Medical Science

AIMS Medical Science, 5 (1): 1–22.
DOI: 10.3934/medsci.2018.1.1
Received: 07 June 2017
Accepted: 08 November 2017
Published: 20 December 2017

http://www.aimspress.com/journal/medicalScience

Review

A survey of state-of-the-art methods for securing medical databases

Andrei V. Kelarev[1,*], Xun Yi[1], Hui Cui[1], Leanne Rylands[2] and Herbert F. Jelinek[3]

[1] School of Science, RMIT University, GPO Box 2476, Melbourne, VIC 3001, Australia
[2] School of Computing, Engineering and Mathematics, Western Sydney University, Locked Bay 1797, Penrith, NSW 2751, Australia
[3] Center for Research in Complex Systems and School of Community Health, Charles Sturt University, Albury, NSW, Australia

*** Correspondence**: E-mail: andrei.kelarev@gmail.com; Tel: +61-3-992-52793; Fax: +61-3-051-9238.

Abstract: This review article presents a survey of recent work devoted to advanced state-of-the-art methods for securing of medical databases. We concentrate on three main directions, which have received attention recently: attribute-based encryption for enabling secure access to confidential medical databases distributed among several data centers; homomorphic encryption for providing answers to confidential queries in a secure manner; and privacy-preserving data mining used to analyze data stored in medical databases for verifying hypotheses and discovering trends. Only the most recent and significant work has been included.

Keywords: medical databases; privacy and security; attribute-based encryption; homomorphic encryption; privacy preserving data mining

1. Introduction

The use of advanced cryptographic methods for protecting the privacy of medical records has been actively investigated for some years, and is still one of the most rapidly developing and changing research domains. To maintain data security and patient confidentiality it is important for clinicians to be aware of state-of-the-art techniques in this field, as they may have significant implications for making decisions about what data and types of data are to be included in patient records, their storage and distribution. Protection of patient privacy is paramount, for both ethical and legal reasons. Personal Health Records (PHRs) and electronic medical databases form an integral

part of e-health systems, and so should be protected. In some cases PHRs and information can be maintained by patients via the Internet, creating further privacy issues. For preliminaries and background information on this broad area we refer readers to [1] and [2].

Cloud computing and storage has attracted attention worldwide in the medical field as an efficient system for storing and accessing data. Many countries including the USA have been developing electronic health management systems to integrate data collected by hospitals for creating more efficient healthcare services. Cloud computing and storage can provide the basis for the integration of such data. A Cloud Service Provider (CSP) provides computing infrastructure, data storage and computing power, which allows access when required from anywhere. The computing infrastructure can be located in one, several or many physical locations. Cloud computing and storage has brought about an improvement in the provision of healthcare in terms of information management, providing low-cost and effective outsourced data storage but with additional privacy requirements.

PHRs are a patient-centric model of health information exchange, with storage of records often outsourced to third parties such as CSPs. Efficient storage, access control and sharing of data in the cloud are highly desirable for modern clinical practitioners; it is easy to access and provides many advantages over paper documents and client-server records. However, there have been widespread security and privacy concerns about entrusting personal health information to third parties as this could increase the risk of unauthorized access to private information and highly sensitive data. Privacy preserving methods are essential for the storage and exchange of private health information.

Cryptography is an essential tool that helps to ensure medical data accuracy and confidentiality whilst retaining data security. Cryptography is the study of disguising messages or information, the *plaintext*, so that only those for whom the messages or information are intended can read them. The process of disguising the plaintext is *encryption*, and the disguised message is the *ciphertext*. *Decryption* is the process of extracting the original message, the plaintext, from the ciphertext. A *key* (think of a very big number) is used together with an encryption algorithm to create the ciphertext. Decryption requires a key and a decryption algorithm; this algorithm and key are used to recover the plaintext from the ciphertext. The processes or algorithms used to encrypt the plaintext and decrypt the ciphertext are not secret; it is the keys that provide security. If the encryption and decryption keys are the same, the key is a symmetric key. If the encryption key is public, and the decryption key is different and private, then anyone can encipher information (for example, their health information) using the public key and send the ciphertext to the owner of the private key (for example, a central health authority) who can decipher all information received using the private key. Such a cryptosystem is a *public key* cryptosystem.

Encryption is a crucial aspect of the preservation of privacy of medical records and PHRs, especially with the increasing integration of data sets, electronic sharing of data and more complicated access requirements. It is not difficult to eavesdrop on messages as they are transmitted or for a dishonest insider in a data center to copy files. If such messages and files are encrypted then the files are unintelligible and of no use to those who have illegally accessed them. Privacy of records is preserved. However, problems of efficiency and scalability with the granting and revoking of access to data remain very important challenges for achieving fine-grained, cryptographically enforced data access control.

Medical wireless sensor networks (WSNs) are another technology used for data transfer and these have been widely used in healthcare applications such as in hospitals and for home patient monitoring systems including, for example, blood pressure monitoring [3]. WSNs are more

vulnerable to eavesdropping, modification, impersonation and replaying attacks than wired networks.

Advanced cryptographic techniques play vital roles in maintaining the privacy of data in a cloud environment and in securing the operation of medical WSNs. Such techniques can facilitate the use of large distributed medical databases for medical purposes, for example, reliable verification of the role of lifestyle and social factors in disease prevention and improved well-being [4–6], the effectiveness of treatment options [7], the investigation of rare genetic diseases [8], as well as for finding attributes that may help in the early diagnosis of various conditions [9–17]. Likewise, medial databases are valuable in the development of automated computer-based diagnostic classification systems for e-health and mobile applications [18–20]. The preservation of privacy could also facilitate the creation of large medical databases combining data available in several countries. This could strengthen research on the automated medical diagnosis of various diseases, including cardiac autonomic neuropathy, which has been investigated recently in [21–29] using only one local database.

The present review article is devoted to recent developments in the area of medical applications of cryptographic techniques. We concentrate on three advanced directions, where active research has been carried out recently. These three directions have significant implications for security of medical data. In the next section, we begin with background and recent results devoted to k-anonymity. Section 3 deals with attribute-based encryption. Homomorphic encryption is presented in Section 4. Finally, a review of recent work on privacy-preserving data mining is given in Section 5. For the fundamentals of computer security and mathematical foundations of classical cryptographic primitives, the readers can turn, for example, to [30–33]. For preliminaries on private information retrieval, we refer to [34].

2. K-anonymity

For medical research it may be essential to use some parts of the medical records of arbitrary selections of patients without requesting the patients to sign legal consent which might otherwise introduce a bias to the study and reduce participation. When only a part of the record is released, to comply with the privacy legislation it is essential to guarantee that the identity of the person cannot be established. To achieve this it is necessary to ensure an important property known as k-anonymity. It is essential for the protection of data being released against the possibility of re-identification of the patients to which the released attributes refer.

To define the notion of k-anonymity, suppose that a database D is to be released and made available for medical research. We assume that the identifiers of patients, such as their name, Medicare number or private health insurance membership number have been deleted. The database D may contain *quasi-identifiers*, i.e., attributes which can be exploited for linking the records to the patients. This means that an attacker can use quasi-identifiers to identify the patients by using other sources of publicly available information. Examples of quasi-identifiers include the date of birth, age, sex, type of employment, postal code of the area, the medical practice name, and so on. If there is a combination of quasi-identifiers, which is unique, then an attacker can determine the identity of the respondent to which the record with this combination of quasi-identifiers refers. The notion of k-anonymity requires that the quasi-identifiers of every record in D be related to no fewer than k patients. More specifically, the database D is said to satisfy the k-anonymity condition if, for every combination of values of quasi-identifying attributes occurring in D, there exist k records in D with exactly the same combination of the quasi-identifiers.

2.1. Previous work on k-anonymity

The main methods for enforcing k-anonymity in medical databases are presented in Table 1.

Table 1. Major methods for enforcing k-anonymity**.**

Method	Summary
Generalization	Replaces values of attributes with generalized version of these attributes. Different values can be generalized to a same value so that the number of occurrences of the new value will increase.
Suppression	Suppresses sensitive information by removing it. Suppression can be applied by deleting a single attribute in one record, or a single attribute in all records, or an entire record. Usually suppression is applied to remove outliers or records without similar records. Suppression makes it possible to reduce the amount of generalization required for achieving k-anonymity.

Suppression and generalization were applied for achieving k-anonymity before sharing medical information in [35]. An extended version of k-anonymity, called a modified entropy 1-diversity model, is introduced and investigated. Several types of linking attacks are considered and it is shown that they can be prevented using this model. It is explained how to use suppression and generalization to achieve the modified 1-diversity.

A new globally optimal de-identification algorithm Optimal Lattice Anonymization (OLA) was developed in [36]. It is designed for use with medical data in order to achieve k-anonymity. A thorough experimental study was carried out to compare OLA with previous techniques proposed by other authors. The results established that OLA was significantly faster than other methods in finding the globally optimal de-identification solution.

Anonymization techniques for PHRs keeping patient data safe while preserving useful medical information were studied in [37]. In order to design an algorithm for privacy protection that reduces the overall information distortion and leads to consistent information loss, the paper [37] introduced a k-member cluster seed selection algorithm (KMCSSA) and applies it in k-member clustering to achieve k-anonymity. The k-member clustering with KMCSSA aims to collect records minimizing the amount of generalization required to achieve k-anonymity. It is well known that many classical clustering techniques involve choosing a random seed, which results in inconsistencies in performance. To overcome this problem, KMCSSA selects the seed based on the closeness of records in order to reduce the information distortion and to produce consistent information loss. Experiments presented in the paper show that KMCSSA is superior to previous algorithms as far as information loss is concerned.

An application of k-anonymity to preserve privacy in wireless sensor network medical environments was considered in [38]. The paper proposes a clustering-based architecture for effective data aggregation and achieving k-anonymity. The architecture is resource aware. It minimizes energy consumption. Experimental results evaluating the energy consumption and network performance of the system are presented.

A k-anonymity study of medical care data is undertaken in [39]. Experimental results discussed in this article have highlighted high risk of disclosure of the identities of the patients in medical

records and the need to develop a robust technique for k-anonymization.

In [40], the authors develop a novel method for achieving k-anonymity while preserving the data distribution by applying dithered quantization and Rosenblatt's transformation. The quality of preserving data distribution is workload driven. This method is then used for real-life publicly available medical datasets in order to solve the medical insurance cost minimization problems.

Distributed randomization is combined with k-anonymity in [41] for the privacy protection of medical data. A clinical dataset of diabetic nephropathy is used to assess novel method for the reduction of the information loss rate proposed in the paper.

An extensive collection of algorithms for k-anonymization is studied in [42]. The results of comprehensive experiments are presented and the best three algorithms are chosen based on their execution time and performance in the degree k of anonymization for various choices of quasi-identifiers.

A scalable k-anonymization approach using MapReduce is proposed in [43]. In the case of very large medical databases, the amount of generalization and suppression required to achieve the same level of k-anonymity reduces considerably because of the well-known large crowd effect. However, the problem of handing big data for anonymization remains challenging. MapReduce can be used to handle large volumes of data. To apply it effectively, it is essential to design scalable algorithms. The paper [43] introduces an algorithm called scalable k-anonymization (SKA) using MapReduce for privacy preserving big data publishing. Experiments comparing it with previous solutions demonstrate that the new algorithm leads to a remarkable improvement in quality of the outcomes and in running time.

The paper [44] proposes a semantic-based k-anonymity scheme for health record linkage. It is applied for linking the original records from sources in situations where direct access to the data is not possible. The semantic-based linkage k-anonymity is proposed for de-identifying record linkage with fewer generalizations at the same time eliminating inference disclosure by means of semantic reasoning.

2.2. Challenges facing k-anonymity

Legislative requirements to obtain consent may be waived if the disclosed data are de-identified. Achieving k-anonymity means that the probability of identifying the patient is reduced to 1/k. Careful legal studies are needed to determine the appropriate value of k that should be used to avoid legal challenges for compliance with legislation during data release in each particular situation.

The concept of k-anonymity is absolutely necessary for the preservation of privacy of the shared data. Therefore the crucial challenge of minimizing the data loss will remain central in future work devoted to k-anonymity for medical databases.

However, k-anonymity is not a sufficient condition, which means that is it does not guarantee that the privacy of patient data is protected. For example, suppose that a sensitive attribute S in a medical database D must be protected and that there is a group of k records in D with the same selection of quasi-identifiers and such that the values of S in all of these k records are similar, or perhaps they may even be equal. In this case, if an attacker can reveal the identity of a patient from these quasi-identifiers, then the attacker will be able to identify all k individuals in this group, and so the attacker will know the value of the confidential attribute S for all k patients. Furthermore, suppose that the HIV status (or the diagnosis of any other disease) with values 0 and 1 is the sensitive attribute S. In this case, since the attribute S can take on only two values, it follows that if the attacker looks at many subsets of k patients with coinciding quasi-identifiers each, then the

probability that the attacker can find k individuals having the same diagnosis and also the same quasi-identifiers rapidly approaches 1 with the growth of size of the medical database. Therefore with a probability approaching 1 the attacker can find and identify a group of k patients with the same HIV status, which is a security risk.

2.3. Future directions for work on k-anonymity

The example given in the preceding section shows that further research is needed to develop more advanced generalized versions of k-anonymity and investigate medical databases satisfying these conditions.

Another promising direction for future work is to investigate ways of combining k-anonymity with cryptographic techniques for the development of hybrid solutions increasing the effectiveness and efficiency of both approaches in preserving the privacy of medical data.

These challenges demonstrate that further research on k-anonymity and its generalizations is required. Currently available stronger privacy protection methods based on cryptography discussed in the following sections can be recommended to the medical practitioners.

3. Attribute-based encryption

For stronger and more reliable protection of data privacy, it is desirable to store all data items in encrypted form, and allow only authorized users to access the original values. Attribute-based encryption (ABE) [45] is a form of public key encryption. It identifies users by attributes such as medical condition, gender, age, hospital, and position. An ABE system enables scalable access control by specifying access polices (or access structures) over encrypted data. The access policies are used to determine what kind of users can decrypt and access a ciphertext stored in the medical database. Following [46], the main types of ABE are presented in Table 2.

Table 2. Main types of abe.

Acronym	Method
CP-ABE	Ciphertext-Policy ABE
KP-ABE	Key-Policy ABE

In CP-ABE, each user is issued with a private key which encapsulates his/her attributes, and each message is encrypted under a specified access policy in terms of attributes. Thus the users can decrypt a ciphertext if and only if their attributes, which are associated with their private keys, satisfy the access policy of the ciphertext. In KP-ABE the situation is reversed. Private user keys are associated with their access policies and the ciphertexts are associated with the attributes. This means that a user can decrypt a ciphertext if and only if the corresponding access policy (associated with their private key) is satisfied by the attributes of the ciphertext.

3.1. Previous work on ABE

This use of attributes makes ABE a successful security application for providing flexible and

fine-grained access control over medical data. It also makes ABE a promising solution for the protection of data privacy in scenarios requiring scalable access control, for example, medical databases. For technical background on ABE the readers are referred to the original articles [45–48].

An important aspect of practical encryption is a trusted key escrow system, that is, a system for storage of decryption keys, in which the keys can be provided to authorized participants, and to authorized third parties in special circumstances. Though ABE has many advantages in terms of access control, it has limitations. Further research is needed to improve aspects of key management including key escrow [49], revocation, efficiency and privacy protection. To solve these issues, various types of ABE schemes have been proposed. Regarding key escrow, where a single trusted party monitors attributes and issues private keys for all users, the notion of multi-authority ABE [50,51] has been introduced in which multiple attribute authorities operate simultaneously, each distributing private key components corresponding to different sets of attributes.

In terms of revocation, there are solutions such as revocable and decentralized ABE [49], which split the task of decryption key generation across multiple attribute authorities, without requiring any central party, so that attribute revocation is achieved by simply stopping the updating of the corresponding private key. Server-aided revocable ABE [52] optimizes user revocation in ABE by delegating almost all of the workload incurred by user revocation to an untrusted server. ABE with granular revocation [53] utilizes the key separation technique to support selective revocation in which a user's attributes can be selectively revoked.

Many ABE schemes with outsourced decryption have been proposed [52,54,55]. Outsourcing increases efficiency as it decreases the computations to be performed by a user when decrypting a ciphertext. With regard to privacy protection, there are ABE schemes with partially hidden access structures [56–58] or fully hidden access structures [59], which hide the sensitive attribute information of users from the access policies included in the ciphertexts.

A novel patient-centric framework and a suite of mechanisms for data access control of PHRs stored in semi-trusted servers are proposed in [60]. To achieve fine-grained and scalable data access control for PHRs, an ABE is used in [60] to encrypt each patient's PHR. The encryption protocol focuses on handling information with multiple data owners. It divides the owners in the PHR system into multiple security domains, which greatly reduces the key management complexity for owners and users. A high degree of patient privacy is guaranteed simultaneously by exploiting multi-authority ABE. The scheme proposed in [60] enables dynamic modification of access policies or file attributes, supports efficient on-demand user attribute revocation and incorporates special procedures for access in case of emergency.

A privacy-preserving PHR scheme is set out in [61] to support fine-grained access control and efficient revocation. It achieves efficient on-demand user/attribute revocation, dynamic policy update, and scalable and fine-grained access control for PHRs by using a multi-authority ABE scheme. The PHRs of patients are encrypted and stored in semi-trusted servers, where access is allowed by multiple data owners. It is demonstrated in [61] where the security of the scheme reduces to the standard decisional bilinear Diffie-Hellman assumption.

Extending patient data transfer from within health care providers with possible access by patients from fixed sources, wireless body area networks (WBANs) have emerged as a new method for e-healthcare. WBANs can greatly improve healthcare quality, because clinicians can give guidance to patients in real-time without frequent face-to-face meetings and measurements. WBANs are designed for storage and processing of personal information and medical data collected by the

body area sensors. Security and privacy are two vital issues for the successful operation of WBANs. A KP-ABE scheme for fine-grained access control in WBANs is designed in [62]. User access can be revoked if necessary. This protects the security and privacy of patients, providing confidentiality and resistance to collusion attack.

A distributed ABE framework is proposed in [63] for sharing records stored in the cloud with other hospitals. The framework is designed to achieve secure and scalable storage of records in the cloud environment. Lambay [64] uses KP-ABE and multi-authority ABE combined with symmetric key cryptography to achieve fine-grained data access control for encrypted PHRs stored on cloud servers. A novel framework for secure sharing of PHRs stored in the cloud is proposed in [65] to fully realize the patient-centric concept. Patients have complete control of their own privacy as their PHRs are encrypted using several encryption algorithms. Another PHR management system is studied in [66]; it guarantees a high degree of patient privacy by exploiting multi-authority ABE. In a multi-authority ABE all users are divided into multiple security domains that greatly reduce the key management complexity for owners and users.

A scalable, flexible and fine-grained access control mechanism for outsourcing health records to cloud storage is provided in [67]. It categorizes users as personal or professional. For the professional domain, hierarchical attribute set based encryption is employed and for the personal domain a KP-ABE [46] is used. This achieves the needed scalability and fine-grained access for the professional domain and simplicity of key management in the personal domain. The sensitivity of outsourced cloud data is upheld in [68] by employing ABE to restrict admission to encrypted data. Data privacy is safeguarded by established key management and indexing methodologies.

A model for a cloud-based PHR system for health information exchange is studied in [69]. It allows PHR owners to securely store their health data with semi-trusted cloud service providers, and to selectively share their health data with a wide range of PHR users. To reduce the key management complexity, the PHR users are divided into public and personal security domains [69]. PHR owners encrypt their health data for the public domain using CP-ABE, while encrypting their health data for the personal domain using an anonymous multi-receiver identity-based encryption scheme. Only authorized users whose credentials satisfy the specified ciphertext-policy or whose identities belong to dedicated identities can decrypt the encrypted health data.

Attribute based broadcast encryption is proposed in [70] to handle workflow access control scenarios by combining ABE with data access rights based on user identity. It achieves fine-grained and scalable data access control for PHR data in a patient-centric framework for the multiple data owner setting. The attribute based broadcast encryption scheme enables dynamic modification of access policies or file attributes and efficient on demand user attribute revocation. A system for scalable and secure sharing of data in the cloud using ABE is discussed in [71]. PHR data is stored in encrypted form and the system allows authorized persons to access details via a health social network and to generate reports in a secure and authenticated manner.

An innovative architecture for collecting and accessing large amounts of data generated by medical sensor networks is proposed in [72]. The architecture makes it easy for clinicians to share medical information in normal and emergency situations. An effective and flexible security mechanism is presented to guarantee confidentiality, integrity, and fine-grained access control to outsourced medical data. This mechanism relies on CP-ABE to achieve high flexibility and performance.

Eom et al. [73] studied a patient-controlled ABE scheme, which enables patients as data owners to control access to their health data while reducing the operational burden for patients. Patients have

control authority over their own health data; they have the final say on the access and its time limitations. The scheme also provides for medical emergencies with the ability to access health data without patient permission only in an emergency.

An extension of the patient-controlled ABE scheme was designed by [74] to determine the credentials to be used to give access to encrypted sensitive PHRs for users including specialist doctors, physicians, family members, and clinic administrative staff.

Cloud servers can be divided into public and private clouds for storage of information according to the sensitivity of data, as suggested by [75]. ABE is employed for fast data retrieval from multiple servers and one-time passwords are generated and emailed to authorized users for security based data access via Smartphones. Utilizing this method, the level of security is increased and trustworthiness is also maintained. Another efficient specialized ABE technique for pharmaceutical databases was introduced in [76].

A framework using ABE to assist CSPs to securely store and share patient data, and for addressing healthcare regulatory requirements, is explored in [77]. The ABE mechanism studied in [77] guarantees authentication, data confidentiality, availability, and integrity in a multi-level hierarchical order. This allows the healthcare provider to easily add/delete any access rule in any order, which is particularly beneficial for medical practices.

A new method of secure fine-grained access control to PHRs is studied in [78]. It is based on ABE primitives and division of the PHR data into privacy levels. KP-ABE provides fine-grained access control storage system for outsourced sensitive data. It can also provide efficient user revocation by using a timestamp in the private key. The construction achieves data confidentiality, preventing unauthorized users from gaining access to the outsourced data.

3.2. Challenges facing ABE

An attribute-based medical system enables fine-grained access control over encrypted PHRs, which provides data privacy such that only privileged data users are able to access the original PHRs, scalability such that the size of each encrypted PHR is independent of the number of targeted data users, and fine-grained access control such that data owners can specify who are able to access the encrypted PHRs in an expressive manner. In addition to these properties born with ABE, an attribute-based medical system should also address the following issues before it can be widely applied in practice.

- Dynamic user management to enable unbounded number of data users in the system and revocation of data users;
- Data owner anonymity and traceability such that it preserves data owner anonymity in normal circumstances while keeping his/her identity traceable by a trusted authority in case that the data owner misbehaves;
- Secure data provenance to provide irrefutable evidence on who creates and modifies the PHRs in the cloud-based medical system.

All these problems have been separately addressed in ABE schemes, and it might be possible to simply apply these solutions to an attribute-based medical system to solve the related problems in an attribute-based medical system. Anyway, to the best of our knowledge, there has not been an attribute-based medical system that considers all the issues mentioned above.

3.3. Future directions of ABE

Recently, the concept of Internet of Things (IoT) has become increasingly popular. We think this development raises new challenges for the current medical systems based on ABE.

Firstly, many IoT devices are resourced-constrained, but existing ABE schemes are very expensive in calculations, and thus it is not suitable for them to access encrypted data generated by an ABE scheme. It seems that ABE with outsourced decryption can solve this computation problem, but it is not clear that in a medical system, who should play the role of the third party (i.e., a proxy or a cloud server) to perform the calculation for the user. In addition, as the third party should not be trusted at all in the cloud storage scenario, it is important to guarantee that any false calculation conducted by the proxy can be easily detected by the user who sends the outsourcing request and all the information held by the proxy should not leak any information about the real data.

Secondly, current medical systems based on ABE only focus on how to specify access policy over recipients, i.e., what kind of users can decrypt the ciphertext. However, in practice, it is also important to clarify that data owners can share the data with what kind of data users. In other words, the access control should also be performed over data owners such that the administrator of the medical system can be convinced that the data owner does not share any data with those who are outsiders of the whole medical system.

4. Homomorphic encryption

Existing encryption systems can protect patient data during transmission and when stored. Having data encrypted protects data from some insider attacks, for example, where an administrator of a medical database or computer system decides to reveal sensitive or private medical data. This is because encrypted data is unintelligible and so unusable to those who do not have decryption keys. However, medical data can be invaluable for research and therefore require different types of encryption systems to allow, for example, statistical analysis of data in a medical database. For this, as with other analyses, computations need to be performed on the data. One way to do this is to have the data decrypted, transmit all relevant data to one place if it is stored in various locations, and then the analyses can be performed and answers obtained. This risks insider attacks from those with access to the computers, and eavesdropping if data transfer is required. These security risks are removed if calculations can be carried out without decrypting the data. Then analysis can be completed by any party, whether trusted or not. Homomorphic encryption enables this. Homomorphic encryption allows computations to be performed on encrypted data; decryption is not needed. The computation generates an encrypted result which, when decrypted, produces the same result as if the calculation had been performed on the unencrypted data.

Table 2. Main categories of homomorphic encryption.

Acronym	Method
FHE	Fully Homomorphic Encryption
PHE	Partially Homomorphic Encryption

Partial homomorphic encryption allows one operation to be performed on encrypted data. Such encryption might allow addition (we get subtraction for free) of encrypted data, but not multiplication. Fully homomorphic encryption (FHE) allows both addition (and subtraction) and multiplication (and division). The Paillier [79] and ElGamal [80] cryptosystems are partially homomorphic and so can be employed to perform some analyses of patient data without compromising patient privacy [81].

For preliminaries on homomorphic encryption, examples, detailed technical information, and historical overview the readers are referred to the monograph [82]. The most widely used homomorphic encryption methods are the ElGamal [80] and Paillier [79] cryptosystems. Open-source FHE libraries, the HElib library [83] and the FHEW library [84], are available for practitioners.

4.1. Previous work on homomorphic encryption

FHE can eliminate privacy concerns in computations involving confidential medical data. However, its current implementations are very slow. This is an active area of research and there is hope for improvement. One option is the use of branching programs to achieve a dramatic increase in processing speed, though these methods lead to restrictions on the types of data elements that can be used in FHE computations [85]. Another technique for increasing the efficiency, and parallelism, of certain algorithms under FHE is proposed in [86]. Simulations show that parallelization increases the processing speed by a factor of about 20. This is a significant step towards practical FHE-based medical remote monitoring.

Personal health monitoring tools, such as commercially available wireless ECG patches, can significantly reduce healthcare costs by allowing patient monitoring outside healthcare institutions. These tools transmit the acquired medical data to the cloud, and so provide an invaluable diagnostic tool for healthcare professionals. Despite the potential of such systems to revolutionize the medical field, the adoption of medical cloud computing in general has been slow due to the strict privacy regulations on patient health information. FHE is used in [87–89] to protect patient privacy during monitoring of their conditions via wearable devices connected to the cloud. The paper [87] describes homomorphic encryption to develop a system for secure assessment of cloud based health monitoring. A feasibility study of an application of FHE to long-term patient monitoring via cloud based ECG data acquisition through existing ECG acquisition devices was conducted in [88]. FHE is applied to perform secure data analysis and may open up this technology to health care providers.

Several PHR access control protocols based on the ElGamal threshold public key encryption scheme are proposed in [90] under the multi-party framework where all PHRs are encrypted with a common public key, and an encrypted PHR can be decrypted only with the cooperation of all parties. In these protocols, multiple parties cooperate to control clinicians' access to PHRs without actually knowing the content of the PHRs. This ensures protection of the patient data from insider attacks as long as at least one party can be trusted. The protocols are built on public key infrastructure, which facilitates clinician registration and revocation.

The use of WSNs to connect wearable monitoring devices in health care is growing rapidly. Numerous applications are ready to use, such as blood pressure monitors and heart rate monitors that have Bluetooth or WiFi capability. Therefore it is important for system designers to consider how to protect patient privacy in WSNs. A data division scheme for WSNs is proposed in [91]. It utilizes homomorphic encryption to achieve stronger protection. In the proposed scheme, even if a

transmitting node in a WSN is compromised, data privacy is preserved. Experimental results show that the scheme provides a good trade-off between resources consumed and system security, and it is efficient for encryption and decryption.

A novel medical cloud computing approach proposed in [92] eliminates privacy concerns associated with the cloud provider. It uses FHE for computations on private health data in encrypted form without observing the underlying data. The article [92] presents a feasibility study with a working implementation of a long-term cardiac health monitoring application using a well-established open source FHE library.

The paper [93] describes an online health analysis system in the cloud that monitors a patient's health regardless of the geographical location of the patient or clinician. The proposed application reduces patients' traveling time and also reduces time spent in taking and delivering medical reports. In the application, PHR data are encrypted using a strong security algorithm and stored in the cloud by applying ElGamal encryption, using its homomorphic property to secure computations.

Mobile medical queries can also pose a threat to patient location privacy because the location of a query may reveal sensitive information about the patient. An efficient solution using the Paillier public key cryptosystem for preserving location privacy for particular queries is given in [94]. An improvement on methods to preserve location anonymity of patients querying a service, while also allowing the owner of the data to maintain control over their data, was developed in [95]. The article [96] allows a patient to retrieve one type of point of interest, for example, hospitals, without revealing to the location-based services provider what type of point of interest is retrieved. The generic solution is built on the Paillier public key cryptosystem and handles multiple discrete attributes of private location-based queries.

Paillier and ElGamal cryptosystems are used in [97] to encrypt patient data including temperature, heartbeat, and blood pressure. The data are sensed using appropriate wearable sensors, encrypted, and stored in the cloud. Patient data are securely distributed by employing the Paillier and ElGamal cryptosystems. To this end, a mobile application based on homomorphic encryption was developed by [98] to protect patient privacy while carrying out analysis of health data in the cloud.

Large amounts of data are invaluable for learning about and understanding many health issues, leading to improved health care. Complex data mining algorithms are needed to maximize the use of available data, however, the privacy concerns of people, society and organizations put at risk the use of such valuable data. Therefore solutions such as those addressed in this paper to secure data and preserve privacy deserve consideration.

4.2. Challenges facing homomorphic encryption

Further work on increasing the speed of FHE and PHE algorithms is the most important challenge in order to make homomorphic techniques suitable for applications in large medical databases.

4.3. Future directions for work on homomorphic encryption

The development of effective FHE and PHE algorithms with split keys is a valuable direction, because split keys are a convenient way of incorporating secret sharing in a cryptosystem.

5. Privacy preserving data mining

Data mining is the process of sorting through very large data sets to identify patterns, establish relationships and gain knowledge. The main tasks of data mining are association rule mining, classification and clustering. Association rule mining looks for associations between items of the form "if X then Y", meaning that if X is present then it is likely that Y is also. For example, it might be found that a large proportion of people with disease X also have characteristic Y. Classification aims to assign items to categories. For example, based on various characteristics it might be possible to assign people as low, medium or high risk for disease X. Clustering is about grouping like items together. Data mining tools allow enterprises to predict future trends. Privacy preserving data mining facilitates the use of large datasets for research, and is important for clinical practice as it can detect trends in confidential data and provide valuable recommendations to health practitioners, while preserving privacy.

5.1. Previous work on privacy preserving data mining

Privacy preserving data mining research has been divided into two areas presented in Table 3.

Table 3. Main categories of privacy preserving data mining areas.

Acronym	Method
Data perturbation approach	It alters the data before applying the data mining algorithm so that real values are obscured, but important statistics remain preserved.
Privacy preserving distributed data mining	Privacy preserving distributed data mining use sensitive data from distributed databases held by different parties, e.g., hospitals.

The first is to alter the data before delivery to the data miner so that real values are obscured. If a random number chosen from a Gaussian distribution is added to each data value, the data miner no longer knows the exact data values. However, important statistics such as average and standard deviation remain preserved. Research has addressed related statistical issues [99]. Data mining techniques on such altered data have been developed for constructing decision trees, which can be used for classification [100,101] and for association rules [102,103]. This data perturbation approach works in the "data warehouse" model of data mining, but trades privacy for accuracy of results.

The second approach is privacy preserving distributed data mining. Distributed data mining applications, such as those dealing with health care, use sensitive data from distributed databases held by different parties. Consider the case where several hospitals wish to mine their patient data jointly for the purpose of medical research. In many countries privacy policies and legal requirements do not allow hospitals to pool their data or to reveal data to each other. Although hospitals may be allowed to release data after identifiers, such as name and address, have been removed, this does not guarantee privacy as sometimes re-identification is possible by linking different public databases to relocate the original subjects [104]. In order to conduct research and allay privacy concerns, protocols are needed for privacy-preserving distributed data mining.

A cryptographic approach was firstly used to build a decision tree [105] in the case where the

data is held by two parties, each with their own database. The databases are not shared and there is no third trusted party involved. Computations are done by both parties and though information is shared (without revealing to each party anything about the data held by the other party), a decision tree on the combined data is constructed. A well-known algorithm (ID3) is used ensuring efficient communication between the two parties.

Data can be separated with different partitions kept on different servers. This could reduce the load on each server. Horizontal partitioning is one way of splitting data such that, for example, on one server are all records for patients under 25, on another for those over 25 and under 50, and a third for those 50 and over. In the case where three or more parties jointly mine data to obtain global association rules on horizontally partitioned data, a protocol to preserve confidentiality was proposed in [106]. In this application the parties learn (almost) nothing beyond the global results. The protocol involves the parties sharing data and encrypting it as it is passed along (data gets encrypted many times). Based on other information that is gathered during this process, association rules are discovered. A particular type of encryption (commutative encryption) must be used for this protocol to work.

Yi et al. [107] present a protocol which is based on a new semi-trusted mixer model, in which one party “the mixer” (perhaps a government agency) receives messages from several other parties and combines, or mixes, the messages to obtain a result. This privacy-preserving distributed association rule mining protocol uses data from the n sites, and can protect the privacy of each database against collusion by a coalition of up to n-2 parties. The protocol requires many rounds of communication between the parties and the mixer, but only two communications between each party and the mixer in each round of data collection. In the protocol, each party performs distributed association rule mining as it would on its local data.

By extending the single semi-mixer model to the multiple semi-mixer model, a two-party protocol and a multi-party protocol for a privacy-preserving naive Bayes classifier for horizontally partitioned distributed data was discussed in [108]. The multi-party protocol is built on the multiple semi-trusted mixer model in which each data site sends messages to two semi-trusted mixers, which run the two-party protocol and then broadcast the classification result. This model facilitates both trust management and implementation. Security analysis has shown that the two-party protocol is a private protocol and the multi-party protocol is a private protocol as long as the two mixers do not collude.

Yi et al. [109] present an equally contributory multi-party k-means clustering protocol for vertically partitioned data where each party equally contributes to k-means clustering. The protocol is built on ElGamal encryption, Jakobsson and Juels's plaintext equivalence test protocol, and mix networks. It protects privacy by k-means clustering for each iteration without revealing the intermediate values.

Reducing the computational burden of applying privacy preserving protocols is an important problem. A cloud computing environment can provide effective solutions to this problem as the computations are outsourced to the CSPs. However, appropriate methods for alleviating privacy and security concerns are needed. The paradigm of data mining-as-a-service in the cloud computing environment has attracted interest. In this paradigm, a client (data owners), lacking data storage, computational resources and expertise, stores data in the cloud and outsources data mining tasks to the cloud servers. In order to protect the privacy of the outsourced database and the association rules mined, k-anonymity, k-support, and k-privacy techniques have been proposed to perturb the data before it is uploaded to the server. These techniques are computationally expensive and it is often

better to execute association rule mining locally. Recently, more efficient techniques that encrypt data before storage and analysis in the cloud have been actively investigated. A scenario where a clinician encrypts and stores data in the cloud and outsources the task to n "semi-honest" servers was investigated in [110]. To mine association rules from the data, the servers cooperate to perform association rule mining on the encrypted data in the cloud and return encrypted association rules to the user. Three solutions are provided in [110] for protecting data privacy during association rule mining. These solutions are built on the distributed ElGamal cryptosystem and achieve item privacy, transaction privacy and database privacy, as long as at least one of the servers is honest. To eliminate the risk that all servers are compromised, users are advised either to use well established creditable CSPs or to combine servers from several different CSPs.

Rao et al. [111] also propose a novel and efficient protocol for privacy-preserving outsourced distributed clustering for multiple users based on the k-means clustering algorithm. The protocol avoids the secure division operations required in computing cluster centers for k-means clustering through efficient transformation techniques. In addition, offline computation and pipelined execution are studied to boost performance. These two strategies combined with parallelism significantly improve the performance of the protocol.

To make it attractive for medical practitioners to outsource data analytics to service providers with powerful platforms and advanced analytics skills, an effective encryption scheme employing homomorphic encryption to perform k-means clustering directly over the encrypted data was reported in [112]. Since the ciphertexts resulting from homomorphic encryption do not preserve the order of distances between data objects and cluster centers, the proposed approach enables the service provider to compare encrypted distances with the trapdoor information supplied by the data owner.

5.2. Challenges facing privacy preserving data mining

The reduction of information loss is the major challenge facing the data perturbation approach in privacy preserving data mining.

Increasing the effectiveness and reducing the running time of privacy preserving data mining algorithms based on cryptographic techniques is also important.

5.3. Future directions for work on privacy preserving data mining

The investigation of privacy preserving ensemble classifiers is a promising new research direction, since ensembles are well known machine learning tools deployed in solutions to various medical problems and the problem of developing privacy preserving ensembles is quite challenging.

The investigation of privacy preserving fuzzy classifiers is also an interesting direction for further research, because fuzzy techniques have been quite effective in medical applications recently, but they have never been applied in privacy preserving data mining yet.

6. Conclusion

This article gives a review of recent work on three advanced research directions, which have significant implication for practical decisions concerning the security of medical databases: attribute-based encryption for enabling secure access to confidential medical databases distributed

among several data centers; homomorphic encryption for providing answers to confidential queries in a secure manner; and privacy-preserving data mining used to analyze data stored in medical databases for verifying hypotheses and discovering trends.

Acknowledgments

This work was supported by the Australian Research Council, Discovery grant DP160100913. The authors are grateful to the anonymous referees for comments and suggestions of improvements that have helped to revise this review article.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Carter JH (2008) Electronic Health Records: A Guide for Clinicians and Administrators. ACP Press.
2. Anderson R (2012) Personal Medical Information: Security, Engineering, and Ethics. Springer, Cambridge.
3. Villalva CM, López-Alvarez XLM, Rodríguez MM, et al. (2017) Blood pressure monitoring in cardiovascular disease. *AIMS Med Sci* 4: 164–191.
4. Kara B, Tenekeci EG, Demirkaya S (2016) Factors associated with sleep quality in patients with multiple sclerosis. *AIMS Med Sci* 3: 203–212.
5. Dillon C, Taragano FE (2016) Special Issue: Activity and Lifestyle Factors in the Elderly: Their Relationship with Degenerative Diseases and Depression. *AIMS Med Sci* 3: 213–216.
6. Wilson D, Keith G, Harpal B, et al. (2017) Therapy through social medicine: cultivating connections and inspiring solutions for healthy living. *AIMS Med Sci* 4: 131–150.
7. Panchal HB (2016) Percutaneous interventions for peripheral vascular disease. *AIMS Med Sci* 3: 234–236.
8. Amraoui H, Mhamdi F, Elloumi M (2017) Survey of metaheuristics and statistical methods for multifactorial diseases analyses. *AIMS Med Sci* 4: 291–331.
9. Petillo D, Orey S, Tan AC, et al. (2014) Parkinson's disease-related circulating microRNA biomarkers – a validation study. *AIMS Med Sci* 2: 7–14.
10. DeMarshall CA, Sarkar A, Nagele RG (2015) Serum autoantibodies as biomarkers for Parkinson's disease: background and utility. *AIMS Med Sci* 2: 316–327.
11. Ervin K, Pallant J, Terry DR, et al. (2015) A descriptive study of health, lifestyle and sociodemographic characteristics and their relationship to known dementia risk factors in rural Victorian communities. *AIMS Med Sci* 2: 246–260.
12. Shinde S, Mukhopadhyay S, Mohsen G, et al. (2015) Biofluid-based microRNA biomarkers for Parkinson's disease: an overview and update. *AIMS Med Sci* 2: 15–25.
13. White VJ, Nayak RC (2015) Re-circulating phagocytes loaded with CNS debris: a potential marker of neurodegeneration in Parkinsons disease? *AIMS Med Sci* 2: 26–34.

14. Fagere MO (2016) Diagnostic utility of pleural effusion and serum cholesterol, lactic dehydrogenase and protein ratios in the differentiation between transudates and exudates. *AIMS Med Sci* 3: 32–40.
15. Khalid KE, Nsairat HN, Zhang JZ (2016) The presence of interleukin 18 binding protein isoforms in Chinese patients with rheumatoid arthritis. *AIMS Med Sci* 3: 103–113.
16. Kirchengast S (2017) Diabetes and obesity—an evolutionary perspective. *AIMS Med Sci* 4: 28–51.
17. Tanhapour M, Vaisi-Raygani A, Khazaei M, et al. (2017) Cytotoxic T-lymphocyte associated antigen-4 (CTLA-4) polymorphism, cancer, and autoimmune diseases. *AIMS Med Sci* 4: 395–412.
18. Fitzmaurice MJ, Adams K, Eisenberg JM (2002) Three decades of research on computer applications in health care: medical informatics support at the agency for healthcare research and quality. *JAMIA* 9:144–160.
19. Hage I, Hamade R (2015) Automatic detection of cortical bone's Haversian osteonal boundaries. *AIMS Med Sci* 2: 328–346.
20. Zhang Q, Zhou D, Zeng X (2017) Machine learning-empowered biometric methods for biomedicine applications. *AIMS Med Sci* 4: 274–290.
21. Abawajy J, Kelarev A, Chowdhury M (2013) Multistage approach for clustering and classification of ECG data. *Comput Meth Prog Biomed* 112: 720–730.
22. Abawajy J, Kelarev A, Chowdhury M, Jelinek HF, et al. (2013) Predicting cardiac autonomic neuropathy category for diabetic data with missing values. *Comput Biol Med* 43: 1328–1333.
23. Stranieri A, Abawajy J, Kelarev A, et al. (2013) An approach for Ewing test selection to support the clinical assessment of cardiac autonomic neuropathy. *Artif Intell Med* 58: 185–193.
24. Abawajy J, Kelarev A, Chowdhury MU, et al. (2016) Enhancing predictive accuracy of cardiac autonomic neuropathy using blood biochemistry features and iterative multi-tier ensembles. *IEEE J Biomed Health Informatics* 20: 408–415.
25. Chowdhury M, Abawajy J, Kelarev A, et al. (2016) A clustering-based multi-layer distributed ensemble for neurological diagnostics in cloud services. *IEEE Trans Cloud Comp*. DOI10.1109/TCC.2016.2567389.
26. Jelinek HF, Abawajy JH, Kelarev AV, et al. (2014) Decision trees and multi-level ensemble classifiers for neurological diagnostics. *AIMS Med Sci* 1: 1–12.
27. Jelinek HF, Abawajy JH, Cornforth D, et al. (2015) Multi-layer attribute selection and classification algorithm for the diagnosis of cardiac autonomic neuropathy based on HRV attributes. *AIMS Med Sci* 2: 396–409.
28. Jelinek HF, Kelarev AV (2016) A survey of data mining methods for automated diagnosis of cardiac autonomic neuropathy progression. *AIMS Med Sci* 3: 217–233.
29. Jelinek HF, Cornforth DJ, Kelarev AV (2016) Machine learning methods for automated detection of severe diabetic neuropathy. *J. Diab Compl Med* 1: 1–7.
30. Menezes AJ, van Oorschot PC, Vanstone SA (2001) *Handbook of Applied Cryptography* (Discrete Mathematics and Its Applications), Fifth Edition, CRC Press, Taylor & Francis Group, London, New York.
31. Pieprzyk J, Hardjono T, Seberry J (2003) Fundamentals of Computer Security. Springer-Verlag, Berlin.

32. Domingo-Ferrer J (2002) *Inference Control in Statistical Databases*. Sixth edition, Springer, Berlin.
33. Batten LM (2013) *Public Key Cryptography: Applications and Attacks*. Wiley-IEEE Press, New York.
34. Yi X, Paulet R, Bertino E (2013) *Private Information Retrieval*. Morgan and Claypool, United States.
35. Zhu Y, Peng L (2007) Study on K-anonymity Models of Sharing Medical Information. International Conference on Service Systems and Service Management. *IEEE*: 1–8.
36. El Emam K, Dankar FK, Issa R, et al. (2009) A globally optimal k-anonymity method for the de-identification of health data. *J Am Med Inform Association* 16: 670–682.
37. Shin M, Yoo S, Lee KH, et al. (2013) Electronic medical records privacy preservation through k-anonymity clustering method. Joint, International Conference on Soft Computing and Intelligent Systems. *IEEE*: 1119–1124.
38. Belsis P, Pantziou G (2014) A k-anonymity privacy-preserving approach in wireless medical monitoring environments. *Person Ubiquitous Comput* 18: 61–74.
39. Panackal JJ, Pillai AS, Krishnachandran VN (2014) Disclosure risk of individuals: a k-anonymity study on health care data related to Indian population. International Conference on Data Science & Engineering. *IEEE*: 200–205.
40. Wei D, Ramamurthy KN, Varshney KR (2016) Health insurance market risk assessment: Covariate shift and k-anonymity. *SIAM Data Mining*: 226–234.
41. Xie Y, He Q, Zhang D, et al. (2016) Medical ethics privacy protection based on combining distributed randomization with k-anonymity. International Congress on Image and Signal Processing. *IEEE*: 1577–1582.
42. Simi MS, Nayaki KS, Elayidom MS (2017) An extensive study on data anonymization algorithms based on k-anonymity. *IOP Conf Ser Mater Sci Eng* 225: 1–10.
43. Mehta BB, Rao UP (2017) Privacy preserving big data publishing: A scalable k-anonymization approach using MapReduce. *IET Software* 11: 271–276.
44. Lu Y, Sinnott RO, Verspoor K (2017) A semantic-based k-anonymity scheme for health record linkage. *Studies Health Technology Informatics* 239: 84–90.
45. Sahai A, Waters B (2005) Fuzzy identity-based encryption. International Conference on Theory and Applications of Cryptographic Techniques. Springer-Verlag. *Lect Notes Comp Sci* 3494: 457–473.
46. Goyal V, Pandey O, Sahai A, et al. (2006) Attribute-based encryption for fine-grained access control of encrypted data. ACM Conference on Computer and Communications Security. ACM: 89–98.
47. Shamir A (1984) Identity-based cryptosystems and signature schemes. *Lecture Notes Comput Sci* 21: 47–53.
48. Waters B (2011) Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization. *Lecture Notes Comput Sci* 2008: 321–334.
49. Cui H, Deng RH (2016) Revocable and decentralized attribute-based encryption. *Comput J* 59: 1220–1235.
50. Chase M (2007) Multi-authority attribute based encryption. Theory of Cryptography. Springer Berlin Heidelberg, 515–834.

51. Muller S, Katzenbeisser S, Eckert C (2008) Distributed attribute-based encryption. Information Security and Cryptology-Icisc 2008, International Conference, Seoul, Korea, December 3–5, Revised Selected Papers. DBLP: 20–36.
52. Cui H, Deng RH, Li Y, et al. (2016) Server-Aided Revocable Attribute-Based Encryption. *Europ Symp Res Comptu Sec*: 570–587.
53. Cui H, Deng RH, Ding X, et al. (2016) Attribute-based encryption with granular revocation. International Conference on Security and Privacy in Communication Systems. Springer: 165–181
54. Green M, Hohenberger S, Waters B (2011) Outsourcing the decryption of ABE ciphertexts. Proc USENIX Security Symposium, USENIX Association.
55. Lai J, Deng RH, Guan C, et al. (2013) Attribute-based encryption with verifiable outsourced decryption. *IEEE Trans Info Forensics Sec* 8: 1343–1354.
56. Camenisch J, Dubovitskaya M, Enderlein RR, et al. (2012) Oblivious transfer with hidden access control from attribute-based encryption. *Int Conf Security Crypt Networks*: 559–579.
57. Cui H, Deng RH, Wu G, et al. (2016) An efficient and expressive ciphertext-policy attribute-based encryption scheme with partially hidden access structures. International Conference on Provable Security. Springer-Verlag New York: 19–38.
58. Liu L, Lai J, Deng RH, et al. (2016) Ciphertext-policy attribute-based encryption with partially hidden access structure and its application to privacy-preserving electronic medical record system in cloud environment. *Security Comm Networks* 9: 4897–4913.
59. Lewko AB, Okamoto T, Sahai A, et al. (2010). Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. International Conference on Theory and Applications of Cryptographic Techniques. Springer-Verlag: 62–91.
60. Li M, Yu S, Zheng Y, et al. (2013) Scalable and Secure Sharing of Personal Health Records in Cloud Computing using attribute-based encryption. *IEEE Trans Parallel Distrib Syst* 24: 131–143.
61. Qian H, Li J, Zhang Y, et al. (2014) Privacy-preserving personal health record using multi-authority attribute-based encryption with revocation. *Int J Inf Sec* 14: 487–497.
62. Tian Y, Peng Y, Peng X, et al. (2014) An attribute-based encryption scheme with revocation for fine-grained access control in wireless body area networks. *Int J Distrib Sensor Networks*: 1–9.
63. Radhini MP, Prabha PA, Parthasarathi P (2014) Encryption for secure sharing of personal medical records in cloud. *Int J Sci Eng Technol Res (IJSETR)* 3: 1308–1414.
64. Lambay MA, Lakshmi MJ, Gamare PS (2014) Sharing of personal health records securely in cloud computing with attribute based encryption. *Int J Comp Sci Info Tech (IJCSIT)* 5: 6864–6866.
65. Gondkar DA, Kadam VS (2014) Attribute based encryption for securing personal health record on cloud. *Int Conf Devices Circuits Systems (ICDCS)*: 1–5.
66. Alias AE, Roy N (2014) Improved security of attribute based encryption for securing sharing of personal health records. *Int J Adv Comp Technol* 3: 1224–1227.
67. Mohanan L, Varghese AB (2015) Flexible, scalable and fine grained access control for medical data in cloud using attribute based encryption. *Int J Appl Eng Res* 10: 43378–43383.
68. Bhuvaneshwari M, Sasikumar S (2015) Secure and isolated personal health records using cipher text policy attribute based encryption. *Int J App Eng Res* 10: 23022–23026.

69. Wang C, Xu X, Shi D, et al. (2015) Privacy-preserving cloud-based personal health record system using attribute-based encryption and anonymous multi-receiver identity-based encryption. *Informatica* 39: 375–382.
70. Raseena M, Harikrishnan GR (2014) Secure sharing of personal health records in cloud computing using attribute-based broadcast encryption. *Int J Comp App* 102: 13–19.
71. Shubhangi G, Priyanka J, Pranjali K, et al. (2015) Scalable and secure sharing of data in cloud computing using attribute based encryption. *Int J Multidisc Res Develop* 2: 416–420.
72. Lounis A, Hadjidj A, Bouabdallah A, et al. (2016) Healing on the cloud: Secure cloud architecture for medical wireless sensor networks. *Future Gen Computer System* 55: 266–277.
73. Eom J, Lee DH, Lee K (2016) Patient-controlled attribute-based encryption for secure electronic health records system. *J Med Syst* 40. Article number 253.
74. Saxena AR, Swarnalatha P (2016) Attribute based encryption and decryption of medical records. *Int J Pharmacy Technology* 8: 22192–22199.
75. Reddy MR, Anusha N, Shankar BNV (2016) Secured health records storage & retrieval system using keyword based key generation and Attribute Based Encryption (ABE). *Res J Pharm Bio Chem Sci* 7: 1420–1426.
76. Saravanan T (2016) Energy efficient attribute based encryption technique for health records via virtual machines in the cloud. *J. Chem. Pharmaceutical Sci* 9: 1654–1657.
77. Elmogazy H, Bamasag O (2016) Securing healthcare records in the cloud using attribute-based encryption. *Comp Info Sci* 9: 60–67.
78. Yan H, Li J, Li X, et al. (2016) Secure access control of e-health system with attribute-based encryption. *Intell Automation Soft Comput* 22: 345–352.
79. Paillier P (1999) Public-key cryptosystems based on composite degree residuosity classes. International Conference on Theory and Application of Cryptographic Techniques. Springer-Verlag: 223–238.
80. ElGamal T (1985) A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans Inf Theory* 31: 469–472.
81. Yi X, Bouguettaya A, Georgakopoulos D, et al. (2016) Privacy protection for wireless medical sensor data. *IEEE Trans Dep Sec Comp* 13: 369–380.
82. Yi X, Paulet R, Bertino E (2014) *Homomorphic Encryption and Applications*. New York, Springer.
83. HElib, An open-source homomorphic encryption library for C++, https://github.com/shaih/HElib.
84. FHEW. An open source homomorphic encryption library for C and C++, https://github.com/lducas/FHEW.
85. Ames S, Venkitasubramaniam M, Kocabas O, et al. (2015) Secure health monitoring in the cloud using homomorphic encryption: a branching-program formulation. *Enabling Real-Time Mobile Cloud Comput Emerg Technol* 1: 116–152.
86. Page A, Kocabas O, Ames S, et al. (2014) Cloud-based secure health monitoring: Optimizing fully-homomorphic encryption for streaming algorithms. Globecom Workshops. *IEEE*: 48–52.
87. Kocabas O, Soyata T, Couderc JP, et al. (2013) Assessment of cloud-based health monitoring using homomorphic encryption. International Conference on Computer Design. *IEEE*: 443–446.
88. Kocabas O, Soyata T (2014) Medical data analytics in the cloud using homomorphic encryption. *Handbook Res Cloud Infrastructures Big Data Analytics*: 471–488.

89. Kocabas O, Soyata T (2015) Medical data analytics in the cloud using homomorphic encryption. *E-Health Telemed Concept Methodolog Tool Application* 2: 751–768.
90. Yi X, Miao Y, Bertino E, et al. (2013) Multiparty privacy protection for electronic health records. *GLOBECOM-IEEE Global Telecomm*: 2730–2735.
91. Wang X, Zhang Z (2015) Data division scheme based on homomorphic encryption in WSNs for health care. *J Med Syst* 39: 1–7.
92. Kocabas O, Soyata T (2015) Towards privacy-preserving medical cloud computing using homomorphic encryption. *Enabling Real-Time Mobile Cloud Comput Emerging Technol* 1: 213–246.
93. Nagapriya G, Retnaraj J (2015) Securing the privacy of sensitive data on health management system using ElGamal encryption. *ARPN J Eng Appl Sci* 10: 5802–5806.
94. Yi X, Paulet R, Bertino E, et al. (2014) Practical k nearest neighbor queries with location privacy. *Proc Int Conf Data Eng*: 640–651.
95. Paulet R, Kaosar MG, Yi X, et al. (2014) Privacy-preserving and content-protecting location based queries. *IEEE Trans Knowledge Data Eng* 26: 1200–1210.
96. Yi X, Paulet R, Bertino E, et al. (2016) Practical approximate k nearest neighbor queries with location and query privacy. *IEEE Trans Knowledge Data Eng* 28: 1546–1559.
97. Vasukidevi A, Jayalakshmi M, Gomathi V (2016) Secure communication between wireless medical sensor networks and data servers using Paillier and ElGamal key cryptosystem. *Int Conf Comp Technol Intel Data Eng*. Article number 7725333.
98. Carpov S, Nguyen TH, Constantino G, et al. (2017) Practical privacy-preserving medical diagnosis using homomorphic encryption. *IEEE Int Conf Cloud Comput*: 593–599.
99. Muralidhar K, Sarathy R, Parsa RA (1999) A general additive perturbation method for database security. *Management Sci* 45: 1399–1415.
100. Agrawal D, Aggarwal CC (2001) On the design and quantification of privacy preserving data mining algorithms. *Principle Database System*: 247–255.
101. Agrawal R, Srikant R (2000) Privacy-preserving data mining. *Proc ACM SIGMOD Conf Management Data*: 439–450.
102. Rizvi SJ, Haritsa JR (2002) Maintaining data privacy in association rule mining. *Proc 28th Int Conf Very Large Data Bases*: 682–693.
103. Evfimievski A, Srikant R, Agrawal R, et al. (2002) Privacy preserving mining of association rules. *Proc 8th ACM SIGKDD Int Conf Knowledge Discovery Data Mining*: 217–228.
104. Sweeney L (2002) K-anonymity: a model for protecting privacy. *Int J Uncert Fuzz Knowledge-Based Syst* 10: 557–570.
105. Lindell Y, Pinkas B (2002) Privacy preserving data mining. *J Cryptology* 15: 177–206.
106. Kantarcioglu M, Clifton C (2004) Privacy-preserving distributed mining of association rules on horizontally partitioned data. *IEEE Trans Knowledge Data Engineering* 16: 1026–1037.
107. Yi X, Zhang Y (2007) Privacy-preserving distributed association rule mining via semi-trusted mixer. *Data Knowl Eng* 63: 550–567.
108. Yi X, Zhang Y (2009) Privacy-preserving naive Bayes classification on distributed data via semi-trusted mixers. *Inf Syst* 34: 371–380.
109. Yi X, Zhang Y (2013) Equally contributory privacy-preserving k-means clustering over vertically partitioned data. *Inf Syst* 38: 97–107.

110. Yi X, Rao FY, Bertino E, et al. (2015) Privacy-preserving association rule mining in cloud computing. *Proc 10th ACM Sym Inf Comp Comm Sec*: 439–450.
111. Rao FY, Samanthula BK, Bertino E, et al. (2015) Privacy-preserving and outsourced multi-user k-means clustering. *Proc IEEE Conf Collab Internet Comp*: 80–89.
112. Liu D, Bertino E, Yi X (2014) Privacy of outsourced k-means clustering. *Proc 9th ACM Symp Inf Comp Comm Sec*: 123–133.

AIMS Press