usage of the computational resources. In the present form, the equation ensures that the SafeBioMetrics system accurately detects the DRHS condition with virtually no false positives, while running only the absolute necessary FHE operations. The data storage and processing backend aggregates the results of the individual comparisons. The client device simply requires a report from the backend considering a certain period of time. The client device decrypts the received result and checks for the presence of at least one bit that is equal to 1. If at least one such bit is found, then this is enough proof that during the given time period the comparison $T_{\mathrm{QTH}} > T_{\mathrm{RRH}}$ was true at least once. Consequently, the DRHS condition occurred with a significant probability at least once.

Detection of the minimum and maximum heart rates

The detection of minimum and maximum heartbeat rates is a functional requirement of the SafeBioMetrics system and is implemented considering the $f_c(.)$ function, which is graphically represented and put in context in Figure 2. This function has the role of converting the input data into a binary format, which is efficiently processed by the SafeBioMetrics system. It has already been shown that the comparison of two numbers, which are defined by n bits, produces a result that is also defined by n bits. If the first number is greater than the other number, then the result will contain a single bit of 1, and $n-1$ bits with a value of 0. If the first number is less than the other number, then the result contains only bits with a value of 0. Furthermore, the SafeBioMetrics system triggers a succession of rotate and select operations. The output of this subroutine is represented by a succession of n bits, each with a value of 1.

The problem of determining the minimum and the maximum values for the cardiac rate is reduced to the problem of determining the minimum and the maximum values from among N_{CT} ciphertexts, which encrypt an array of messages that are composed of n bits. Consequently, the correct computation of the minimum and maximum values for the cardiac rate is based on the successive application of the following functions: $\min(f_c(.))$ and $\max(f_c(.))$. In this context, the initial calibrated level L of the fully homomorphic encryption computation is calculated according to the following reference formula: $L > (\log_2 n + 2) \times \log_2 N_{\mathrm{CT}}$.

Evaluation of the practical system performance

System architecture

Let us recall that the SafeBioMetrics system architecture is graphically displayed in Figure 1. The system is able to accommodate any kind of mobile data collection device, provided that it is technically capable of gathering the required personal health information. The structural versatility and stability of the system, which is also suggested in Figure 1, is determined by the fact that only the client-side data collection devices may vary. Thus, any technically suitable client-side device is able to communicate with the system and send the data to the data storage and processing backend, without any hardware topology changes.

The client software module, which is installed on the user's mobile device, is capable of sending the collected data to the backend in real time. If the data connection is not available, then the collected data is stored locally, and immediately transferred to the backend as soon as a working data connection becomes available.

We have tested a variety of personal cardiac rate sensors, and we determined that the most accurate device is the Polar H7 [34]. Thus, it has been decided to use this device in order to collect the cardiac rate data. The personal health information, which is required to test the system's ability to detect the delayed repolarization of the heart syndrome (DRHS), is provided by a medical data set that includes 500 patients. The Polar H7 sensor has been applied on an experimental population sample, which is composed of 45 individuals. The Polar H7 device was used in order to assess the system's ability to properly collect, process, and store the data, while the dataset of 500 patients was used in order to assess the system's ability to detect the DRHS medical condition.

The system architecture is composed of the following software and hardware components. The cardiac rate data is collected by the Polar H7 personal sensor. The collected data is sent to each person's Android smartphone. The SafeBioMetrics client application is installed on the

Table 1 The performance metrics values.

Data reading interval	N_{CT}	*Level L*	$XFER_{IN}$ (GB)	$XFER_{OUT}$ (GB)	S_R	P_S
One min	2	12	4.8	2,886.3	32.1	0.54
Five min	12	15	5.9	1,147.8	39.4	0.24
Fifteen min	40	18	6.4	608.2	47.5	0.23
Thirty min	44	20	9.7	1,003.5	88.3	0.36
One hr	86	21	7.4	592.8	91.4	0.35
Three hr	258	24	8.9	201.6	101.2	0.37
Six hr	519	25	10.1	98.9	108.5	0.36
Twelve hr	1,021	26	11.2	42.7	117.4	0.39
One day	2,099	28	14.3	24.6	128.1	0.42

smartphone. It collects the data, which is transmitted by the personal sensor, properly encrypts it, and sends it to the data storage and processing backend, which is stored inside the IBM Bluemix infrastructure.

The software component of the backend is implemented using a modified version of the fully homomorphic encryption library that is described in [30]. This version includes the optimizations that have been described in the previous section. The backend is deployed to the Bluemix infrastructure using a proper buildpack. The Apache Spark Bluemix service is used in order to optimize the data access layer of the backend. The data that is collected from the client software modules is stored using the IBM Cloudant platform. This is a non-relational database engine, which proves to be suitable for the large amounts of data that the SafeBioMetrics system generates. The arrival of new health data in the cloud is detected by the IBM OpenWhisk Bluemix programming service. Following this, the proper event handlers are triggered, so that the newly arrived data is stored by the IBM Cloudant platform. Additionally, any data request that comes from the client devices is processed by the backend considering the algorithms and data flows that are presented in the previous sections.

Performance metrics

The system's performance assessment considers three relevant metrics. The first performance metric is represented by the *network capacity* that is used in order to transfer the data between the client software modules and the backend, in both directions. This metric is particularly relevant in the case of fully homomorphic encryption-enabled systems because of the large amount of data that must be transmitted over the network. Let us define two performance indicators in this context. Thus, the $XFER_{IN}$ represents the amount of data that is transferred from the client devices to the backend, while the $XFER_{OUT}$ denotes the amount of data that is transferred from the backend to the client devices.

The second performance metric is represented by the *storage ratio* (S_R). This assesses the amount of storage that is necessary to store one byte of plaintext data in a fully homomorphic encryption format. As an example, if $S_R = 500$, then it is clear that for one byte of plaintext data, there are a necessary 500 B in order to store the fully homomorphic encrypted byte.

The third performance metric is determined by the *processing speed* (P_S). This metric is defined through the following ratio: $P_S = P_{TO}/P_{IN}$. Here, the numerator represents the amount of time to send the data from the client device to the backend, while the denominator is the amount of time that is required by the backend to process the received data.

Results of the performance analysis

The system's practical performance assessment was conducted considering the dataset of 500 patients in order to detect the delayed heart repolarization condition. Additionally, the Polar H7 devices were applied to 45 individuals over a period of one month.

The values of the performance metrics recorded during the detection of the heart rates are presented in **Table 1**. The table columns are structured in such a way so that each of them offers essential information regarding the state of the system's basic parameters. Thus, the table columns present, in this order, the following set of system parameters and performance metrics values: the time period that is considered when reading the client-side input data, the number of the ciphertexts N_{CT}, the value of the calibrated level L, the amount of data that is transferred to the backend, the amount of data that is transferred from the backend, the values of the storage ratio parameter, and the values of the processing speed parameter. The performance results prove that the SafeBioMetrics system functions in a more efficient manner than existing similar approaches, such as the one that is presented in [30]. The similarity with other systems only pertains to the fully

Table 2 The performance assessment regarding the detection of the DRHS condition.

Data reading interval	N_{CT}	*Level L*	$XFER_{IN}$ (GB)	$XFER_{OUT}$ (GB)	S_R	P_S
One min	2	5	1.1	1,102.3	10.1	0.06
Five min	8	6	1.9	314.8	12.7	0.07
Fifteen min	22	8	2.4	108.5	14.5	0.10
Thirty min	41	10	2.8	83.5	16.3	0.11
One hr	85	11	3.1	69.8	29.4	0.32
Three hr	256	12	3.6	61.8	37.3	0.28
Six hr	517	14	4.8	30.2	42.5	0.29
Twelve hr	1,023	15	6.2	17.7	49.4	0.33
One day	2,079	17	7.3	9.6	53.6	0.35

homomorphic primitives, as the SafeBioMetrics system is one of the few that offers this platform for the personal health information collection in a perfectly safe and private fashion. Additionally, it is worth noting that the well-balanced (the amount of resources used is proportional to the amount of processed data) values of the performance metrics suggest that the system is scalable.

The system's performance metrics values, which pertain to the detection of the delayed heart repolarization medical condition, are presented in **Table 2**. It is relevant to mention that the values of the $XFER_{IN}$ and $XFER_{OUT}$ performance metrics demonstrate the suitability of the system's deployment in the cloud environment, which the data storage and processing backend uses. Thus, the cloud service providers usually charge for the uploaded $(XFER_{OUT})$ data stream, while the downloaded data $(XFER_{IN})$ is usually not monitored regarding the amount of the transferred data. Furthermore, the number of the ciphertexts (N_{CT}) is maintained at the minimum possible level, while the value of the level L is also computed in an optimal fashion. Additionally, it is relevant to note that the finer the time period granularity is, the greater the amount of the uploaded data becomes. Nevertheless, this performance metric's value increases according to an arithmetic model, and it is perfectly balanced relative to the quantity of the encrypted personal health information, which the backend provides as response to the client software module's requests.

Conclusion

The efficient collection of personal health data has been increasingly important during the past 15 years. As a consequence of the technological advancements, it has relatively recently become possible to collect the personal health data considering a continuous and unobtrusive monitoring process. Thus, the amount of the collected personal data is significant and poses numerous administrative and legal challenges. The main administrative challenge is connected to the necessity to efficiently extract relevant medical knowledge out of the vast amount of stored personal health information. The legal constraints principally pertain to the imperative requirement to safely collect, transfer, store, and process the personal health information.

This paper described the SafeBioMetrics system, which addresses the entire palette of requirements that have been mentioned. Considering its flexible and decoupled architecture, the system is capable of accommodating most of the existing and, with a high probability, future client-side data collection devices. In this context, the term *decoupled architecture* refers to the functional autonomy of the system's components. The system's validity and efficiency are tested considering real personal health information and a real-world use case scenario. The outcome of this assessment demonstrates that the SafeBioMetrics system is capable of sustaining a perfectly functional and secure data flow between the client data collection devices and the data storage and processing backend, in both directions. This result is worth mentioning because this is one of the few integrated systems that offer the full range of personal health information collection, storage, and processing functional capabilities. Furthermore, it is significant to mention that this contribution proves that the fully homomorphic encryption can be used in order to secure a complex system like the SafeBioMetrics. In this context, the complexity especially denotes the data buses and the related data processing modules, which are in charge of delivering and processing a large amount of data. Thus, the system proves to be perfectly usable considering real-world use case scenarios. Additionally, the backend system uses the processing and storage capabilities that are offered by the IBM Bluemix ecosystem, but it can be also deployed on other cloud platforms that offer similar data storage and processing services.

The existing software system's architecture is sufficiently flexible and decoupled, so that it can be easily extended with new functional capabilities. In the short term, we will add to the system the ability to perform cognitive computing operations at the backend side, which will be based on machine learning algorithms. These capabilities will be used in order to check that the monitored users rigorously follow a healthy lifestyle, which includes a certain amount of physical activity. This functional requirement is expressly defined by certain insurance policies providers, which offer a certain discount if the beneficiary abides to a healthy lifestyle. The cognitive computing capabilities will also ensure that the collected data is generated by the legitimate user, and not by a friend or even active pet. Furthermore, the fully homomorphic encryption processing primitives will be optimized in order to reduce the data transfer, data storage, and processing resources that the SafeBioMetrics system requires. The system has the potential to address current and future similar computational use case scenarios. Therefore, it will be maintained and continuously improved.

References

1. C. Gentry, *A Fully Homomorphic Encryption Scheme*. Stanford, CA, USA: Stanford Univ., 2009.
2. Q. Li, G. Cao, and T. La Porta, "Efficient and privacy-aware data aggregation in mobile sensing," *IEEE Trans. Dependable Secure Comput.*, vol. 11, no. 2, pp. 115–129, Mar. 2014.
3. R. Zhang, J. Shi, Y. Zhang, et al., "Verifiable privacy-preserving aggregation in people-centric urban sensing systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 268–278, Sep. 2013.
4. J. Zhou, Z. Cao, X. Dong, et al., "PPDM: Privacy-preserving protocol for dynamic medical text mining and image feature extraction from secure data aggregation in cloud-assisted e-healthcare systems," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1332–1344, Oct. 2015.
5. E. Shi, T.-H. H. Chan, E. G. Rieffel, et al., "Privacy-preserving aggregation of time-series data," in *Proc. Netw. Distrib. Syst. Security Symp.*, 2011, vol. 2, p. 4.
6. F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2010, pp. 327–332.
7. R. Gennaro, C. Gentry, and B. Parno, "Non-interactive verifiable computing: Outsourcing computation to untrusted workers," in *Proc. Annu. Conf. Adv. Cryptol.*, pp. 465–482, 2010.
8. S. Benabbas, R. Gennaro, and Y. Vahlis, "Verifiable delegation of computation over large datasets," in *Proc. Annu. Conf. Adv. Cryptol.*, 2011, pp. 111–131.
9. D. Fiore and R. Gennaro, "Publicly verifiable delegation of large polynomials and matrix computations, with applications," in *Proc. 2012 ACM Conf. Comput. Commun. Security*, 2012, pp. 501–512.
10. C. Papamanthou, R. Tamassia, and N. Triandopoulos, "Optimal verification of operations on dynamic sets," in *Proc. Annu. Conf. Adv. Cryptol.*, 2011, pp. 91–110.
11. L. Guo, Y. Fang, M. Li, et al., "Verifiable privacy-preserving monitoring for cloud-assisted mHealth systems," in *Proc. INFOCOM Conf.*, 2015, pp. 1026–1034.
12. G. Zhuo, Q. Jia, L. Guo, M. Li, et al., "Privacy-preserving verifiable proximity test for location-based services," in *Proc. IEEE Global Telecommun. Conf.*, 2015, pp. 1–6.
13. D. Fiore, R. Gennaro, and V. Pastro, "Efficiently verifiable computation on encrypted data," in *Proc. ACM SIGSAC Conf. Comput. Commun. Security*, 2014, pp. 844–855.
14. T. Jaeger and J. Schiffman, "Outlook: Cloudy with a chance of security challenges and improvements," *IEEE Security Privacy*, vol. 8, no. 1, pp. 77–80, Jan./Feb. 2010.
15. M. Kuzu, M. Saiful Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in *Proc. IEEE Int. Conf. Data Eng.*, 2012, pp. 1156–1167.
16. N. Cao, C. Wang, M. Li, et al., "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 222–233, Jan. 2014.
17. C. Orencik and E. Savas, "An efficient privacy-preserving multi-keyword search over encrypted cloud data with ranking," *J. Parallel Distrib. Databases*, vol. 32, no. 1, pp. 119–160, 2014.
18. J. Yu, P. Lu, Y. Zhu, et al., "Toward secure multikeyword top-k retrieval over encrypted cloud data," *IEEE Trans. Dependable Secure Comput.*, vol. 10, no. 4, pp. 239–250, Jul./Aug. 2013.
19. A. Boldyreva, N. Chenette, Y. Lee, et al., "Order-preserving symmetric encryption," in *Proc. 28th Conf. Theory Appl. Cryptograph. Techn.*, pp. 224–241, 2009.
20. G. Breiter and M. Behrendt, "Life cycle and characteristics of services in the world of cloud computing," *IBM J. Res. Dev.*, vol. 53, no. 4, pp. 3:1–3:8, 2009.
21. Z. Brakerski and V. Vaikuntanathan, "Efficient fully homomorphic encryption from (standard) LWE," in *Proc. Annu. Symp. Found. Comput. Sci.*, 2011, pp. 97–106.
22. M. van Dijk, C. Gentry, S. Halevi, et al., "Fully homomorphic encryption over the integers," in *Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Techn. Conf.*, 2010, pp. 24–43.
23. J. Coron, A. Mandal, D. Naccache, et al., "Fully homomorphic encryption over the integers with shorter public keys," in *Proc. 31st Annu. Conf. Adv. Cryptol. Conf.*, 2011, pp. 487–504.
24. C. Gentry, S. Halevi, and N. P. Smart, "Fully homomorphic encryption with polylog overhead," in *Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Techn. Conf.*, 2012, pp. 465–482.
25. Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "Fully homomorphic encryption without bootstrapping," in *Proc. Innov. Theor. Comput. Sci. Conf.*, 2012, pp. 309–325.
26. C. Gentry, A. Sahai, and B. Waters, "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based," in *Proc. Annu. Conf. Adv. Cryptol.*, 2013, pp. 75–92.
27. S. A. Immanuel, A. Sadrieh, M. Baumert, et al., "T-wave morphology can distinguish healthy controls from LQTS patients," *Physiol. Meas.*, vol. 37, no. 9, pp. 1456–1473, 2016.
28. P. Kogge and H. Stone, "A parallel algorithm for the efficient solution of a general class of recurrence equations," *IEEE Trans. Comput.*, vol. C-22, pp. 783–791, Aug. 1973.
29. H. C. Bazett, "An analysis of the time-relations of the electrocardiograms," *Ann. Noninvasive Electrocardiol.*, vol. 2, no. 2, pp. 177–194, 1997.
30. S. Halevi and V. Shoup, "Algorithms in HElib," in *Proc. Annu. Conf. Adv. Cryptol.*, 2014, pp. 554–571.
31. IBM Corporation. IBM Bluemix Cloud Infrastructure. [Online]. Available: https://www.ibm.com/cloud-computing/bluemix
32. IBM Corporation. IBM Cloudant Storage Service. [Online]. Available: https://cloudant.com
33. IBM Corporation. IBM OpenWhisk Service. [Online]. Available: https://developer.ibm.com/openwhisk
34. Polar H7 Sensor. [Online]. Available: https://www.polar.com/en/products/accessories

Received February 23, 2017; accepted for publication March 23, 2017

Razvan Bocu *Department of Mathematics and Computer Science, Transilvania University of Brasov, Brasov 500091, Romania (razvan.bocu@unitbv.ro).* Dr. Bocu received a B.S. degree in computer science, a B.S. degree in sociology, and an M.S. degree in computer science from Transilvania University of Brasov, Romania, in 2005, 2007, and 2006, respectively. He also received a Ph.D. degree from the National University of Ireland, Cork, in 2010. He is a Research and Teaching Staff Member in the Department of Mathematics and Computer Science at the Transilvania University of Brasov. He is author or coauthor of 23 technical papers, together with four books and book chapters. Dr. Bocu is an editorial reviewing board member of two technical journals in the field of information technology and biotechnology.

Cosmin Costache *IBM Romania, Brasov 500152, Romania (cosmin.costache@ro.ibm.com).* Dr. Costache received a B.S. degree in computer science, an M.S. degree in digital communication networks, and a Ph.D. degree from the Transilvania University of Brasov in 2005, 2009, and 2015, respectively. He is a Researcher and Software Engineer at IBM Romania in Brasov. He is author or coauthor of eight technical papers, together with four books and book chapters.