Secure e-Health System for the Integrated Management of Personal Health Data Collected by IoT Devices

Razvan Bocu[a], Maksim Iavich[b], Sergiy Gnatyuk[c], Dinara Ospanova[d], and Yuliia Sotnichenko[c]

[a] *Transilvania University of Brasov, 29 Eroilor ave., Brasov, 500036, Romania*
[b] *Caucasus University, 1 Paata Saakadze str., Tbilisi, 0102, Georgia*
[c] *National Aviation University, 1 Liubomyr Huzar ave, Kyiv, 03058, Ukraine*
[d] *Kazakh Humanitarian Juridical Innovative University, 11 Mengilik str., Semey, 070000, Kazakhstan*

Abstract

The definition of a smart city as a broad concept values the versatile acquisition, storage, and processing of relevant data for the city's community. In this context, health data occupies a privileged place. The reliable gathering of personal health information has become recently possible through wearable medical devices. These devices usually do not store the data locally and offer, in the most favorable case, limited basic data processing features, and virtually no advanced processing capabilities for the collected personal health data. This paper describes an integrated distributed e-Health system, which collects health data from the enrolled city residents, and allows secure storage and processing of medical data in the cloud by using a comprehensive encryption model to preserve the data privacy, which is based on the NTRU public-key cryptosystem. The correct assignment of the medical data to the respective person is verified by the usage of a hash-based digital signature mechanism. The system collects the user data through a client application module that is installed on the user's smartphone or smartwatch and securely transports it to the cloud backend. The homomorphic processing of the encrypted data is performed using the Apache Spark service. The event-based handlers are triggered by the IBM OpenWhisk programming service. The prototype has been tested using a real-world use case, which involves five hundred residents of Brasov City, Romania.

Keywords

Cloud-based processing and storage, homomorphic encryption, distributed computing, client-server systems, hash-based digital signature, wearable sensors.

1. Introduction

The consideration of personal mobile devices fundamentally modifies the way personal health information is acquired. These devices are featured with an array of sensors, which include several biomedical sensing components. Furthermore, the individual subjects' physiological parameters may be conveniently monitored, while the personal health information is collected and used for medical or research purposes. This process inherently produces large amounts of personal health information. Considering that wearable devices possess limited storage and computational capabilities, the local processing of the collected data is not practical. Consequently, the data must be stored and processed on external systems. The sensitive nature of medical data imposes safe and anonymous handling. In this context, safety comprises two perspectives. First, the communication channel that connects the user-side mobile device to the processing and storage backend should securely transport the data. Second, the backend should perform the required computations on the gathered data without knowing

Cybersecurity Providing in Information and Telecommunication Systems, January 28, 2021, Kyiv, Ukraine
EMAIL: razvan@bocu.ro (A.1); miavich@cu.edu.ge (B.2); s.gnatyuk@nau.edu.ua (C.3); dinara.ospanova@mail.ru (D.4); y.sotnichenko@gmail.com (C.5)
ORCID: 0000-0001-6577-1904 (A.1); 0000-0002-3109-7971 (B.2); 0000-0003-4992-0564 (C.3); 0000-0002-2206-7367 (D.4); 0000-0002-1281-9238 (C.5)

about the individual's identity or personal information. The system that is described in this paper proposes a computationally efficient NTRU-based fully homomorphic encryption system, which is capable to preserve personal data privacy. The homomorphic encryption allows the computations to be conducted on encrypted data, without revealing the original data. The results of the computations are also encrypted and completely preserve the significance of the original bit values [1, 2].

This type of data processing, which preserves privacy, is useful in the context of the system that is described in this paper. It can be also suggested that it is a necessary approach when working with personal health and physiological information in the field of mobile sensing and eHealth applications. The next paragraphs will summarize the related work done in the field of secure personal information processing systems.

2. Review of Related Approaches

The existing similar approaches [3–7] may often be unsuitable, as they are, for the construction of an efficient system like the one that is reported in this paper, while considering all the four major perspectives: the biomedical data is gathered at the user's end, it is transferred to the storage and processing backend where it is properly and securely stored, and the data is processed while completely preserving the privacy of the personal data. The e-Health 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 capability to easily attach a variety of medical sensors and data collection devices at the client-side. Additionally, the backend component can make use of cloud storage and processing services, which 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, which essentially makes it the only existing distributed system of this kind that processes large amounts of personal medical data in a timely manner, while completely preserving the data privacy.

The rest of the paper is structured considering three main sections. The next section describes the e-Health system considering its architectural components, with an emphasis on the NTRU-based encrypted data processing component. The following section presents an assessment regarding the system's validity and appropriateness for the intended scope by presenting the real-world usage scenario. The last section of the paper is dedicated to the description of the subsequent system optimization plans. It also concludes the paper.

3. System Architecture

The standard encryption models, such as AES (Advanced Encryption Standard), do not allow for arithmetic operations to be conducted over the encrypted data. In this case, the only valid operation is the plain decryption using the secret decryption key. Consequently, the standard encryption schemes define an environment that securely stores the data, but it is not able to allow for its processing to occur.

The Fully Homomorphic Encryption (FHE) schemes offer the possibility to perform computation operations over the encrypted data. The e-Health system is based on the utilization of the NTRU-based fully homomorphic encryption schemes, which ensure that personal health information (PHI) is securely gathered and analyzed. The personal data is processed by the backend in its encrypted form. Thus, the level of privacy is optimal, and the overall user experience performance is essentially not affected.

The system architecture is presented in Fig. 1. Data privacy is considered during the four phases that define the data transmission pipeline. The first phase is represented by the data collection that is performed through each individual's wearable or portable device. Then, the second phase involves the data is safely transmitted to the data storage and processing backend. The third phase is represented by the actual storage of the patient data, while the last phase implies safe data processing using the fully homomorphic encryption-based approach. The data storage and processing backend are deployed considering the IBM Bluemix infrastructure. Thus, the collected data is efficiently stored

using an IBM Cloudant-based storage module. Additionally, the required fully homomorphic encryption computations are performed using the Apache Spark platform, which is also provided by the IBM Bluemix infrastructure. The processing events are detected and the proper handlers are triggered using the IBM OpenWhisk programming service. The following sections offer more details on this storage and processing infrastructure.

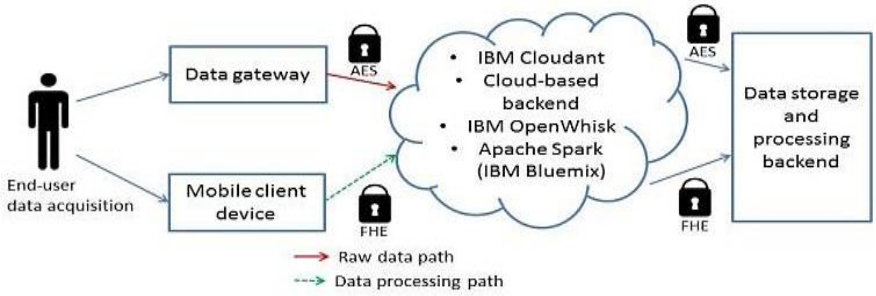

Figure 1: The e-Health System Architecture

The data transmission pipeline is closely related to the cloud-based infrastructure considering the last two phases: data storage, and safe data processing. The results of the data processing are sent back to the client devices on request. It is essential to note that the transmission of the results is performed while the data exists in a fully homomorphic encrypted format.

The homomorphic encryption functions usually multiply the amount of the processed data by a few orders of magnitude compared to the plain text original data. The architecture of the e-Health system defines a data pipeline, which is composed of two distinct data buses. Considering Fig. 1, the top data bus is intended for storage purposes, while the bottom data bus is intended for data transfers, which support the fully homomorphic encryption operations. Thus, the data processing path is used during the NTRU-based fully homomorphic encryption computations, while the raw data path is used during normal patient data retrieval. Therefore, this data transmission topology ensures that health status reports are efficiently obtained by fetching only the necessary personal health data through the storage data bus. This implies that only the relevant data chunk is processed and secured using the NTRU-based fully homomorphic encryption mechanism, while the results are safely transmitted to the requesting mobile client device.

4. Optimization Model

The existing FHE schemes are assessed considering simulated test settings. We have found that most of the considered FHE schemes are highly resource-intensive, even in the case of capable hardware, especially considering the related noise elimination (recrypt) operations, which are conducted after each multiplication operation [1,8]. We have optimized and extended this base model by implementing the computationally effective NTRU-based processing capabilities. This contribution allows for the personal health information to be processed in the encrypted form, while efficiently considering the limited processing capabilities of the wearable devices, which collect the health data, and also the finite computational power that the cloud-based processing infrastructure offers. Furthermore, it allows for cascaded homomorphic multiplications (Xh) to be performed, while avoiding the risk to encounter decryption errors. This type of processing problem would be catastrophic for a system like e-Health, as inaccurate personal health assumptions could be inferred, or even the mapping between the PHI and the respective individual could be rendered impossible.

The next paragraph describes the four types of FHE operations, which are considered by the e-Health system. The system relies on a parameter L (the Level), which must be computed before the initialization of any computation instruction. The level L is calibrated relative to the depth of the multiplication operations that are performed in the given computational context.

The first type of FHE operation that the e-Health system supports is the homomorphic addition (+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 the e-Health system supports is homomorphic multiplication (Xh). This operation considers 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 the rotate (<<<h, >>>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 the select (selmask), is important in order to correct the potentially altered slots (bits) of the data elements after the rotate operation. Therefore, the select operation has the role to maintain data consistency during the fully homomorphic encryption process.

The e-Health system considers the efficient usage of the data storage and processing backend, which has the role to safely process personal health information. The efficient incorporation of the NTRU-based improved fully homomorphic encryption primitives into the e-Health system relies on the utilization of the communication data path, which is presented in Fig. 2. Thus, each bit of the plaintext data is accurately embedded into the corresponding plaintext message. The ciphertext is generated considering a fully homomorphic encryption model considering the steps specified by the top data path. The direct processing of the encrypted data is the essential feature of this computational model. Thus, the bottom data processing path, which is represented in Fig. 2, implies that the input data is translated into a binary format, which is efficiently understood by the central processing unit. This is realized using the computation (fc(.)) and aggregation (fa(.)) functions that are illustrated as the first components of the bottom data processing path. Furthermore, the binary data is optimally processed using a parallel single instruction, multiple data (SIMD) model, and all the four types of operations that have already been described are implemented.

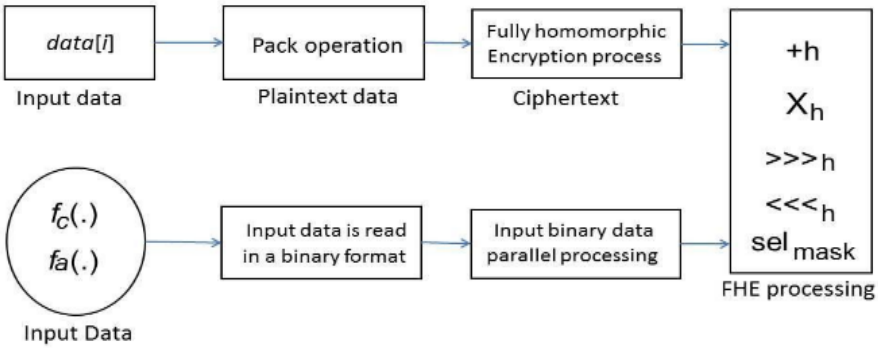

Figure 2: The distributed e-Health system basic data flow

The system's architecture is sufficiently flexible in order to address any use case scenario that implies the client data collection through a mobile device, and its safe transportation, storage, and processing at the backend. Therefore, it is suitable for the implementation of the smart city medical data system, which is presented in this paper. The system is configurable, and it can accommodate existing and future mobile devices, which gather the health data at the user's end. The speedup that is induced by the improved NTRU-based data processing module is demonstrated using the metrics that have been considered during the assessment of the first version. Thus, the system is used in order to collect the cardiac rhythm data. It is relevant to note that the data storage and processing backend is able to store the cardiac rhythm data and provide it on request to the entitled end-user through the raw data path that is presented in Fig. 1. Moreover, the system is capable to detect the delayed repolarization of the heart syndrome (DRHS) [9] and properly report this condition. The following section presents the most relevant implementation details, which are connected to this specific use case scenario.

5. Implementation Details

The computational behavior of the e-Health system, which implements the NTRU-based optimized fully homomorphic encryption scheme relies on three calibration variables. First, the value of the level L, which determines the operation of the FHE model, is an important performance factor. Second, the system performance relates to the number of multiplication and rotation operations, which are inherently expensive from a computational perspective. The multiplication is also relevant considering that it determines the calibration of the level L. The computational core of the e-Health system considers that the level L is kept at the optimal level, and it also ensures that the number of FHE operations is kept at the possible minimal value. Thus, it is important to assert that the system is designed in order to calculate the optimal value of the level L. This operation considers a number of NCT ciphertexts, which have the role to encrypt an array with n bits that store cardiac rhythm data. Furthermore, the improved real-world performance of the model is also related to the inclusion of the NTRU-based fully homomorphic encryption capabilities into the core of the system.

The proper assignment of the processed medical data to the respective person is validated through the usage of a hash-based digital signature mechanism. This is specifically developed for the e-Health system and is based on the usage of photon-based generators that produce the required seed, which is used to uniquely sign each patient's medical record. The in-depth presentation of the hash-based digital signature mechanism would constitute the subject of a separate paper. Therefore, this article does not present the mathematical details of this digital signature model. Nevertheless, it is important to describe the fundamental logic that determines the correct identification of the patients' data through the usage of the digital signature mechanism, which is novel, as it differs from all the similar existing approaches that this research reviews.

Thus, the size of the signature tree has to be H>=2, and by means of one public key, 2H documents have to be signed. We use a photon-counting generator to produce the required seed, as light waves' structures represent a valuable source of randomness. The hashing routines receive the seed as the input, and consequently, the hash-based signature and verification keys Xi, Yi, where 0<=i<=2H, are generated. Here, Xi is the signature key, and Yi is the verification key. In order to calculate the leaves of the tree, the signature keys have to be hashed by means of the hash function h:{0,1}*➔{0,1}n. The calculation of the parent node supposes that the concatenation of the two previous nodes must be hashed. The root of the tree is the public key of the public signature. Furthermore, in order to sign a message of any size, it is transformed to size n by means of the hashing operation. Thus, h(m) = mhash, and in order to sign the message, an arbitrary one-time-key X any is used. This key is calculated using the same seed that is received from the photon counting generator. The signature is a combination of the one-time signature, the one-time verification key, the index of a key, and all sibling nodes according to the selected arbitrary key, which is featured by the index "any". Thus, the following relation stands: msignature= (msig||any|| Yany||authn0, authn1,…,authnH-1). In order to verify the signature, the one-time signature is checked using the chosen verification key. If the validation has passed successfully, then all the required nodes must be computed. If the root of the tree coincides with the public key, then the signature is correct. Consequently, the patient data is properly assigned to its bearer before the results of the cloud-based data processing are sent back to the respective user's mobile device.

It is significant to note that the speed assessments that were conducted prove that the digital signature, which is generated for one patient that is administered by the e-Health system determines the following execution times. The average public key generation time is 0.018247955289241242 seconds, the average signature generation time is 0.0007876256585058229 seconds, and the average signature verification time is 0.011725638324516478 seconds. The length of the respective messages is 128 bits. This ensures completely secure identification of the involved patient data before the cloud-based software components send the data back to the user's mobile device while inducing virtually negligible in-creases of the overall execution times. This is essential for the e-Health system, as the homomorphic encryption-related processes are sufficiently time-consuming by their nature.

5.1. The Detection of the Average Heart Rate

The calculation of the average heart rate considers the storage of the encrypted values in NCT ciphertexts. The optimized implementation of the e-Health system considers three main types of improvements. The first one considers the reduction of the computationally expensive multiplication operations. The second one relies on the reduction of the computation operations depth so that the level L is calibrated at the optimal level. Finally, the design and implementation of the NTRU-based encryption routines further improve the efficient usage of the computational resources, including the processor time allocation.

The addition operation is optimized considering two mechanisms. These two mechanisms are called additive compression and prefixed parallel addition. The additive compression transforms three data inputs (H,M,F), each of them composed of n bits, into two outputs. These are represented by the AR (addition result), and LOVER (leftover). Thus, the $A_R = H\Delta M\Delta F$, and $L_{OVER} = [(H \times M)\nabla(H \times F)\nabla(M \times F)] \ll 1$. Here, Δ denotes an additive single instruction multiple data (SIMD) operation, while the nabla operand (∇) also represents a SIMD operation, which is conducted on all n bits of the input data in a parallel manner. The prefixed parallel addition has been designed and implemented considering the algorithmic model presented in [10].

The computation of the average heart rate considers the NCT ciphertexts, which encrypt the input messages that are represented on n bits. Thus, the first step of this data flow relies on the usage of additive compression in order to transform NCT ciphertexts into two ciphertexts. Consequently, the resulting two ciphertexts are added using the prefixed parallel addition operation.

5.2. The Detection of the Delayed Repolarization of the Heart

The detection of this abnormal cardiac condition is based on the usage of the scientific model that is presented in [11]. The basic equation that is presented in [11] is optimized. Thus,

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

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

Here, the expressions ${T_{QT}}^2 = T_{QTH}$ and $T_{RR} \times 225{,}625 = T_{RRH}$, 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. In principle, T_{QT} represents the time taken for ventricular depolarisation and repolarisation, 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. The optimized version of equation (1), which has been obtained after an extensive set of calibration tests, has been considered during the implementation of the e-Health system. It is relevant to note that the equation is optimized regarding the accuracy of the detection and the efficient usage of the computational resources. The equation ensures that the e-Health system accurately detects the DRHS condition with virtually no false positives while running only the absolute necessary NTRU-based FHE operations. The data storage and processing backend aggregate the results of the individual comparisons. The client device requests a report from the backend considering a certain period. 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 during the given period the comparison $T_{QTH} > T_{RRH}$ was true at least once. Consequently, the DRHS condition occurred with a significant probability at least once.

5.3. The Detection of the Minimum and Maximum Heart Rates

This is a functional requirement of the system and is implemented considering the (.) function, which is graphically described in Fig. 2. The function is intended to convert the input data into a binary format, which is efficiently processed by the system. Let us recall that the comparison of two numbers, which are defined by n bits, produces a result that is also determined 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 that have a value of 0. Furthermore, if the first number is less than the other number, then the result contains only bits with a value of 0. Furthermore, the e-Health system triggers a succession of rotate and selects 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. As a consequence, the proper calculation 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 NTRU-based fully homomorphic encryption computation is calculated according to the following reference formula: $L > (\log_2 n + 2) \times \log_2 N_{CT}$.

6. Practical System Performance

Let us consider the e-Health system's distributed architecture, which is graphically presented in Figure 1. It is important to assert that one of the system's important features is its ability to accommodate any kind of user-side (client-side) 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 user-side data collection devices may vary. Thus, any technically suitable user-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 deployed on the user's mobile device, is able to send 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.

The practical performance assessment procedure considers a sample of 500 citizens of Brasov City. We have tested a variety of personal cardiac rate sensors and determined that the most accurate device is the Polar H7. Consequently, it has been decided to use this device in order to gather the cardiac rate data, which is necessary to assess the system's ability to detect the delayed repolarization of the heart syndrome (DRHS).

Thus, the system architecture is determined by the following software and hardware components. The cardiac rate data is gathered by the Polar H7 personal sensor. The collected data is sent to each person's Android smartphone. The e-Health system's client application is installed on the 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 [8]. This version supports the optimizations that have been described in the previous section, including the NTRU-based processing routines that further enhance the practical system speed. The optimized version of the backend is deployed to the Bluemix infrastructure using a proper build pack. 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 e-Health system generates. The arrival of new health data in the cloud is detected by the IBM OpenWhisk Bluemix programming service. Following, the proper event handlers are triggered, so that the newly arrived data is properly stored by the IBM Cloudant. Additionally, any data request that comes from the client devices is properly

processed by the backend considering the algorithms and data flows that are presented in the previous sections.

6.1. The 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 has to 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 (SR). This calculates the amount of storage that is necessary to store one byte of plaintext data in a fully homomorphic encrypted format. As an example, if $S_R = 500$, then it is clear that for one byte of plaintext data, there are necessary 500 bytes in order to store the fully homomorphic encrypted byte.

The third performance metric is determined by the processing speed (PS). This metric is defined through the following ratio: $PS = PTO/PIN$. 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.

Let us recall that the system's practical performance assessment has been conducted considering the dataset of 500 participants.

The values of the performance metrics recorded during the detection of the heart rates are presented in Table 1.

Table 1. The performance metrics values

Data Reading Interval	N_{CT}	*Level L*	$XFER_{IN}$ (GB)	$XFER_{OUT}$ (GB)	S_R	P_S
One minute	2	12	4.8	2886.3	32.1	0.54
Five minutes	12	15	5.9	1147.8	39.4	0.24
Fifteen minutes	40	18	6.4	608.2	47.5	0.23
Thirty Minutes	44	20	9.7	1003.5	88.3	0.36
One hour	86	21	7.4	592.8	91.4	0.35
Three hours	258	24	8.9	201.6	101.2	0.37
Six hours	519	25	10.1	98.9	108.5	0.36
Twelve hours	1021	26	11.2	42.7	117.4	0.39
One day	2099	28	14.3	24.6	128.1	0.42

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 period that is considered when reading the client-side input data, the number of the ciphertexts NCT, 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 proposed e-Health system functions more efficiently than existing similar approaches, such as the one that is presented in [11]. It is relevant to note that the presented e-Health system is unique in the context of other similar systems considering its distributed architecture, and also the ability to assure complete safety for the collected data, both during the data transmission and processing stages. Furthermore, the optimization of the computing core through the implementation of the NTRU-based routines further enhances the system's runtime efficiency, as it is demonstrated by the obtained values of the processing speed. The similarity with

other systems only pertains to the fully homomorphic primitives, as the e-Health system is one of the few that offers this platform for personal health information collected 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 while keeping the usage of the computational resources at the minimum possible level using the NTRU-based fully homomorphic encryption routines.

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 XFERIN and XFEROUT 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 (XFEROUT) data stream, while the downloaded data (XFERIN) is usually not monitored regarding the amount of the transferred data. Furthermore, the number of the ciphertexts (NCT) is maintained at the minimum possible level, while the value of the level L is also computed in an optimal fashion.

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 minute	2	5	1.1	1102.3	10.1	0.06
Five minutes	8	6	1.9	314.8	12.7	0.07
Fifteen minutes	22	8	2.4	108.5	14.5	0.10
Thirty Minutes	41	10	2.8	83.5	16.3	0.11
One hour	85	11	3.1	69.8	29.4	0.32
Three hours	256	12	3.6	61.8	37.3	0.28
Six hours	517	14	4.8	30.2	42.5	0.29
Twelve hours	1023	15	6.2	17.7	49.4	0.33
One day	2079	17	7.3	9.6	53.6	0.35

Additionally, it is relevant to note that the finer the 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 a response to the client software module's requests [13–15].

6.2. Comparative Performance Analysis

The flexible system architecture allowed us to easily co duct a comparative system performance analysis by modifying the homomorphic encryption core's implementation. This homomorphic model has been chosen because, according to our performance assessment, it is one of the most efficient fully homomorphic encryption schemes that is able to accommodate the requirements of the e-Health system in terms of data processing speed and used resources.

The same set of performance metrics have been used in order to ensure a uniform comparative performance analysis. The homomorphic model that is presented in [12] has the advantage that it doesn't necessarily require the computation of the level variable L, in order to prevent the alteration of the computed encrypted data. Furthermore, the two approaches demonstrate similar behaviors concerning the storage ratio, while the NTRU-based homomorphic encryption core, which is described in this paper, produces superior values for the processing speed and both data transfer performance metrics. Consequently, it can be asserted that while the homomorphic model suggested in [12] implies a slightly simpler code implementation, the NTRU-based approach performs better, considering the most significant performance metrics, in the context of the e-Health system [16, 17].

7. Conclusions and Future Developments

The efficient collection of personal health data has been increasingly important during the past fifteen years. As a consequence of the technological advancements, it has relatively recently become possible to collect 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 problem becomes even more important in the context of the e-Health system, which is designed to collect personal medical data from a wide urban area. Furthermore, 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 personal health information.

This paper describes the e-Health system, considering the improved NTRU-based fully homomorphic encryption core, which addresses the entire palette of requirements that have been mentioned. Considering its flexible and decoupled architecture, the system is capable to accommodate 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 results of this assessment demonstrate that the system is capable to sustain a perfectly functional and secure data flow between the client data collection devices and the data storage and processing backend, in both directions, for the entire sample of 500 citizens from Brasov. This result is worth to be mentioned because this is one of the few integrated systems, which offers the full range of personal health information collection, storage, and processing functional capabilities. Furthermore, it is significant to mention that this contribution proves that fully homomorphic encryption can be used in order to secure a complex system, while the NTRU-based fully homomorphic encryption model proves that it is useful in order to further enhance the system's efficient behavior. In this context, the complexity particularly 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. The performance tests prove that the system is scalable, which would potentially justify its deployment on full large urban areas.

The field of homomorphic encryption still needs to evolve in order to support general-purpose large-scale big data systems that value data privacy. Thus, it is intended to continuously improve the homomorphic encryption core of the e-Health system. Additionally, it was determined that the NTRU-based core behaves better than another promising homomorphic encryption model.

The software system's architecture allows for the e-Health platform to be easily extended with new functional capabilities. The implementation of the NTRU-based fully homomorphic encryption routines determines an efficient computational behavior of the system, including the computational time that is used. Therefore, in the short term, we'll concentrate on extending the system with new functional requirements. Thus, we'll 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 by 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 an active pet. The system has the potential to address current and future similar computational use case scenarios. Therefore, it will be maintained and continuously improved.

8. Acknowledgments

The work was financed by Shota Rustaveli National Science Foundation and was conducted in the frame of the [CARYS-19-121] grant.

9. References

[1] C. Gentry, A Fully Homomorphic Encryption Scheme, PhD Thesis, Stanford University, Stanford, CA, USA, 2009.

[2] I. Kuzminykh, et al., Investigation of the IoT device lifetime with secure data transmission, Internet of Things, Smart Spaces, and Next Generation Networks and Systems 11660 (2019) 16–27. doi: 10.1007/978-3-030-30859-9_2.

[3] Q. Li, G. Cao, T. F. L. Porta, Efficient and Privacy-Aware Data Aggregation in Mobile Sensing, IEEE Transactions on Dependable and Secure Computing 11 (2014) 115–129. doi:10.1109/TDSC.2013.31

[4] R. Zhang, J. Shi, Y. Zhang, C. Zhang, Verifiable privacy-preserving aggregation in people-centric urban sensing systems, IEEE Journal on Selected Areas in Communications 31 (2013) 268–278. doi:10.1109/JSAC.2013.SUP.0513024.

[5] D. Fiore, R. Gennaro, Publicly Verifiable Delegation of Large Polynomials and Matrix Computations, with Applications, in: Proceedings of the 2012 ACM Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY, USA, 2012, pp. 501–512. doi:10.1145/2382196.2382250.

[6] C. Papamanthou, R. Tamassia, N. Triandopoulos, Optimal Verification of Operations on Dynamic Sets, in: Advances in Cryptology—CRYPTO 2011, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 91–110.

[7] G. Zhuo, Q. Jia, L. Guo, M. Li and Y. Fang, Privacy-preserving verifiable proximity test for location-based services, in: Proceedings of GLOBECOM Conference, 2015, pp. 1–6.

[8] C. Orencik, E. Savas, An efficient privacy-preserving multi-keyword search over encrypted cloud data with ranking, Journal of Parallel and Distributed Databases 32 (2014), 119–160.

[9] N. Cao, C. Wang, M. Li, K. Ren, W. Lou, Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data, IEEE Transactions on Parallel and Distributed Systems. 25 (2014) 222–233. doi:10.1109/TPDS.2013.45

[10] D. Fiore, R. Gennaro, V. Pastro, Efficiently Verifiable Computation on Encrypted Data, in: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY, USA, 2014, pp. 844–855. doi:10.1145/2660267.2660366.

[11] T. Jaeger, J. Schiffman, Outlook: Cloudy with a chance of security challenges and improvements, IEEE Security and Privacy 8 (2010) 77–80. doi:10.1109/MSP.2010.45

[12] M. Kuzu, M.S. Islam, M. Kantarcioglu, Efficient similarity search over encrypted data, in: In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, IEEE Computer Society, 2012, pp. 1156–1167.

[13] M. Zalisky, R. Odarchenko, S. Gnatyuk, Y. Petrova, A. Chaplits, Method of Traffic Monitoring for DDoS Attacks Detection in e-Health systems and networks, 2020. CEUR-WS.org, online CEUR-WS.ORG/Vol-2255/paper18.pdf

[14] M. Iavich, A. Gagnidze, G. Iashvili, S. Gnatyuk, V. Vialkova, Lattice based Merkle, 2019. CEUR-WS.org, online CEUR-WS.ORG/Vol-2470/p6.pdf

[15] G. V. Hindumathi, D. L. Bhaskari, Message Based Key Distribution Technique for Establishing a Secure Communication Channel in IoT Networks, International Journal of Computer Network and Information Security(IJCNIS), International Journal of Computer Network and Information Security(IJCNIS) 11 (2019) 28–35. doi:10.5815/ijcnis.2019.11.04

[16] S. Gnatyuk, B. Akhmetov, V. Kinzeryavyy et al, New Secure Block Cipher for Critical Applications: Design, Implementation, Speed and Security Analysis, Advances in Intelligent Systems and Computing 1126 (2020), pp. 93–104.

[17] S. Kamel, S. Elhamayed, Mitigating the Impact of IoT Routing Attacks on Power Consumption in IoT Healthcare Environment using Convolutional Neural Network, International Journal of Computer Network and Information Security 12(4) (2020) 11–29. doi:10.5815/ijcnis.2020.04.02.