Information Fusion 62 (2020) 32–46

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

DependData: Data collection dependability through three-layer decision-making in BSNs for healthcare monitoring

Tao Hai[a,1], Md Zakirul Alam Bhuiyan[b,1], Jing Wang[a,d], Tian Wang[c], D. Frank Hsu[b], Yafeng Li[a,*], Sinan Q Salih[f,g], Jie Wu[e], Penghui Liu[d]

[a] School of Computer Science, Baoji University of Arts and Sciences, China 721007
[b] Department of Computer and Information Sciences, Fordham University, 10458 USA
[c] College of Computer Science and Technology, Huaqiao University, China, 361021
[d] Faculty of Computer Systems & Software Engineering, University Malaysia Pahang, 26300 Malaysia
[e] The Department of Computer and Information Sciences, Temple University, 10121 USA
[f] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
[g] Computer Science Department, College of Computer Science and Information Technology, University of Anbar, Ramadi 31001, Iraq

ARTICLE INFO

Keywords:
Body sensor networks
Healthcare monitoring
Decision-making
Dependability
Data dependability
Security and privacy
Data quality

ABSTRACT

Recently, there have been extensive studies on applying security and privacy protocols in Body Sensor Networks (BSNs) for patient healthcare monitoring (BSN-Health). Though these protocols provide adequate security to data packets, the collected data may still be compromised at the time of acquisition and before aggregation/storage in the severely resource-constrained BSNs. This leads to data collection frameworks being meaningless or undependable, i.e., an undependable BSN-Health. We study data dependability concerns in the BSN-Health and propose a data dependability verification framework named DependData with the objective of verifying data dependability through the decision-making in three layers. The 1st decision-making (1-DM) layer verifies signal-level data at each health sensor of the BSN locally to guarantee that collected signals ready for processing and transmission are dependable so that undependable processing and transmission in the BSN can be avoided. The 2nd decision-making (2-DM) layer verifies data before aggregation at each local aggregator (like clusterhead) of the BSN to guarantee that data received for aggregation is dependable so that undependable data aggregation can be avoided. The 3rd decision-making (3-DM) layer verifies the stored data before the data appears to a remote healthcare data user to guarantee that data available to the owner end (such as smartphone) is dependable so that undependable information viewing can be avoided. Finally, we evaluate the performance of DependData through simulations regarding 1-DM, 2-DM, and 3-DM and show that up to 92% of data dependability concerns can be detected in the three layers. To the best of our knowledge, DependData would be the first framework to address data dependability aside from current substantial studies of security and privacy protocols. We believe the three layers decision-making framework would attract a wide range of applications in the future.

1. Introduction

Recently, we have witnessed the emergence of Body Sensor Networks (BSNs) for patient healthcare monitoring applications where the body sensors are coupled with the Internet of Things (IoT) and cloud platforms. Body sensors in BSNs acquire patients health data (such as temperature, heart beat, ECG, Oximetry, or fetal status), may process a certain part locally, and transmits the data to a designated intermediate node (such as aggregator, gateway, or Fog) [1,2]. The data then travels from BSNs to a remote computing platform (such as medical cloud) and arrives at a remote healthcare data user (also referred to as the patient/provider/doctor/receiver), who will access the data using the IoT-enabled interfaces (such as mobile). The BSN-enabled health monitoring applications have already demonstrated great potential for significantly improving the quality of patient health, healthcare facilities, and well-being [3–8].

In spite of the pleasing potentials of the BSN for patient health monitoring, the incorporation of body sensors and IoT into the Internet brings security and privacy concerns for the reason that other IoT devices and network infrastructures around a BSN may interfere with body sensors in the BSN. They may make happen various privacy and security threats to body sensors data. We have witnessed some recent complicated security

[*] Corresponding author.
[1] Co-first authors.
E-mail addresses: zakirulalam@gmail.com (M.Z.A. Bhuiyan), wangtian@hqu.edu.cn (T. Wang), liyafeng770729@126.com (Y. Li).

https://doi.org/10.1016/j.inffus.2020.03.004
Received 24 April 2019; Received in revised form 16 February 2020; Accepted 19 March 2020
Available online 2 April 2020

attacks even after using cryptography techniques, particularly during data transformation into the cipher and after the transmission of the cipher. More security attacks including DDoS, insider, and physically compromised attacks on-body sensors are normal types of attacks in BSNs for healthcare [7,9,33]. In addition to that, there would be kinds of security threats/attacks that are closely connected to the transmission, such as snooping, masquerade, data integrity, compromised signal penetration, data breach, collusion, and so on. These kinds of threats/attacks lead to some concerns; for example, data can be compromised before or after the data transmissions, by which patients health signals can be altered. A number of latest investigations have established that introducing security/privacy attacks into the BSN and IoT for healthcare might lead to disastrous circumstances and life-threatening situations [1,10]. In addition, latest technologies such as remote IoT storage and cloud increase the body sensor data security concerns to another level.

To deal with the situations above, numerous compelling security protocols and algorithms have been used to provide health data protection during computing functions (processing, storing, and transmitting and decision-making) in a BSN for healthcare applications. These protocols and algorithms include cryptography and authentication algorithms, private and public-key generation, data anonymization, tokenization, MAC algorithms, and the like [9,11,12]. Similar protocols provide a series of security capabilities that safeguard the communication while maintaining the functionality, convenience and flexibility in the cloud. These protocols/algorithms also have diverse constraints such as computation cost, computation complexity, energy consumption, and real timeliness. However, many of the protocols/algorithms are often suggested for sensor networks with higher resources and capabilities than those of the BSN. In many cases, BSNs cannot secure the data properly or make the data vulnerable to malicious attacks due to severe resource constraints (high-rate data acquisition, energy, processing, communication). Particularly, a tiny body sensor with a micro-battery does not provide enough energy to run a complicated encryption algorithm with a bigger-sized secrete key, while data transmission with a light-weight secret key is often vulnerable.

Furthermore, there exist some security concerns in BSNs that directly interfere with health data in BSNs. For example, ECG sensors in a BSN are shown to be susceptible to data manipulation attacks on the measurements. Such measurements misrepresent the current health state of a patient. An attacker may alter the measured ECG signal by manipulating the sensor, thus introducing a wrong view of the patients health condition [12]. Attackers are around in the relative proximity (say IoT network) to the healthcare provider/user such that the security threats/attacks can be launched directly on body sensor signal outputs, such as ECG, heart beat, and tri-axial accelerometer. Through hardware or physical attacks, an attacker may also physically compromise one or more body sensors or one of the sensing units of a sensor, and/or afterward alter their firmware. An alternative attack is to replace a trustworthy sensor of the BSN with a comprised one, setting the same id or similar functional features. An adversary may inject low-quality, unrelated, or compromised signals through variety types of stimuli. These include electromagnetic induction, light, and acoustic wave [9,12].

Regarding the critical situation above, one of the foremost hurdles in adopting BSNs together with IoT and cloud for healthcare application is “data dependability.” Data dependability can be defined as how exact and trustworthy the data is, which reflects whether the data has been acquired with any security and privacy compromises and violations. A motivating example of data dependability is illustrated in Fig. 1. In a BSN, we verify the patient ECG data dependability in three layers of the BSN-Health. We find that with different rates of data dependability confirmation in the three layers, we can guarantee to achieve the data quality of up to 93% under security attacks. That is to say, the data collected under security and privacy attacks often have dependability concerns. This is a critical concern for every sensor in a BSN as the dependability of their collected data is highly important with respect to the risks of patients lives. Patients everyday health status (temperature, ECG, pressure, accelerometer) depends on the dependable signal collection. Existing security protocols mainly provide protection to computing functions for patients health data, but they do not address the concerns verifying if the collected data used in these functions is undependable. Also, they often do not validate whether or not there is an amount of undependable data at the time of data acquisition, before aggregation, before storing at a remote storage, or before utilizing/viewing the data by a healthcare data user. Current multi-sensor fusion techniques for BSNs also do not address the concerns [3,4,6].

To address the concerns above, in this paper we study data dependability concerns in the BSN-Health and propose a data dependability verification framework named DependData with the objective of verifying data dependability in the three-layer decision-making framework, as shown in Fig. 2. We present the 1st decision-making (1-DM) layer to verify signal-level data at each health sensor of a BSN locally to guarantee that collected signals ready for processing and transmission are dependable. That is, undependable processing and transmission in the BSN can be avoided. We then present the 2nd decision-making (2-DM) layer to verify data before aggregating at each local aggregator of the BSN to guarantee that data received for aggregation is dependable. That is, undependable data aggregation in the BSN can be avoided. We think

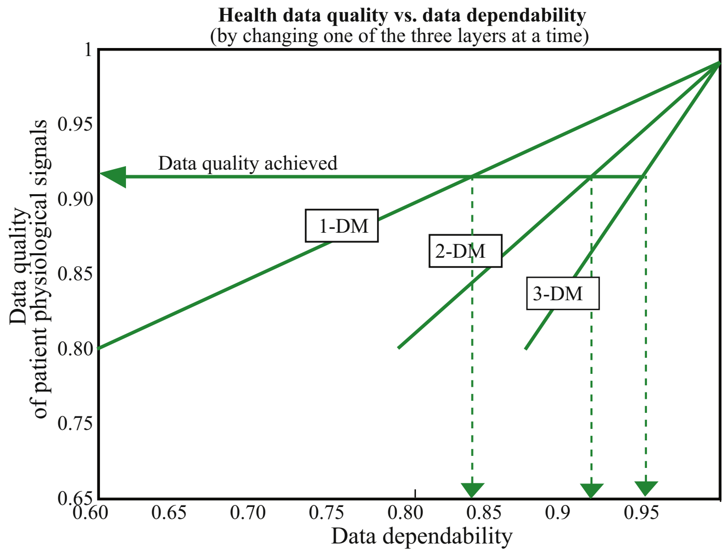


Fig. 1. The rate of data quality achieved through guaranteeing the data dependability in three-layer decision-making under security attacks on the ECG signal measurements.

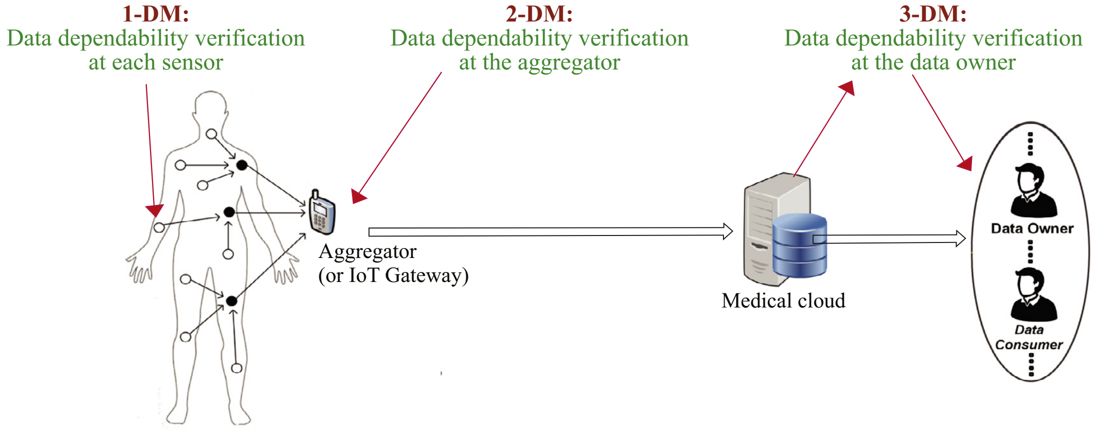
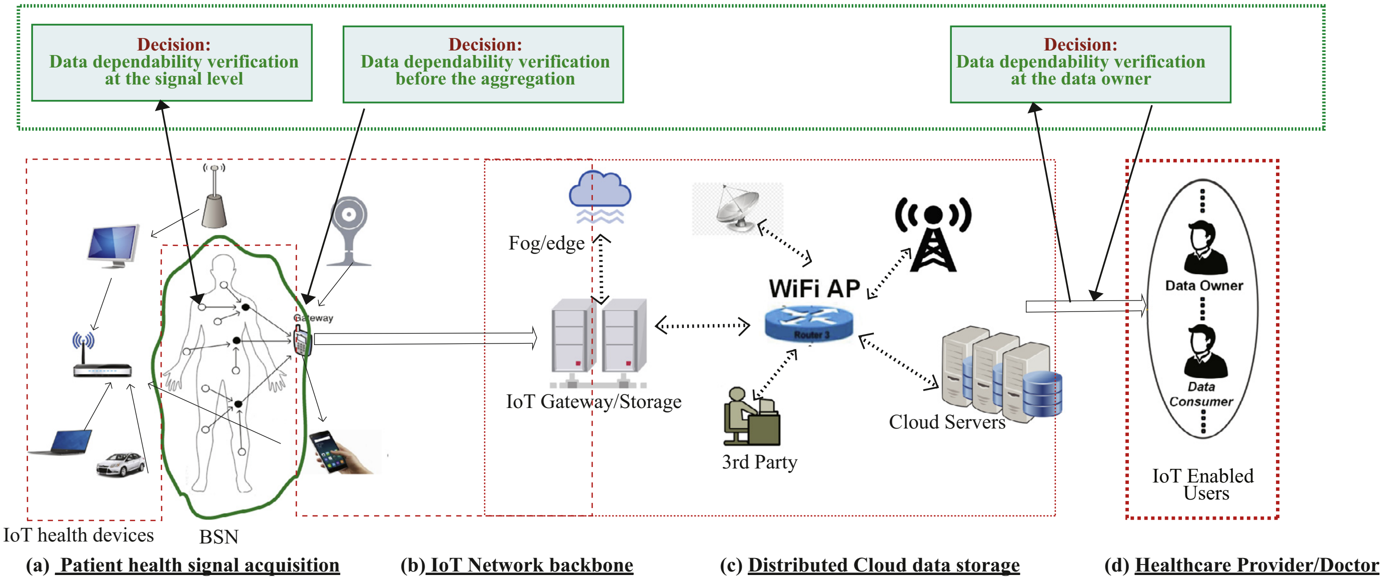

1-DM:
Data dependability verification at each sensor
2-DM:
Data dependability verification at the aggregator
3-DM:
Data dependability verification at the data owner
Aggregator (or IoT Gateway)
Medical cloud
Data Owner
Data Consumer

Fig. 2. Proposed three-layer decision-making framework in DependData framework for data verification in a BSN.



(a) Patient health signal acquisition (b) IoT Network backbone (c) Distributed Cloud data storage (d) Healthcare Provider/Doctor

Fig. 3. A healthcare monitoring infrastructure incorporating a BSN surrounded by an IoT network and integrated with the cloud: (a) sensors in the BSN layer for patient body monitoring and signal level data dependability verification; at this layer (b) aggregators of the BSN integrated with the IoT network for data collection, data dependability verification, and the aggregation; (c) permanent data storage on the cloud; (d) IoT-enabled devices where the healthcare data user can view the patient health information, where data dependability verification is required for the patient data quality at the user.

that acquired data is vulnerable to being altered before/after transmission. For the need of the 2-DM, we think that acquired signals are vulnerable to alteration before/after transmission and before reaching an aggregator. We enable the aggregators in the BSN to verify body sensor signal features before aggregating data in a neighborhood.

The 1-DM assures that the acquired data has no dependability concerns, which is verified at sensor level, while the 2-DM assures that the transmitted data is not altered before/after transmission, which is verified at the aggregator. To guarantee the data dependability, a measurement model adopted from Mutual (Signal) Information Independence (MSII) is exploited. The idea is to find independence within any two signals either from two different sets of signals of the same sensor or different sensors, or from history. To do that, we consider the ground truth and signal correlation analysis between the two signals. The MSII is utilized as an index to verify the data dependability of any acquired or received data. Both 1-DM and 2-DM can help to drop a significant amount of undependable data before transmission, resulting in a reduction of the resource consumption of the tiny body sensors of the BSN.

We think that all the aggregators in the BSN transmit the collected data, which pass through the 2-DM with no dependability concerns, towards a remote data storage (medical cloud). A healthcare data user can pool the data from the cloud for health status analyses using an IoT-enabled interface and this data might have been compromised before/after being stored into the cloud and before the user can use it. We apply the 3rd decision-making (3-DM) layer to verify the data in order to guarantee that the available data is dependable so that undependable data can be avoided. Data dependability is calculated by the data dependence, data quality, etc. To the best of our knowledge, DependData would be the first framework addressing data dependability in three layers aside from the current extensive studies of security and privacy protocols. The three-layer decision-making framework would attract a lot of applications in the future.

In summary, the contributions of this paper are four-fold.

- We study data dependability concerns in the BSN-Health and proposed DependData framework for data dependability verification at different layers of a BSN.
- We propose two decision-making layers (1-DM and 2-DM) for the BSN to make a decision on the data dependability at each sensor signal level locally and at each aggregator in the distributed manner in the BSN.
- We present the idea of data dependability at the healthcare data user through the 3rd decision-making (3-DM) layer, which can help the user to know the quality of the data they are going to view.

- Finally, we evaluate the performance of DependData through simulations regarding 1-DM, 2-DM, and 3-DM and show that up to 92% of data dependability concerns can be identified at the three-layers, according to the dependability computational theory when computing dependability of each of the decision-making three layers in series.

The remainder of the paper is summarized as follows. We cover the background and closely related work in Section 2. Section 3 offers the proposed design of DependData. In Section 4, Section 5, and Section 6, we provide 1-DM, 2-DM, and 3-DM layers, respectively. Section 7 presents the evaluation of DependData. Finally, Section 8 offers the conclusion of this paper and future directions.

2. Background and related work

In this section, we first explain the data dependability in decision-making in the BSN. Then, we relate our work to existing work regarding fusion techniques, data security, privacy protocols, and algorithms in the BSN-Health.

2.1. Data dependability in BSNs for right decision-Making

The basis of an adequate decision-making framework is the dependability of the data gathered to make decisions. When the data cannot be trusted, a proper decision cannot be made. Dependability of the data can simply be as exact as the sensors in a normal state in BSNs or the health exam used to acquire the data. Irrespective to the specialty of the healthcare users and providers (doctors/nurses), they are also able to make decisions on the health diagnosis and treatment of a patient. This is proved by the outcomes obtained from a few tests to which the patients body is subjected [13]. To guarantee that the patient gets the right diagnosis and the right treatment at the right time, the doctors/nurses have to first make sure that the exam outcomes used to get decisions are dependable and correct. How can a user make a decision if the outcome being used as the basis of that process cannot be trusted [13,14,34]?

Now, more than ever, there is a vital demand for sharing the data among various sensors of the BSN and IoT devices with other networks and healthcare systems so that specialists and healthcare decision-makers may evaluate the data and make the right decisions. Meantime, in theBSN for health monitoring systems, a great quantity of data bear vital information for making a crucial decision is collected from sensors in BSNs. As a result, it is significant that dependability concerns related to data dependability, quality, provenance, privacy, must be studied for BSN-Health data sharing, health situation estimation, and multi-sensor data integration to support decision makers and analysts.

Without data dependability verification, once something wrong occurs to the gathered data in a BSN, patient health exam outcomes can be influenced. Whats more, if some attackers add incorrect or irrelevant information to data and it not identified by the healthcare data user, there can be a huge influence on the actual information of patient health state and the actual health state. A technique to address the concerns can be to appraise the collected data dependability and know the degree of the data quality and data dependence.

2.2. Existing fusion schemes in BSNs for healthcare monitoring

There is a decent amount of promising work on multi-sensor fusion for the BSN. A comprehensive overview of fusion schemes in BSNs can be found in [2,3]. A decision-making scheme using trust relationship is proposed for Health loT executed by followers of an environmental health society of interest that include risk grouping, trust reliability, and loss of information of health likelihood as the three dimensions in the development of the decision-making. One typical multi-sensor data fusion scheme can be to calculate the risk through the function of input memberships regarding the amount of significant signs in BSNs [4]. The degree of risk stands for the calculation of risk levels, which are within 0 and 1. The greater the risk degree, the more severe the patients health state is and the more action it needs in health awareness [4]. Another work proposed a fusion framework for activity evaluation on the basis of heart beat and accelerometer data of rescuers in the case of urgent interventions [15]

2.3. Existing data security and privacy protocols and algorithms in the BSN-Health

Guaranteeing protection to the patient body data over the transmission channels from the BSN to a remote storage (cloud) is addressed. This supports end-to-end (E2E) security, which is permitted by allocating encryption keys among the BSN and the cloud. As a result, the body data can be encrypted, and it has a secure integrity operation. Moreover, the secret key can also be used for a joint authentication to the transmission. However, successful dissemination of secret keys introduces significant overhead overheads in assisting the transmission security [9,10,16]. A large number of researchers focus on protecting communication links disjointly. There are also healthcare solutions preserving public and private in IoT devices, device-to-device, device-to-phone, phone-to-medical cloud [17]. It presents a major operating cost in the resource-constrained BSN. Numerous works present extra hardware in the BSN-Health system merely to obtain security between different things. But this approach is limited to securing sensor-sensor transmission and may not support E2E security. PEES lessens this limitation to some extent by offering transparent E2E transmission security between a body sensor in the BSN and the cloud [11].

Similarly, a good number of convincing security protocols has been used to offer data security during computing functions (processing, storing, and transmitting and decision-making) in BSNs for healthcare applications, including encryption and authentication algorithms, private and public-key cryptosystems, anonymity, tokenization, MAC algorithms, data usability, data auditing, and the like [18]. Similar protocols provide a series of security capabilities that safeguard the communication while maintaining the functionality, convenience and flexibility in the data cloud. Security control protocols are also utilized for controlling the data access. These protocols/algorithms also have diverse constraints, such as computation cost, computation complexity, energy consumption, real timeliness. However, many of the protocols/algorithms are often suggested for sensor networks with higher resources and capabilities than those of the BSN. Body sensors in the BSN are tiny. In many cases, BSN cannot secure the data strongly due to severe resource constraints (power, processing, communication). As a result, such tiny body sensor does not have enough energy to run a sophisticated encryption algorithm with a bigger-sized secrete key. To avoid this, when we use a light-weight secret key, data transmission often becomes often vulnerable.

Tremendous work has also dealt with trustworthiness in various functions, particularly in secure communication. A trustworthiness management and trust management protocol for data collection is popular. Trustworthiness proposed for big data collection has two functions for user familiarity and user similarity in order to detect malicious users [19]. To achieve data that is authentic, the idea of trustworthiness is suggested and numerous associated methods are used to compute the trustworthiness [20].

Furthermore, there would be some security concerns in BSNs that directly interfere with health data. For example, ECG sensors in a BSN are shown to be susceptible to data manipulation attacks on the measurements. Such measurements misrepresent the current health state of a patient. An attacker alters the ECG signal intensity measured by manipulating the sensor, thus introducing a wrong view of the patients health condition [12]. Malicious users can be around the relative proximity (say IoT network) of the healthcare data user (also referred to as the patient/provider/doctor) such that the attacks can be launched directly on body sensor signal outputs, such as ECG, heart beat, and tri-axial accelerometer. Through hardware or physical attacks, an attacker may also physically compromise one or more body sensors or one of the sensing units of a sensor, and/or afterward alters their firmware. An alternative attack is to replace a trustworthy sensor of the BSN with a comprised one setting the same id or similar functional features. An adversary may inject low-quality, unrelated, or compromised signals through a variety type of stimuli. This includes electromagnetic induction, light, and acoustic wave.

Our work fully differs from the previous work above regarding the way we designed the proposed fusion frameworks for verifying the data dependability in BSNs. A thorough search of the relevant literature yields that this work might be the first work to use fusion frameworks in different layers of the BSN-enabled healthcare application and address the data dependability. This work also stands aside from current studies of security and privacy protocols for BSNs. Each of the layers is used to guarantee the data dependability, which finally reduces a large amount of undependable (so unnecessary) data transmission and processing in BSNs.

3. The design of dependdata

We first describe the data dependability concerns in the BSN-Health applications regarding numerous threat/attack surfaces in this section. Then, we present the DependData framework for detection data with dependability concerns data during data collection.

3.1. Network architecture

Similar to the traditional BSN-Health, we consider a BSN integrated with IoT and cloud for a healthcare application having several layers of data collection, processing, transmission, and decision-making layer. We illustrate a representative BSN-enabled network architecture, as shown in Fig. 4. This three-layer network architecture is made similar to the traditional BSN-based data collection architecture: data acquisition layer (IoT devices level), IoT network layer (intermediate data processing level at cluster heads or aggregators), and high-end layer (data processing and storage layer, such as cloud).

3.1.1. BSN-Health Model and data

The BSN-Health system consists of a number of health status measurement, computing and communicating sensors in a BSN, as shown in Fig. 4(a). This includes medical sensors together with IoT networks. In DependData, in the case of patient health monitoring, sensors continuously sense the patients bodies to measure physiological states of the patient, including the status of the blood pressure, body activities, heart beat, ECG, and more, which are utilized by doctor or healthcare experts. If a sensor in the BSN needs to transmit any data, it forwards the collected health data or a partial local decision to an aggregator (also can be a clusterhead, gateway, sink). The sensors in the BSN are mixed with sensors of the IoT network.

3.1.2. IoT Networks

An IoT network consists of a number of sensors, including diverse sensors, wearable things, health equipment, network sensors, and our daily-life appliances, and so on. The IoT network maintains the connectivity between the sensors of the BSN, stores data, and maintains data exchange. On the one hand, the IoT network also maintains communication with the remote computing and storage environments such as cloud and other networks, including cell networks, as shown in Fig. 4b. On the other hand, BSNs also forward the data to the cloud and provides facilities for different application platforms [9,17,21]. In terms of communication, sensors in the BSN have short-range low-power wireless communication components and micro-battery. Besides other constraints, the energy cost of wireless communication of sensors of the BSN is limited [32].

3.1.3. Healthcare data users

A designated data user, receiver, or viewer say a healthcare application specialist or provider (doctor) with the Internet collects the stored data from the IoT storage or cloud through IoT-enabled user interfaces. Using some security access control [22,23], the healthcare providers get access to the patients data and monitor the patients health performance. It is expected that they want to view the data that should be dependable.

3.2. Security attack models

We assume the *data manipulation attacks* occurred on the measurements of body sensors in the BSN [12]. An attacker alters the signals (such as ECG) measured by body sensors by manipulating the sensors thus introducing a wrong view of the patients' health conditions. An effective data manipulation may cause severe harm to patients' health,

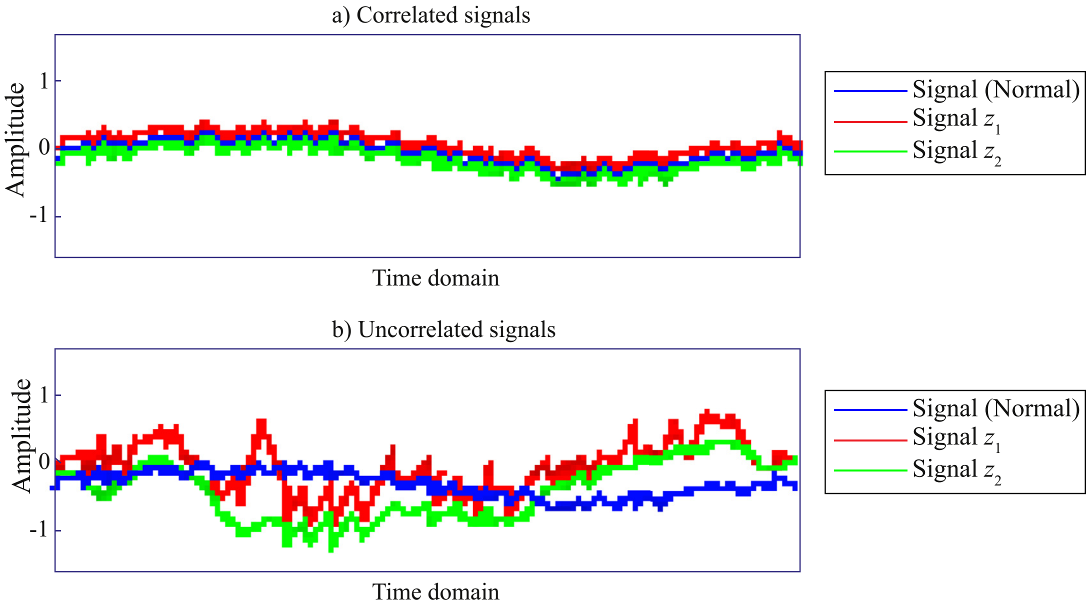


Fig. 4. Representative correlated and uncorrelated signals of three signals under both normal and compromised one.

resulting in undependable data. This leads to useless health attention for the patient or wrong health diagnosis by the doctors. Due to the undependable data, a compromise happens to a sensor's signals that may have the results similar to the results from the data manipulation attack. The *signal manipulation attack* leverages measured signals by applying inter-relationship between patient physiological signals (ECG). We also consider that ECG measurements in the BSN can be compromised by: (1) adding random noise to the actual ECG signals, and (ii) replaying historical ECG signals obtained from the healthcare data user previously [11].

Such attacks on medical sensors can lead to transmission vulnerabilities. These affect the performance of devices, such as pacemakers and insulin pumps [12]. For example, BSNs are vulnerable to the entire category of sensor communication channel threats/attacks. These interfere with the body sensors transducers and bring random sensing estimation into the sensing. These kinds of attacks ought to not only be applied to mitigate the measured signals (such as accelerometer signals), but also allow random code expectation in the case of a particular condition. Some IoT-based healthcare organizations and industries do not usually anticipate the software/firmware update and records in the course of on-field updating. Then, an attack on the manufacturers servers may be applied to attack the body sensor devices at the time of software updating.

We also consider some related attacks such as *hardware-based attacks* and *physical attacks*. An attacker may physically compromise one or more sensors in the BSN or one of the sensing units of a sensor or afterward alters their firmware. An alternative attack called *service integrity* is to replace a sensor with a comprised one by setting the same id or similar functional features. An example of such attacks can be on fitness monitors, including Fitbit which can be installed with malicious file via a public Bluetooth port. Also, we assume *signal injection attack* in the BSN, where an adversary may inject low-quality, unrelated, or compromised signals through various types of stimuli, including electromagnetic induction, light, and acoustic wave.

3.3. Data dependability model

Though a BSN and an IoT network help consolidate the whole health infrastructure on a single platform and facilitate patient health monitoring security, and privacy of the IoT health sensors, and its network, and their communications, data exchanges are still serious concerns to the users [11]. We assume that the authenticity and confidentiality of patient health data transferred via the common channel are the two utmost vital security requirements in every sensor application. Cryptography is used as the key solution to tackle these two issues in the BSN [9,11]. However, we assume that protection to data, computing, and communication is not guaranteed, even we are still far from guaranteeing data security and quality, due to various constraints with the tiny sensors, including computational time and complex algorithm, radio communication, resource as well as cost limitations.

Thus, it is significant to address data dependability concerns, in terms of data accuracy, data quality, real-time data, and secured data sharing, real-time situation assessment, data usefulness, and many other issues needed to assist the decision-makers and healthcare experts. Lacking the data dependability, the usefulness of data in BSN-enabled healthcare applications turns out to be weakened as patient health situations or decisions on the patient health status made from the collected data cannot be trusted with dependence/confidence. In order to make decisions whether or not the collected data is "truth" or "good" to the healthcare users, we consider addressing data dependability problem with the decision-making in three layers.

3.3.1. Decision-making on the data dependability at acquisition

It is important to guarantee the signals acquired by the BSN based healthcare sensors (such as body sensors) are the truth data, that is, we do not like to have patient health information (ECG signals, pressure signals) compromised at the time of acquisition due to some security attacks such as data integrity, data manipulation, system integrity, and sensor integrity problems. We consider data dependability to verify the "truth" of acquired signals in order to answer, "is the signal acquired by the sensor is really what it is supposed to be?. This layer of signal dependability verification is done before the signals is locally stored, processed, or transmitted to the aggregator. We apply a three-layer decision-making framework on signals that are being acquired by a sensor of the BSN to guarantee that the collected signals are dependable.

3.3.2. Decision-making on the data dependability at the local aggregator

As shown in Fig. 4, after the signal dependability verification above, we often aggregate the collected data from a set of sensors at an aggregator. Therefore, in data transmission, other networked sensors in the BSN may interfere with the sensors and the data may face integrity problem. Snipper radio sensors can be manipulated. There can also be compromised hardware sensors with the IoT healthcare sensors that can modify the transmitted data during the aggregation locally, inject suspicious information into the data packet, adding instructions to the process so that the transmitted data may become meaningless. They may bring various security threats to IoT health sensors. Some sensors constantly offer dependable data, while other sensors may produce biased, compromised, or even fake data [1]. There can be a question among patients' or healthcare users: "is my data protected? Without identifying any alterations of the transmitted patients data towards the local aggregator, effective aggregation may not happen. A decision-making layer can be applied on data being received by the local aggregator to verify that data received by the local aggregator is dependable.

3.3.3. Decision-making on the data dependability at the data user

After the data dependability verification in aggregation, the aggregated data will be forwarded to IoT storage or cloud server over the cloud service provider (CSP) for further processing and storing. The BSN-enabled healthcare users will access the data using security protocols and utilize the data for patient health monitoring and make decisions on the patient health. There could be various security concerns over the long-way data transmission due to numerous security threats/attacks. Whether or not the transmitted data is verified as dependable before the aggregation, the data might be modified again at the time of data transmission from the BSNs to an IoT storage or cloud environments over the third-party service provider, and before the data access by the data user.

A man-in-the-cloud attack can easily happen in the cloud. Such an attack puts more emphasis on the data manipulation and theft of a user's cloud synchronization token. The victim data user is normally triggered with malware through some malicious website or email. Then, the attacker can gain access to the data when a user download forms the cloud. Moreover, there could be a hash value manipulation attack at the files stored in the cloud. There could also be data Breaches, data loss or errors, data compromised by the hijacked cloud accounts or compromised credentials, compromised user interfaces or API, and DDoS attacks. Without verifying any changes in the transmitted health data, the health data accessed and used by the healthcare provider or doctor can be undependable. Undependable health data negatively influences the overall patient health monitoring quality and patients wellbeing. A decision-making framework should be used on the data being received by the user/owner layer to guarantee that data received by the user/owner is dependable so that undependable data access can be avoided.

The objectives of this work are to increase the data dependability in patients health status monitoring and reduce false decision-making in the patient health diagnosis by healthcare providers, and meaningless data transmission.

4. 1-DM: The 1st-Layer decision-Making for data dependability verification

4.1. Main idea

Generally, data from sensors in the BSN can be compromised at many layers, namely during the data acquisition, processing, transmission, storing, aggregation, and decision-making. Among them, the first and foremost layer is the data dependability verification at the signal measurement layer, which can confirm the status of sensor signals in the BSN in the initialization. However, the quality of signal measurements may be influenced by the application-specific requirements, such as signal measurement environment, data resolution, etc. We attempt to consider similar situations with signal measurements.

The main idea is to analyze each of the measured signals of a sensor compared to a set of latest signals (maybe also signals from history), and a set of signals from a neighboring sensor for a signal measurement cycle. Based on the measured signals, each sensor itself and its neighboring sensor can indicate if there is an irregular signal due to security attacks on the signals.

4.2. Signal measurement in the BSN

We suppose that a health sensor may have one to multiple sensing units collecting the multiple signals. Sensor i compares its signals collected in a measurement cycle with the measured signals of the latest cycle, and current signals of one of its neighboring sensors. We assume that each cycle is broken down into a set of discrete small sampling intervals. In each interval, body sensor i compares signals with own latest signals and unicast its acquired signals to a one-hop neighboring body sensors.

Let z_i^t denote the actual signal of sensor i at each discrete sampling cycle, $t \in T$, $t = 1, 2, \ldots$, where T_t is a signal collection round and t is a discrete signal collection cycle. Then, the measured signals of sensor i are specified as follows:

$$z = z_i^t + \epsilon \tag{1}$$

The signal measurement noise denoted by ϵ for a sensor might be random noise in practical health signal measurement. We take z measured signals to be transmitted to one-hop neighboring sensors. But it may also be influenced to some extent by a possible security attack on the signals.

Assume that a subset of signals is not security compromised at a cycle t in patient body health monitoring application. In the BSN-enabled healthcare application, when all of the measured signals are not security compromised, it is no problem to make a decision on the accurate signals and decide that the acquired signals are dependable. However, if one or a subset of measured signals is compromised, a sensor i may decide the measured signals is security compromised by analyzing latest signals, such as those from history or signals of its one-hop neighboring sensor, and by using correlation statistics and the extent of ϵ.

Correlation of measured signal is a metric of two acquired signals of the patient health of the BSN. Correlation is a broadly used notion in signal processing, which is a degree of in what way two signals are equivalent. It can be calculated by multiplying the two signals and summing the result over a given signal acquisition cycle. As shown in Fig. 4, the two signals look indistinguishable and, therefore, their correlation is within -1 or 1. But the two signals of patient body are uncorrelated, so distinguishing one of the signals cannot offer any information on the other.

4.3. Signal dependability verification

Mutual Signal Information Independence (MSII). *It is a function denoted by* $\Phi()$ *that calculates the amount of correlation or statistical dependence between two signals sets Z and Z′. In other words, the function is used to quantify the deviation between signals from a signal correlation pattern.*

We first think that there can be security attacks on the patient body signal measurements. Consider that a security attack on patient body signals may occur at any time t. A subset $D \subset P$ of signals of patient body is perhaps security compromised, which are mixed with other collected signals making signal collection undependable. To make a decision on identifying compromised signals, the followings are used to distinguish the signals:

Z= body signals supposed to be *dependable*

Z’ = signals supposed to be compromised (*undependable*)

$Z \cap Z' = \{\}$ and $Z \cup Z' = D$, where $D \subset P$

Normally, we assume that other signals of patent health in a cycle may be irregular due to attack or many reasons. However, regarding the patient health monitoring, we focus more on uninterrupted monitoring of signals of a sensor whose signal behaviors are altered remarkably (due to an attack, fault, or other reason).

We apply mutual signal information independence (MSII) as a function of indirect patient body signal measurement. We suppose that a correlation pattern CP of measured signals exists [24]. We can provide CP as a standard signal subset, which is temporally buffered data in the sensors memory acquired when there are no security attacks. The MSII amongst two signals q and r at any time t in subset of signals D, is given by:

$$\Phi(z_q, z_r, CP) \tag{2}$$

We take the measured signals in Z and Z' as the latest and current health signals of a patient, respectively. Hence, MSII of signals of Z and Z' can be specified as follows:

$$\Phi(z_Z, z_{Z'}, CP) \tag{3}$$

Given Q consecutive signals, the MSII is used to estimate the correlation between z_Z and z_Z at time t as:

$$\Delta(Z, Z') = \sum_{t=1}^{Q} \Phi(z_{Z'}, z_Z, CP) - \sum_{t=1}^{Q} \Phi(z_Z, z_Z, CP) \tag{4}$$

$\Delta(Z, Z')$ is obtained by decreasing the difference amongst normal signals in Z, and by increasing the difference amongst normal signals in Z and compromised signals in Z'. D is regulated by the BSN user regarding signal density. Intuitively, normal signals should be consistent with each other, while the compromised should be inconsistent. Note that for generalization we do not adopt that the compromised signals can be uncorrelated.

4.4. Decision-making on the signal dependability verification

This subsection explains normal signals collection and an algorithm to detect compromised signals of patient health, according to the model described earlier. The layer is described in Fig. 5.

Body health sensors signals are likely to be compromised by security attacks. We apply a joint Gaussian distribution-based correlation pattern. We find that multivariate Gaussian distribution is applied to precisely pattern the correlation of different kinds of signals in the literature [25]. Every measure signal is compared with the latest signals and the signal unicasted to its one hope neighbor. Let sensor ith signals be $z_i^t \in z_D^t$ and jth signals be $z_j^t \in z_D^t$; $i, j \in D$ and $Z \subseteq D$ and $D \subset P$. For the convenience's sake, z_i^t as q and z_j^t as r are denoted hereafter.

We calculate the statistical signal independency amongst signals q and r that are described in the joint probability density $p(q, r)$ of measured signals in the following:

$$p(q, r) = \frac{1}{2\pi\tau_q\tau_u\sqrt{1-\rho_{qr}}} e - \frac{1}{2(1-\rho_{qr}^2)} \left[\left(\frac{q-\mu_q}{\tau_q^2} \right)^2 - 2\rho_{qr} \frac{(q-\mu_q)(r-\mu_u)}{\tau_q\tau_u} + \left(\frac{r-\mu_u}{\tau_u^2} \right)^2 \right] \tag{5}$$

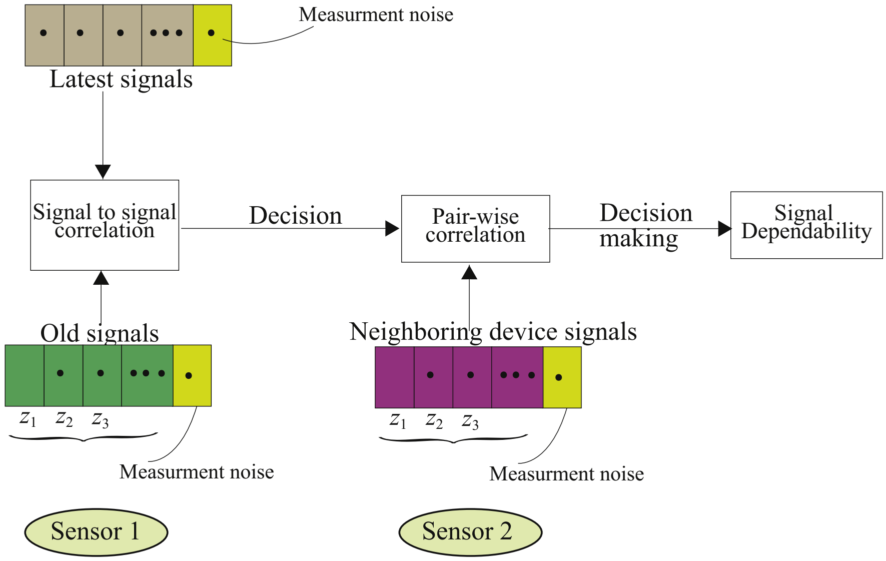


Fig. 5. 1-DM: the first layer decision-making.

Here, the averages and the standard deviations of the measured signals q and r are denoted as μ_q, μ_u, τ_q, and τ_u, respectively. We compute the correlation coefficient amongst two signals denoted by ρ_{qr}. We present it as follows:

$$\rho_{qr} = \frac{E\{(q-\mu_x)(r-\mu_y)\}}{\tau_q \tau_u} \tag{6}$$

We can also utilize the correlation coefficient to decide if signals q and r are statistically independent. If $|\rho_{qr}| = 1$, we think that there is a sufficient correlation amongst q and r. Alternatively, if $|\rho_{qr}| = 0$, signals q and r have no sufficient correlation. This type of correlation is assumed as poor statistical dependency between signals. According to literature [26], it can be seen that two random variables, which have no correlation amongst them, are to be statistically related or dependent. Regarding this situation, we put emphasis on the statistical dependency or independency amongst signals. We take ρ_q and ρ_u of the signals q and r of the product of the marginal densities, respectively, which is stated as follows:

$$p(q,r) = p(q)p(r) \tag{7}$$

To make a decision on the signal dependency/independency, we compare the (5) and (7). If the outcome of (7) is identical to the outcome of the marginal densities in (9), we decide that the measured signals can be solely independent. To confirm this, we quantify the statistical dependency amongst signals q and r, which is to calculate the MSII of q and r and is given in the following:

$$\Phi(q,r,CP) = \int\int p(q,r)\log\frac{p(q,r)}{p(q)p(r)}dq\,dr \tag{8}$$

The logarithm basis regulates the units in which signal information is measured. Based on the information in (3), it can be depicted that if signals q and r are not dependent, Φ happens to be zero.

We think of a forward approach to split the range of q and r into some finite bins and we quantify the amount of sampled signal pairs of $h_o = (q_o, r_o), o = 1, 2, \cdots, n$, which fall into the finite bins. The quantification permits nearly estimating the probabilities through substituting (9) by the finite sum:

$$\Phi_{bin}(q,r,CP) = \sum_{a,b} p_{qr}(a,b)\log\frac{p_{q,r}(a,b)}{p_q(a)p_r(b)} \tag{9}$$

Here $p_q(a) \approx n_q(a)/Z$ and $p_q(b) \approx n_q(b)/Z$ are the likelihoods on the amount of points $n_q(a)$ and $n_r(b)$ that falls into ath bin of q and the bth bin of r, respectively. Then, we can have the joint probability as $p_{qr}(a,\ b) \approx n(a,\ b)/Z$ taking the amount $Z(a,\ b)$ of points falling into the box number a and number b. The MSII should be non-negative and symmetric:

$$\Phi(q,r,CP) = \Phi(r,q,CP) \geq 0 \tag{10}$$

The MSII for all likely mixtures of outcomes of BSN-enabled health device signals as z_u and z_v (with the exception of $u = v$, where $i = 1, 2, \cdots, u, j = 1, 2, \cdots, v$) is estimated. This guides to an Φ-matrix for the total possible mixtures of r and s. The main concept here is that the MSII alters when f_v is appeared, which is due to be a security attack on sensors signals or other reasons. We infer that f_v can be in rth channel or index:

$$\tilde{z}_u = z_u + f_v \tag{11}$$

This change happens just in the rth channel. As a result, we can find that all possible combinations with index r have to demonstrate a drop in the Φ. This help us make a decision on the compromised signals of a sensor. More signals can be concurrently identified, accordingly. We can visualize the compromised signals by using the relative change as a compromised signal indicator denoted by $\Theta^{\Phi}_{z_u}$, given by:

$$\Theta^{\Phi}_{z_u} = \frac{|\Phi_{z_u} - \Phi_{nor}|}{\Phi_{z_u}} \tag{12}$$

where z_u is a real signal set and the lower index *nor* is one normal reference signal set. The layer grounded on the MSII is capable to identify compromised sensors signals in different combinations of them.

4.5. Algorithm 1: Undependable signal detection

In every decision-making cycle, a body sensor makes a decision on the compromised signals, mainly on the basis of current signals, k latest signals, and signals obtained from the one-hop neighboring device.

In Algorithm a, if a decision made locally on a sensor measured signals, $\Theta^{\Phi}_{z_u} > 0.5$, then the device's some or all of its signals might be compromised. We imply that MSII can be highest on the device's collected signals. This decision-making requires at least a neighboring sensor to be synchronized. In addition, the decision-making on signal detection is nearly direct and real-time, since a sensor does not demand to keep waiting for the signals of a neighboring sensor. Furthermore, the identified set of compromised signals is not transmitted towards the aggregator, therefore, the transmission cost may become comparatively low.

The MSII should not depend on a specific kind of security attacks. Algorithm 2 depending on the MSII may identify different kinds of compromised signals. Note that, the MSII might fail to identify missing signals from a body sensor or sensor failing.

Algorithm 1 1st-Layer Decision-making on the Signal Dependability.

```
Decision: $(\Theta^{\Phi}_{z_u} \le 0.5$: dependable signal), $(\Theta^{\Phi}_{z_u} > 0.5$: undependable signal)
for device $i \in Z$ where $Z \subseteq D$ do:
    $(\Theta^{\Phi}_{z_u})_i \longleftarrow 0$ // each sensor $i$'s signal is dependable
loop
  for every device $i$ do:
    $(\Theta^{\Phi}_{z_u})_{j(neighboringSensor)} \longleftarrow$ receive neighboring $j$th signals
    samples $\longleftarrow$ $Z$ samples from $(\Theta^{\Phi}_{z_u})_{j(neighbor)}$
      for every sample $u$ do: // $u$ is an index
        $G_u \longleftarrow$ normal signals
        $C_u \longleftarrow$ compromised signals
        $C_u \longleftarrow (G_u, C_u)$ //Eq. 3
  for every device $i$ do:
    if $\Theta^{\Phi}_{z_u} > 0.5$ then
        $i$ indicates as producing compromised signals
    if $i$ is not acting for signals then
        $j$ indicates $i$ as a compromised signal producing device
end
```

Algorithm 2 Signal Collection at the Aggregator for Dependability Verification.

```
Input: Signal transmission within a given neighborhood range
Output: Decision on the undependable signals in the neighborhoods
        $t \longleftarrow$ signal transmission at any time slot
       $i$ transfers own signals to the aggregator device $j$
       for each signal collection cycledo:
         for everybody sensor $i$ do:
           If  $i$'s signal is transmitted to $j$ in $t$ then
              transfers a new signal to $j$
       run Algorithm 3
```

5. 2-DM: The 2nd-Layer decision-Making for data dependability verification

In the previous section, we have verified data dependability at the signal level and then the body sensors are supposed to transmit the data to an aggregator. In this section, we have verified data dependability at the local aggregator, which is done before making the aggregation. A decision can be made in two phases: (1) examining whether or not the data sent by the sensors in the BSN is authentic; (2) guaranteeing that the data sent by the sensors in the BSN is not altered after data acquisition. However, we do not cover the authentication phase in this paper. We assume that there are conventional security mechanisms and protocols used in data transmission from the BSN-enabled health device toward an aggregator.

In this section, we carry out data dependability verification of the data transmitted by the individual BSNs. The verification is carried out before the data aggregation with the aim that the data found with dependability concern should be included in the data aggregation (such as sum, count, avg). A number of body sensors can in parallel send the collected data toward the aggregators. With the data dependability verification, patients health data aggregation can be effective.

The local aggregator calculates the MSII for signals, and selects the signals with the greatest independence for dealing with the security attack. Distributed decision-making on the signal dependability is not appropriate for resource-limited IoT health devices. If every IoT device demands to transmit all of the acquired signals, including compromised ones to the central station (where each sequence of signals can be from $X0$kb to $X000$kb, $X = 1, 2, \ldots$), the centralized BSN-enabled healthcare application might not be capable to function properly for a given period of time. On the contrary, a decision on the compromised signal acquisition may run distributedly when each of the device has a decision on the acquired signals locally, as shown in Fig. 5. The decision also indicates signal dependability of an IoT device based on the signals from a neighboring device and decides whether the devices collected data is compromised.

We apply theoretical procedures to verify the data in one-hop to multi-hop communication in the DependData framework. In the patient healthcare monitoring scenario, the wireless transmission range of health sensors can be constrained, say a few centimeters to several meters in a local neighborhood, in which the IoT health sensors send the signals towards a local aggregator. We bound the sensors to transmit the signals within 1-hop neighboring sensors assuming that transmission within two or more hops is not efficient. In decision-making on the transmitted data dependability, we take sensors in one-hop away from the local aggregator so that a sensor can transmit the data straightforwardly.

As shown in Fig. 6 we see the second-layer of decision-making on the dependability of collected patient health data. Corresponding to the 2nd-layer of Fig. 5, Fig. 6 shows the decision-making framework. Algorithm 2 is used for data collection and decision making, which shows data collection based on the neighborhood.

The signal collection algorithm functions in 2 stages as follows. In Stage 1, each sensor transmits the collected signals to the aggregator, while it also gets acquired signals of the neighboring sensors. In stage 2, each device carries out Algorithm 3, which is called “decision-making for the signal dependability verification. Whether signals are faulty or not can be known through this algorithm.

Algorithm 3 Decision-Making on the Undependable Collected Data.

```
Decision: $(\Theta^{\Phi}_{z_u} \le 0.5$: dependable), $(\Theta^{\Phi}_{z_u} > 0.5$: undependable)
loop:
  for Signals of each body device $i$ in the neighborhooddo:
    $(\Theta^{\Phi}_{z_u})_j \longleftarrow$ receiving signal from $j$
    Signal set $\longleftarrow$ $Z$ signals from $(\Theta^{\Phi}_{z_u})_{j)}$
      for every signal $u$ do:
        $G_u \longleftarrow$ dependable signals
        $C_u \longleftarrow$ undependable signals
        $C_u \longleftarrow (G_u, C_u)$
  for signals of each $i$ do: // at the aggregator
    if $\Theta^{\Phi}_{z_u} > 0.5$ then
       the signal of $i$ is indicated as undependable
    if a signal from $i$ is not received at the aggregator then
       the signal was undependable (then dropped) or was lost
end
```

5.1. Algorithm 3: Data dependability verification

When applying a central signal monitoring framework, the base station or sink node can easily verify compromised signal identification procedures. In a decision-making cycle, the sink can conduct a decision on compromised or abnormal signals, on the basis of k latest signals transmitted by each of the BSN health devices. The sink estimates the MSII for the signals and indicates the signals with the signal independence. The decision is that the highest signal independence indicates the high-sensitivity to the compromised or abnormal signals. However, regarding the IoT-based health monitoring scenario, we think distributed signal verification should be suitable in terms of various constraints and signal verification quality. Also, BSN-enabled health device is severely resource-constrained. It may be infeasible if each body device transmits all its signals to the sink. Rather, we perform compromised signal detection (see Algorithm 3) that runs distributedly.

In the algorithm, the distributed decision-making on the signals is made. If $\Theta^{\Phi}_{y_u} > 0.5$, a signal seems to be compromised or abnormal. This indicates that when the MSII is the highest on the acquired health signals, these signals are compromised. The distributed framework only involves neighboring sensors signals to be synchronized, since an aggregator requires to keep waiting for the signals from local sensors.

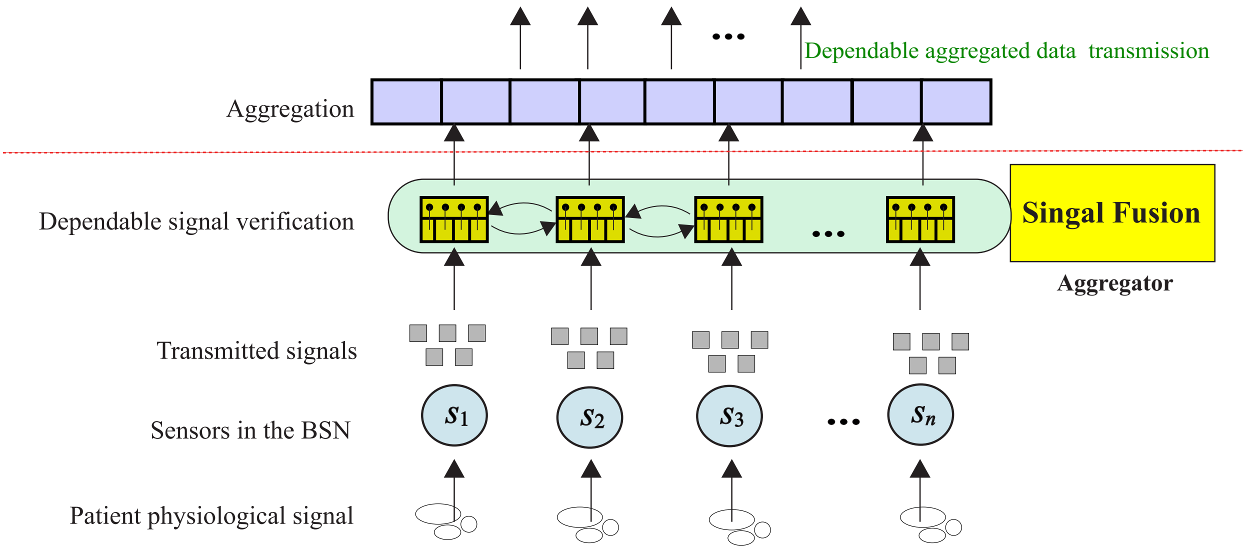

Dependable aggregated data transmission
Aggregation
Dependable signal verification
Singal Fusion
Aggregator
Transmitted signals
Sensors in the BSN
s_1
s_2
s_3
s_n
Patient physiological signal

Fig. 6. 2-DM: the 2nd-layer decision-making.

Furthermore, the set of identified undependable signals are not sent. As a result, the transmission energy can be comparatively minimal.

MSII does not depend on particular security attacks or device abnormalities. Algorithm 3 using the basis of MSII is capable to identify different types of undependable signals (as modeled before).

6. 3-DM: The 3rd-Layer decision-Making for data dependability verification

In the previous sections, we have verified data dependability at the two decision-making layers. In this section, we verify data dependability at the data user layer.

6.1. The 3rd layer decision-making

It is important to offer dependable data to BSN-enabled healthcare users using the BSN-enabled user interface. This is an intrinsically tough problem demanding promising and effective solutions. This needs to associate numerous protocols and schemes, including security protocols, access control protocol, privacy control, etc. At the current state of research, more than ever, data dependability in data collection and data viewing at the data user should be important. We propose a 3rd-layer decision-making for data dependability in DependData for ensuring dependable data access from the cloud storage.

Fig. 7 shows an overview of proposed scheme for ensuring data dependability to the healthcare users. The scheme depicts how data collected from the IoT data storage (cloud) are processed, data dependence status are calculated, and presented to healthcare users. The proposed scheme is comprised of 3 main modules: trustworthiness evaluation, user query and query policy evaluation, and the data quality management. The function of each module is given in the following.

Data dependability evaluation is comprised of data dependence levels with the data stored in the cloud. A data dependence level is a numerical value varying from 0 to 1. Here, 0 implies the lowermost dependability and 1 implies the uppermost dependability. Such a data dependence level can be a crucial concept of DependData framework, as it specifies dependability of the data item. This may be applied for data similarity or quality rating. Data dependence levels would be gained by means of numerous features such as the dependability of data providers and the layer in which the data is acquired. In DependData, we mainly put our emphasis on data dependency.

6.2. Data dependence level

The first phase is to calculate the patient health data dependence level in BSN-enabled healthcare. This calculation is based on the association among all the stored patient health data items in the data storage. This points out the dependability of each data item in the user interfaces. Data dependence level ranges between 0 and 1. A decision on the data dependency level can be as follows.

$$D_j = \begin{cases} 0 & if\ the\ data\ dependency\ level\ is\ poor \\ . & \dots \\ . & \dots \\ 1 & if\ the\ data\ dependency\ level\ is\ noble \end{cases} \tag{13}$$

The goal of knowing this level is to realize the quality of the stored data before a BSN-enabled healthcare data user accesses the data, and it can also be used for data ranking. This level calculation is applied alongside with more circumstances (such as the context of data, real-world factors, history of data) for determining the usage of the data items. The data dependence level calculation scheme is due to the idea of data provenance, which can be utilized as the proof about the data origin, i.e., where and how the data is produced.

The concept behind the level calculation is that the more value of D_j a data provider (such as a cloud server) provides, the more dependable data provider is. Based on this, an *interdependency property* can be found between health data providers and health data items regarding the calculation of the data dependency level. This is to say, the data dependency level value of the collected health data influences the dependable data of healthcare providers, which has initiated and controlled the data, etc. Therefore, the data dependability calculation module is calculated by:

- d as the dependence level value of the data items on the basis of data providers, and
- d_p as the dependence level value of the data providers on the basis of those of data items.

For resolving the multiple data conflict from data providers or multi-sources, we use the truth discovery approach [27,28], which helps to verify the data truth based on the reliability of data provider. This can help mitigate issues, such as when the sum (or average) value of data

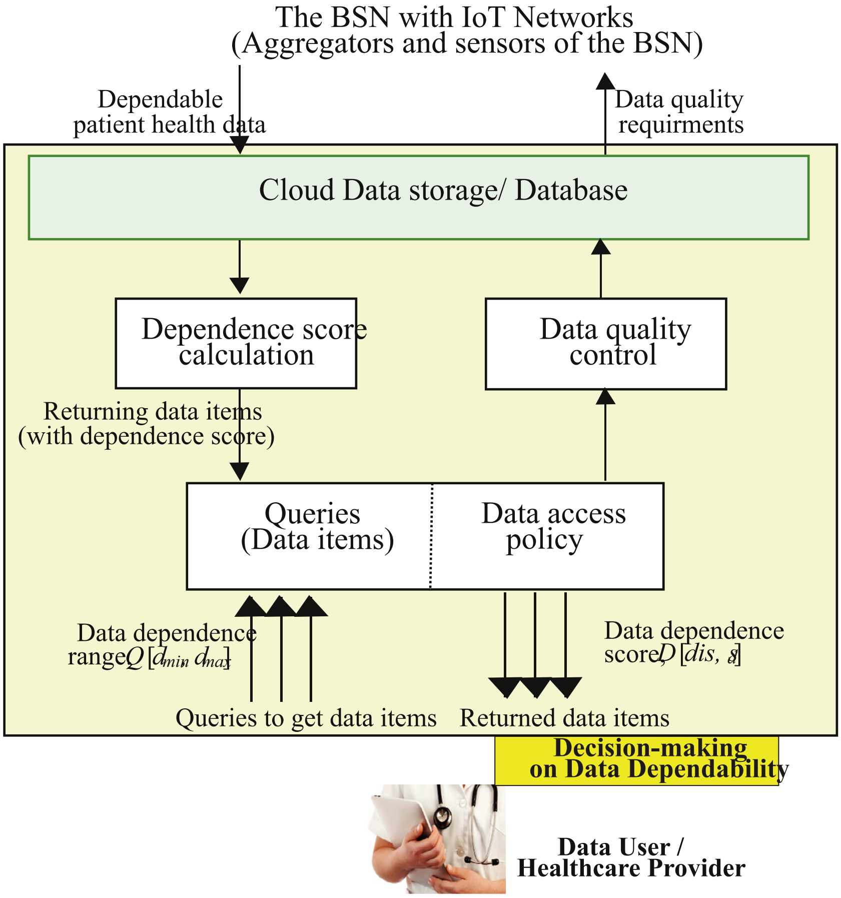

The BSN with IoT Networks
(Aggregators and sensors of the BSN)
Dependable patient health data
Data quality requirments
Cloud Data storage/ Database
Dependence score calculation
Data quality control
Returning data items (with dependence score)
Queries (Data items)
Data access policy
Data dependence range$Q$ [d_{min} d_{max}]
Data dependence score$D$ [dis, s_d]
Queries to get data items
Returned data items
Decision-making on Data Dependability
Data User / Healthcare Provider

Fig. 7. 3-DM: the 3rd layer decision-making scheme.

providers becomes big, even when the data is not dependable. To stay with the space limitation, we do not discuss it in this paper.

6.2.1. Data confidence policy

The second phase is to specify the confidence policy [9,29], which can also be a part of the data dependability. The confidence policy executes user query and data access policy. It can be done some existing work such as [30]. However, different IoT-based healthcare organization and applications has different policies. Nowadays, policies are a big part of the data dependability assurance. Such a policy identifies an array of data dependence level. We apply the policy to a data item, or a set of data items in the application or task. For example, a BSN-enabled health user make a query denoted by Q with a confidence query policy metric $[q_{\max}, q_{\min}]$, then get returned data items where data dependence level ranges from $q_{\max}$ to $q_{\min}$.

6.2.2. The control of the quality of data

It ensures the health data quality in a BSN-enabled healthcare application. The quality can be included in the calculation of data dependability. Based on the healthcare application user, a user may adjust the expected data quality and send a request to sensors to satisfy the data requirement, as shown in Fig. 7.

According to the structure in Fig. 7, whenever a healthcare data user needs to pool patient data from the cloud, it sends a data query to the cloud. The 3rd layer decision-making scheme runs the data dependence range $Q[d_{\min}, q_{\max}]$. With this range, the user can expect to data dependability of the returned data. Then, the request goes to the dependence score calculation. Based on the availability of the data in the cloud regarding $Q[d_{\min}, q_{\max}]$, the available data is selected to be returned to the user. Here, we think different healthcare users may have different data access policies. The returned data then is filtered through the data access policy function. Only the data that is in $D[dis, s_d]$ with the policy is returned to the user.

7. Implementation and evaluations

7.1. Methodologies and datasets

We have implemented `DependData` to validate the performance evaluation in simulations. We have fully developed the first two-layer of decision-making in the BSN-enabled healthcare application. We have partially developed the third layer of decision-making because of limitations in the evaluation environment. We used a computer Core i5 processor, 4 GB of RAM, Win OS 10 with 64-bit and Python language to write programs.

We consider a BSN consisting of 40 body sensors. We have adopted two real data sets of patient health status collected by the body sensors in the proposed BSN-enabled healthcare application. The data sets are from MIT BIH database [31] and IMPACT [17]. / These data sets are used for the 40 sensors of the BSN in our simulations. We divided them into two classes: one includes 20 tri-axial accelerometers and the other one contains 20 ECG sensing sensors. The background patient health signals are simulated as patient health data influenced by the sensors in the BSN. An arbitrary Gaussian noise is injected into all the signals. The average of the noises is set to zero. The standard deviation is set to 10 percent (of actual health status signals of patients). We considered a set of patient health signals from the data sets as *a set of reference signals* to train the joint distribution. We considered another set of similar data for testing. These reference signals are regarded as normal health signals that incorporate signals of patients who have not suffered from any ailments or abnormal sinus rhythm ECG. Both data sets have noise. As a result, the model of trained correlation can reflect the noises.

From both of the patients ECG and tri-axial accelerometer signals, we have randomly selected 4 patients ECG signals with abnormal morphological information and 4 patients accelerometer signals, respectively, and injected compromised signal information (regarding the data manipulation attacks, signal compromised attacks) into signals of those sensors.

In order to detect exactly what happened in the signals, that is to say, to verify the received signal dependability, we used the proposed decision-making in three layers. In the first layer, each sensor used 1-DM to make the signal level data dependability verification locally. All of the data passed through the 1-DM, including the signals that appear to have data dependability concerns are fed into the upstream sensors (aggregators) of the BSN. Aggregators received signals from its sensors in the neighborhood and used the 2-DM to make the data dependability verification in a distributed manner. All the normal and compromised data are stored in a database. When a user requests to get data with a data dependence, the 3-DM is used to make data dependability verification at the user/owner layer.

7.2. Performance measures

For comparison, we did not find any appropriate existing scheme to compare to. Using the simulation results, we evaluate and compare the performance of DependData in several measures under the security attack information injections:

- *Performance of the MSII values*. MSII values of any two signals are calculated by the mutual dependence between the two signals. More precisely, it quantifies the “amount of information” gained about one signal through monitoring the other signal.
- *Data dependability in terms of detection.* It is defined by the detection ability of DependData that can provide us an indication of how much the DependData can cope with the security attacks on the signals. Here, the detection ability is the percentage of the detection of compromised signals over the total amount of injected compromised signals.
- *Data Dependability Importance.* To determine the overall dependability of DependData, we consider the role of identifying the least dependable decision-making functions of DependData in order to improve future BSN for healthcare system design. By means of the reliability importance measures, we find the relative importance of each of the decision-making of DependData with respect to the overall dependability of the DependData. According to the reliability importance [14], the data dependability is given by:

$$I_{D_l} = \frac{D_d}{D_{l-(DM)}}, \qquad l = 1, 2, 3 \tag{14}$$

where: D_d is the data dependability of DependData and D_l is the dependability of each of the decision-making layers.
However, we consider DependData’s three-layer decision-making scheme reliability-wise in series and then dependability of DependData can be calculated as:

$$I_{D_l} = D_{1-DM} * D_{2-DM} * D_{3-M} \tag{15}$$

7.3. Simulation results

7.3.1. Performance of the signal analysis and MSII values

In the initial set of simulations, we can see real ECG output and tri-axes accelermeter output. Fig. 8 demonstrates the signal output waveforms of ECG and accelerometer, respectively. The ECG signals deliver a record of compound electrical events happening in the heart.

We execute the first two layers of decision-making under the compromised (undependable) signal injection detection. Each sensor of the BSN acquires signals and receives from its 1-hop neighboring sensors. This includes undependable ECG signals with morphological abnormalities and tri-axial accelerometer signals. Each sensor updates its MSII based on its signals, previous signals, and received signals. Fig. 9 shows the MSII achieved through the 1-DM and 2-DM layers in the first two successful simulations (Sim1, Sim2).

There are no undependable signals in body sensors #1, #2, #4, #6, #7. This implies that the gathered data is distorted nearly in most of the sensor’s signals by undependable signals sharing between any two sensors’ signals except for such sensors as #3, #9, #14, #19, which show the highest signal information. The reason is that the signals #9, #14 only influenced by the signal of their neighboring sensors. It is observed that high MSII values exist at some sensors of the BSN, including sensor #3, sensor #7, and sensor #13. Their ECG signals become untrustworthy or partially changed that is evidently noticed. In the 2-DM layer, when the amount of undependable signals rises, it can be seen that the values of the MSII at body sensors of the BSN in the neighborhood become maximum. This justifies the precision of the undependable signal detection at the aggregator each time the aggregator receives such data. The dependable signals are detected after executing the 1-DM or 2-DM.

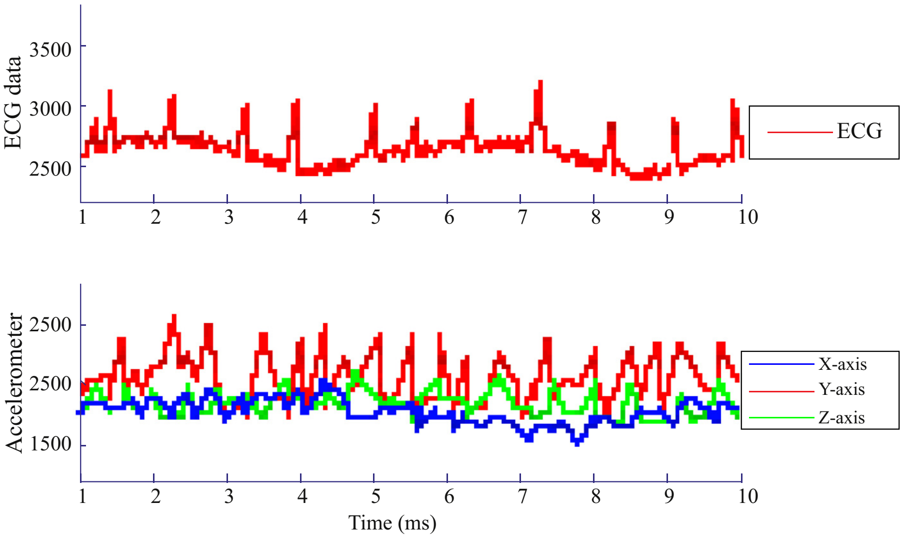


Fig. 8. Snapshots of the real ECG and accelerometer signals, respectively, under a patient’s normal health states.

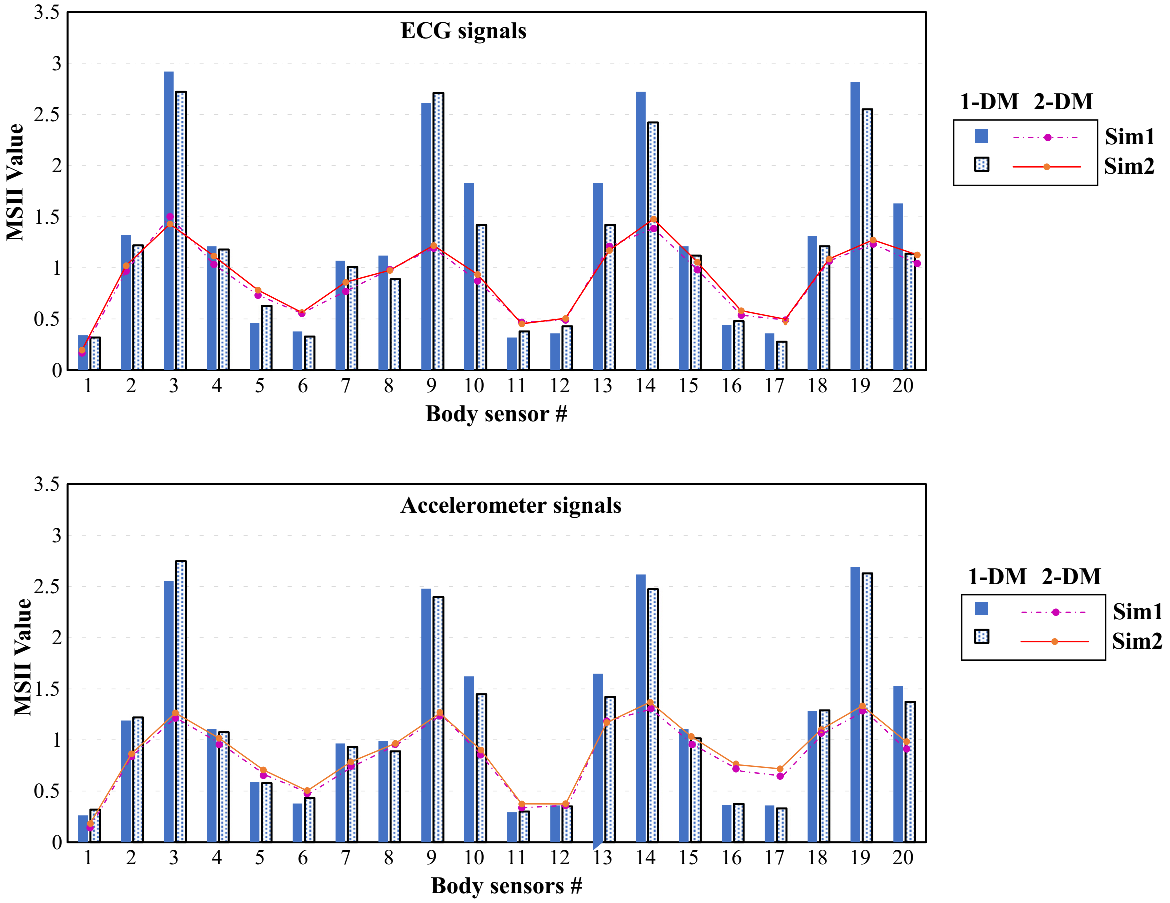


Fig. 9. Performance of both 1-DM and 2-DM layers analyzed through the MSII under security attacks.

These signals may be dropped before aggregation or may be dealt with by other approaches such as signal reestablishment process to reestablish the signals.

7.4. Data dependability verification

Given the dependability nature of compromised signal injections on the sensors, we justify to identity the potential changes to the ECG signals and tri-axial accelerometer signals very quickly so as to evade potential harm to the patients, as shown in Fig. 8. To detect exactly what happened in the ECG and tri-axial accelerometer signals, we can notice that the acquired signals of the compromised sensors is included in the patient health status. Thus, the health status corresponding to the compromised signals are altered drastically. Considering compromised signals, we want to realize accurately what happened in patient health status so that we can realize whether the BSN-enabled health monitoring is dependable or not.

We next observe the ECG collection in simulations. In the result analysis, we have applied a mixture of the true positive and true negative detection outputs in the device signal attack detection accuracy estimation. We gather all the false positive and false-negative cases that appeared (obtained from a total of 30 simulation runs), and we obtain an average. Then, we compute the data dependability in terms of compromised signal detection ability rate as:

$$1 - [(false\ positive\ rates + false\ negative\ rates)] \tag{16}$$

This is shown in Fig. 10. We can see the impact on the signals, in which the compromised signals are distorted under the security attacks and can see the detection ability of 1-DM, 2-DM, and 3-DM. This indicates that, if there is no proper undependable signal identification technique as well as protection to the undependable signals, achieving successful BSN-enabled patient health monitoring operations will be difficult. We can say the BSN-enabled health monitoring without addressing the data dependability will not be dependable.

7.5. Dependability of $\mathtt{DependData}$

Finally, the dependability importance values of the decision-making at the three layers are computed. As shown in Fig. 11, the data dependability is achieved by the dependability of each of the decision-making layers. The increase of dependability in one of the decision-making results contributed to the overall reliability of DependData. The dependability of three-layer decision-making scheme for a given time is 74%, calculated by:

$$D_{D_{1-DM}} * D_{D_{2-DM}} * D_{D_{3-DM}} = 0.74 \tag{17}$$

where $D_{1-DM} = 0.93$, $D_{2-DM} = 0.91$, and $D_{3-DM} = 0.88$.

We observe that data dependability concerns appear in all of the three-layers. If these concerns are not addressed properly, the data quality of the patient physiological signals can be achieved up to 88%. In fact, it can be much lower than 80% if the data dependability rate is low in any of the decision-making layers. It is evident that the dependability of each of the DependDatas decision-making layers demand to be increased in order for the BSN-enabled healthcare application to meet its goal.

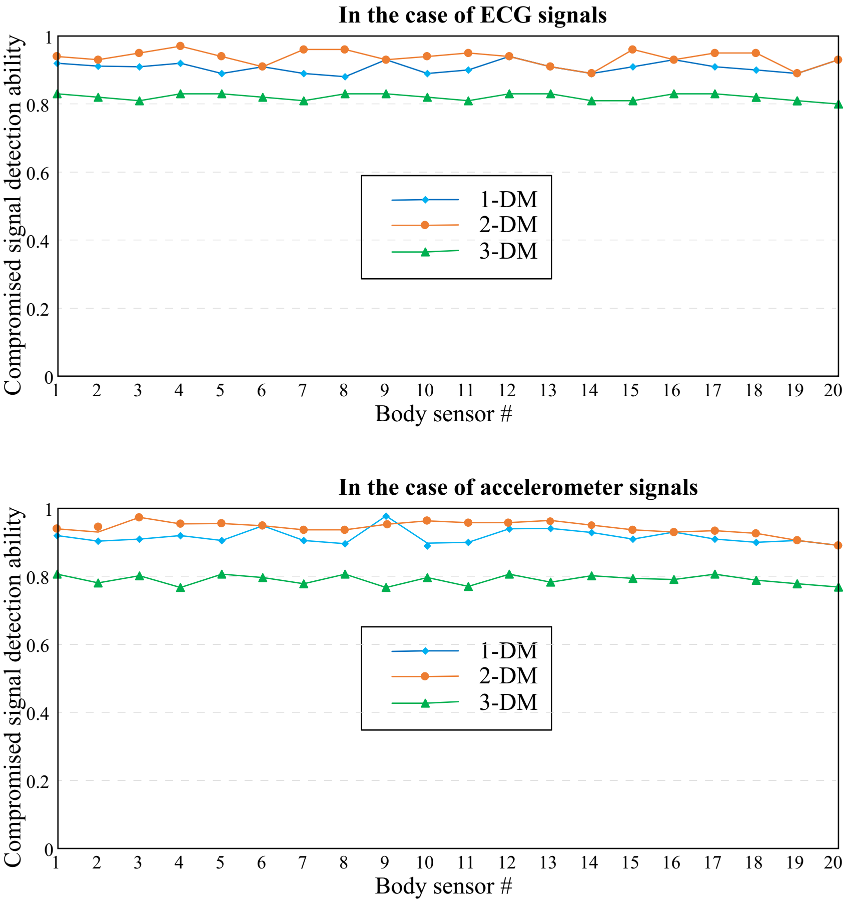


Fig. 10. Compromised data detection ability through the decision-making in three layers of `DependData`: in the case of ECG and accelerometer signals.

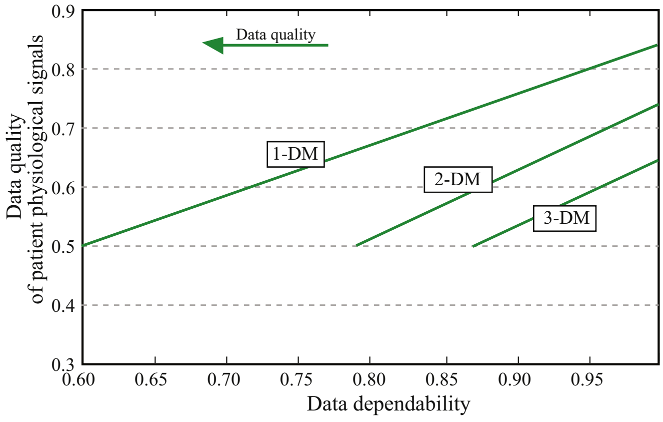


Fig. 11. Overall Dependability of `DependData`: this illustrate data quality of the collected patent physiological signals.

8. Conclusions

In this paper, we have introduced a comprehensive dependable data verification framework named `DependData` to guarantee data dependability in the BSN-enabled healthcare application. `DependData` has come with a novel concept: data dependability verification before data utilization, i.e., the healthcare data user can check whether or not the collected data is dependable and whether they will use this data or not. Alongside with current studies on security and privacy protocols and algorithms, `DependData` has attempted to verify the data dependability concerns regarding that the data may still be compromised at the acquisition and before aggregation/storing in severely resource-constrained BSNs. This leads to data collection scheme becoming meaningless or undependable i.e., an undependable BSN-Health. `DependData` includes three-layer decision-making to verify data dependability in the three layers: signal level dependability at each sensor locally; data dependability at each of the aggregator of the BSN in a distributed manner; data dependability verification before the user views the data in IoT-enabled interfaces. It is worthwhile to note that, in the `DependData` framework, a significant amount of undependable (untrustworthy, meaningless) data can be reduced before processing and transmission in all three layers. `DependData` is very general and can easily be applied to different application areas.

Disclosure of Conflicts of Interest

The authors declare that they have no known competing financial interests or prsonal relationships that could have appeared to influence the work reported in this papaer.

CRediT authorship contribution statement

Tao Hai: Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Md Zakirul Alam Bhuiyan:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Jing Wang:** Data curation, Investigation, Methodology. **Tian Wang:** Investigation, Validation. **D. Frank Hsu:** Data curation, Writing - review & editing. **Yafeng Li:** Methodology, Project administration, Supervision. **Sinan Q Salih:** Investigation, Validation. **Jie Wu:** Conceptualization, Writing - review & editing. **Penghui Liu:** Resources.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (No. 61971005) and the Natural Science Basic Research Plan in Shanxi Province of China (No. 2018JM6043), in part by the Shaanxi Provincial Department of Education Project (16JK1051) and the Doctoral scientific research initial funding project of Baoji University of Arts and Sciences (ZK2018062), and in part by the Key Research and Development Program in Shaanxi Province (2019GY-131).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.inffus.2020.03.004.

References

[1] E. Luo, M.Z.A. Bhuiyan, G. Wang, M.A. Rahman, J. Wu, M. Atiquzzaman, Privacyprotector: privacy-protected patient data collection in iot-based healthcare systems, IEEE Commun. Mag. 56 (2) (2018) 163–168.

[2] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.

[3] G. Fortino, S. Galzarano, R. Gravina, W. Li, A framework for collaborative computing and multi-sensor data fusion in body sensor networks, Inf. Fusion 22 (2015) 50–70.

[4] Health risk assessment and decision-making for patient monitoring and decision–support using wireless body sensor networks, Inf. Fusion 47 (2019) 10–22.

[5] G. Yang, L. Xie, M. Mntysalo, X. Zhou, Z. Pang, L.D. Xu, S. Kao-Walter, Q. Chen, L. Zheng, A health-iot platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box, IEEE Trans. Ind. Inf. 10 (4) (2014) 2180–2191.

[6] E. Kanjo, E.M.G. Younis, N. Sherkat, Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach, Inf. Fusion 40 (2018) 18–31.

[7] R.A. Khan, A.-S.K. Pathan, The state-of-the-art wireless body area sensor networks: a survey, Int. J. Distrib. Sens. Netw. 14 (4) (2018) 1–22.

[8] T. Wang, M.Z.A. Bhuiyan, G. Wang, M.A. Rahman, J. Wu, J. Cao, Big data reduction for a smart citys critical infrastructural health monitoring, IEEE Commun. Mag. 56 (3) (2018) 128–133.

[9] I. Masood, Y. Wang, A. Daud, N.R. Aljohani, H. Dawood, Towards smart healthcare: patient data privacy and security in sensor-cloud infrastructure, Wirel. Commun. Mobile Comput. 2018 (2016) 1–23.

[10] M. Chen, Y. Qian, J. Chen, K. Hwang, S. Mao, L. Hu, Privacy protection and intrusion avoidance for cloudlet-based medical data sharing, IEEE Trans. Cloud Comput. (2019) 1–9.

[11] A. Banerjee, S.K.S. Gupta, K.K. Venkatasubramanian, Pees: Physiology-based end–to-end security for mhealth, in: Proceedings of the 4th Conference on Wireless Health, in: WH '13, 2013, pp. 2:1–2:8.

[12] H. Cai, K.K. Venkatasubramanian, Detecting data manipulation attacks on physiological sensor measurements in wearable medical systems, EURASIP J. Inf. Secur. 2018 (1) (2018) 13.

[13] C. Boswell, S. Cannon, Introduction to Nursing Research: Incorporating Evidence-Based Practice https://en.wikipedia.org/wiki/Mutual_information.

[14] A. Mettas, in: Reliability allocation and optimization for complex systems, 2000, pp. 216–221.

[15] D. Curone, A. Tognetti, E.L. Secco, G. Anania, N. Carbonaro, D. De Rossi, G. Magenes, Heart rate and accelerometer data fusion for activity assessment of rescuers during emergency interventions, Trans. Info. Tech. Biomed. 14 (3) (2010) 702–710.

[16] X. Yi, A. Bouguettaya, D. Georgakopoulos, A. Song, J. Willemson, Privacy protection for wireless medical sensor data, IEEE Trans. Dependable Secure Comput. 13 (3) (2016) 369–380.

[17] P.K. Sahoo, Efficient security mechanisms for mhealth applications using wireless body sensor networks, Sensors 12 (9) (2012) 12606–12633.

[18] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, G. Wang, Security and privacy in the medical internet of things: a review, Secur. Commun. Netw. 2018 (2018).

[19] J. Yu, K. Wang, P. Li, R. Xia, S. Guo, M. Guo, Efficient trustworthiness management for malicious user detection in big data collection, IEEE Transactions on Big Data (2018). 1–1.

[20] T. Wang, Y. Li, G. Wang, J. Cao, M.Z.A. Bhuiyan, W. Jia, Sustainable and efficient data collection from WSNs to cloud, IEEE Trans. Sustainable Comput. (2018) 1–14.

[21] M.Z.A. Bhuiyan, G. Wang, J. Wu, J. Cao, X. Liu, T. Wang, Dependable structural health monitoring using wireless sensor networks, IEEE Trans. Dependable Secure. Comput. 14 (4) (2017) 363–376.

[22] F. Al-Turjman, S. Alturjman, Context-sensitive access in industrial internet of things (IIot) healthcare applications, IEEE Trans. Ind. Inf. 14 (6) (2018) 2736–2744.

[23] T. Wang, G. Zhang, A. Liu, M.Z.A. Bhuiyan, Q. Jin, A secure IoT service architecture with an efficient balance dynamics based on cloud and edge computing, IEEE Internet Things J. 6 (3) (2019) 4831—4843.

[24] P. Schaffer, I. Vajda, in: Cora: correlation-based resilient aggregation in sensor networks, 2007, pp. 373–376.

[25] J. Eldridge, Mutual information, http://samples.jbpub.com/9781284079654/9781284108958_CH12_Pass03.pdf.

[26] L. Lu, X. Zhu, X. Zhang, J. Liu, M. Bhuiyan, G. Cui, One Intrusion Detection Method Based On Uniformed Conditional Dynamic Mutual Information, 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (Trust-Com/BigDataSE), 1236–1241, 2018.

[27] M.Z.A. Bhuiyan, T. Wang, L. Qi, G. Wang, J. Wu, T. Hayajneh, Preserving balance between privacy and data integrity in edge-assisted internet of thing, IEEE Internet Things J. (IEEE IoT-J) (2019) 183–196, doi:10.1109/JIOT.2019.2951687.

[28] C. Miao, W. Jiang, L. Su, Y. Li, S. Guo, Z. Qin, H. Xiao, J. Gao, K. Ren, Cloud-enabled privacy-preserving truth discovery in crowd sensing systems, in: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, in: SenSys '15, 2015, pp. 183–196.

[29] E. Bertino, C. Dai, H.-S. Lim, D. Lin, High-assurance integrity techniques for databases, in: A. Gray, K. Jeffery, J. Shao (Eds.), Sharing Data, Information and Knowledge, 2008, pp. 244–256.

[30] C. Dai, D. Lin, M. Kantarcioglu, E. Bertino, E. Celikel, B. Thuraisingham, Query processing techniques for compliance with data confidence policies, in: W. Jonker, M. Petković (Eds.), Secure Data Management, 2009, pp. 49–67.

[31] A. Banerjee, S.K.S. Gupta, K. Venkatasubramanian, PEES: Physiology-based End–to-end Security for mHealth, Proceedings of the 4th Conference on Wireless Health, WH'13, 2013, p. 2:1-2:8. ISBN 978-1-4503-2290-4.

[32] S. George, D. Nikos, S. Rosario, L. Valeria, F. Giancarlo, A. Yiannis, Decentralized Time-Synchronized Channel Swapping for Ad Hoc Wireless Networks, IEEE Trans. Vehicular Technology 65 (10) (2016) 8538–8553 In this issue, doi:10.1109/TVT.2015.2509861.

[33] F. Giancarlo, et al., BodyCloud: A SaaS approach for community Body Sensor Networks, Future Generation Computer Systems 35 (2014) 62–79 In this issue, doi:10.1016/j.future.2013.12.015.

[34] H. Mohammad Mehedi, A. Md. Golam Rabiul, U. Md. Zia, H. Md. Shamsul, A. Ahmad, F. Giancarlo, Human emotion recognition using deep belief network architecture, Information Fusion 51 (2019) 10–18, doi:10.1016/j.inffus.2018.10.009.