Control and Cybernetics

vol. **49** (2020) No. 4

Privacy preservation for the health care sector in a cloud environment by advanced hybridization mechanism*

by

Annie M. M. Alphonsa and N. Mohanasundaram

Karpagam Academy of Higher Education, Salem - Kochi Highway, Eachanari, Coimbatore, Tamil Nadu 641021, India
annie.alphonsa@yahoo.com

Abstract: Cloud computing is a very popular computing model, which grants a manageable infrastructure for various kinds of functions, like storage of data, application realization and presenting, and delivery of information. The concept is therefore very dynamically advancing in all kinds of organisations, including, in particular, the health care sector. However, effective analysis and extraction of information is a challenging issue that must find adequate solutions as soon as possible, since the medical scenarios are heavily dependent on such computing aspects as data security, computing standards and compliance, governance, and so on. In order to contribute to the resolution of the issues, associated with these aspects, this paper proposes a privacy-preserving algorithm for both data sanitization and restoration processes. Even though a high number of researchers contributed to the enhancement of the restoration process, the joint sanitization and restoration process still faces some problems, such as high cost. To attain better results with a possibly low cost, this paper proposes a hybrid algorithm, referred to as Glow Worm Swarm Employed Bee (GWOSEB) for realization of both data sanitization and data restoration process. The proposed GWOSEB algorithm is compared as to its performance with some of the existing approaches, such as the conventional Glowworm Swarm Optimization (GSO), FireFly (FF), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), Genetic Algorithm (GA), and Genetically Modified Glowworm Swarm (GMGW), in terms of analysis involving the best, worst, mean, median and standard deviation values, sanitization and restoration effectiveness, convergence analysis, and sensitivity analysis of the generated optimal key. The comparison shows the supremacy of the developed approach.

Keywords: medical data, privacy preservation, Glowworm Swarm Optimization, data hiding, data restoration

*Submitted: April 2020; Accepted: March 2021

1. Introduction

The cloud computing technology provides an easy and relatively inexpensive on-demand network access (Sahi, Lai and Li, 2016), which is also simple to install. Generally, cloud computing is an evolving substructure and software model for enabling virtually unlimited access to the shared pools of configurable properties, like computer networks, network services, applications of the network, and so on. The technique currently becomes a vigorous technology landmark, and numerous researchers and scientists have declared that cloud computing already altered the computing processes and the respective Information Technology (IT) markets. Accessing via cloud computing can allow the users to utilize the comprehensive resource sets in terms of accessing various platforms, storage capacities, etc. via the internet, and also the services granted by cloud providers. The National Institute of Standards and Technology (NIST) admits that cloud computing is the most advantageous model for the operation of multiple computer resources and also for other new practical applications in the IT world (Zhou, et al., 2015a; Gatzoulis and Iakovidis, 2007).

However, many of the potential cloud customers are still not confident enough to secure for themselves the advantages from the cloud computing process, as security and privacy in the cloud are suffering from many issues (Takabi, Joshi and Ahn, 2010; Zissis and Lekkas, 2012; Grobauer, Walloschek and Stocker, 2011). In particular, privacy protection (Zhang et al., 2014a; Bianchini et al., 2017) is considered as one of the crucial issues in this model. Recently, many companies and especially the entities involved in the health care sector (Liu, et al., 2016; Barua, et al., 2011; Viswanathan, Chen and Pompili, 2012), like hospitals and clinics, have organized their medical services software and applications in the cloud, like, for example, Microsoft Health Vault (Lee, Song and Kim, 2016). The data sets that are involved in cloud applications include also sensitive data (Zhang et al., 2013). Achieving better access to health-related data (Azadeh et al., 2008; George and Rajakumar, 2013) is the most vital requirement for both medical doctors and researchers in medical and pharmaceutical sciences, and is of paramount importance for the in-depth studies of particular diseases (Manfredini et al., 2016; Gracco et al., 2007; Lombardo et al., 2014; Bossolasco and Fenoglio, 2018; Manassero et al., 2014). Nowadays, the cloud computing technique has significantly enhanced its accessing services, this applying also to the health care area and provision of health care related information (Zhang et al., 2014 a; Barbosa et al., 2019), consequently ensuring high-quality on-demand services with minimum cost (Wang et al., 2015; Moreira et al., 2019).

Although the cloud provides numerous helpful healthcare oriented services, some associated privacy issues are still being deliberated by both government and private entities (Wang, Chen and Zhang, 2015). Privacy risk increases when outsourcing a person's health care records, which are highly sensitive (Wang, Chen and Zhang, 2015; Iakovidis, 1998), to the cloud. Further, the cloud is char-

acterized in terms of an honest-but inquisitive model, this aspect being often perfected in the implementation of protocol requirements, regarding the mining of the private Personal Health Information (PHI) of patients from interfaces. Thus, the design of privacy-preserving, health-care-dedicated methods and applications (see Lu et al., 2012; Li et al., 2014; Lu, Lin and Shen, 2010; Shi et al., 2010, 2011) for data mining and mining of images constitutes the biggest issue that needs to be effectively solved today (Zhou et al., 2015a).

The privacy services are more required in granting privacy preserving solutions for cloud users. However, this is not so simple, since the communication between users and providers implies a lot more of consequences than just the transmitted contents. For instance, if a client allows a particular service to interconnect (and perform a service) with a specific user, it must be ensured that the additional ("other") information, related to the person, not be revealed (like his/her behavior, activities, social connections, etc.) to the cloud providers. Hence, the protection of clients' security needs not only to be achieved by doing just the encryption of data, but requires also some additional privacy defense measures in the cloud environment. There is now a number of privacy-preserving models, such as the data portioning model, 'Preserving cloud computing Privacy (PccP)' model, One Ring to Rule Them All (ORUTA) model (Nallakumar, Sengottaiyan and Arif, 2014), etc., which have been proposed for securing the privacy in the cloud. Yet, in general terms, the attainment of truly satisfactory results in data privacy is still a challenging problem, and hence, in particular, an effective privacy-preserving model is required for better preservation of medical (health care) data.

The main contributions of the research work here reported are as follows:

- Proposal of an effective privacy preservation model using a hybrid algorithm, referred to as Glow Worm Swarm Employed Bee (GWOSEB).
- The hybridization model proposed is used to generate the key for both data sanitization and the data restoration process.

The rest of the paper is arranged as follows: Section 2 reviews the literature of the subject. Section 3 details the privacy preservation framework for medical data. Section 4 explains the proposed key extraction procedure. Section 5 discusses the obtained results, and Section 6 concludes the paper.

2. Literature survey

2.1. Related work

Zhou et al. (2015a) developed a secure and efficient privacy-preserving technique, involving the extraction of features from the image model, termed Privacy-Preserving Data Mining (PPDM). First, an efficient privacy-preserving complete data aggregation method was proposed. Then, an early intervention and out-

sourced disease modeling technique was proposed, which turned out to be successful due to the introduction of an operative privacy-preserving correlation function that coordinated with PPDM1 from dynamic medical text mining and PPDM2 from feature extraction of the medical image. The proposed methodology was compared with conventional approaches in terms of their performance. The obtained results have shown the superiority of the proposed methodology over the other ones considered as displaying higher security levels. Further, the performance was also demonstrated regarding both computational and communication overhead.

Liu et al. (2016) developed a new privacy preserving approach termed 'patient-centric clinical decision support system', which is meant to aid clinicians in determining patient diagnosis of disease, especially in private practice. In this approach, past patients' details were saved in the cloud, and the respective data could also be used to help in training the classifier (naïve Bayesian) without revealing the data privacy of patients. Next, the trained classifier could evaluate the risk of diseases for new patients. Also a novel 'additive homomorphic proxy aggregation model' was designed for the protection of patients' data. The overall analysis has assured the patients' data privacy in the developed model with high accuracy.

Wang, Chen and Zhang (2015) developed a privacy preservation method to transit insensitive information to the public cloud and the residual data to the vital private cloud. In this model, two protocols were designed to provide personalized privacy protection. The developed model also protected against collusion between the service providers and users. The authors have derived an evident privacy assurance and restricted the alteration to prove the proposed protocols. The results of investigations were verified against a real-time scenario. The outcome has shown the usability of the developed model.

Zhang et al. (2014 a) developed a 'priority-based health data aggregation (PHDA)' method in the context of privacy preservation for the enhancement of aggregation among several health-related data types. Initially, they explored social spots for helping forward medical data and have allowed the users for selecting the optimal relay under social ties. The analysis of the developed method was carried out, and it was established that the PHDA could achieve data identity and 'data privacy preservation.' Further, it was shown that the developed model can resist some kinds of attack, like a forgery attack. Finally, the analysis of performance showed that the PHDA could achieve the required delivery ratio with minimum communication costs.

Sahi, Lai and Li (2016) made a survey on the literature dealing with security as disaster recovery strategy and privacy preservation, especially in the electronic-health cloud domain. These authors reported on having developed two models and one 'disaster retrieval plan'. The privacy model was robust

enough to assure the security and integrity of e-Health data. The developed model has an efficient authentication technique, which is supposed to protect the security of the person's medical (health) data. Further, the authors quoted introduced a 'three-party password-based authentication key exchange protocol (3PAKE)', and finally, they have shown the ideas for disaster recovery, meant to assure the security along with the cloud reliability. The model provided a feature named 'break-glass access' that could be helpful in disaster situations.

Waqar et al. (2013) reviewed the possibility of exploiting the metadata that has been stored in the cloud's database for preserving the data privacy of the cloud users. Then they have adopted a framework that belongs to the database schema. With the use of sensitivity parameterization of parent class membership, they have altered the database method by using cryptographic and interactive privacy preservation procedures. At the same time, the unchanged file access of the database was assured for the cloud provider to have aided in dynamic reconstruction (metadata). This was aimed at implementing the restoration process. Moreover, the assurance of suitability of the developed model was processed by evaluating its corresponding steps.

Zhang et al. (2013) developed an efficient quasi-identifier index-based approach for assuring privacy preservation, which worked well for the distributed as well as incremental data sets on the cloud. 'Quasi-identifiers' represent the group of anonymized data, which are indexed for effectiveness. Additionally, an algorithm was designed to implement the concept developed. Along with this, a locality sensitive hash function was utilized for placing the same quasi-identifier. The efficiency of the developed privacy preservation method was shown to be significantly better when compared against the conventional models.

Chandramohan et al. (2017) reviewed the security-related solutions and the corresponding issues regarding both intellectual and confidential data that are owned and used by different sectors, including insurance and finance. Security issues in the business fields are related to the legal and financial sensitivity of the data in question. With the expansion of cloud applications, the privacy of users became a big question also in this field. Furthermore, the specific characteristics of the cloud provider, reliability as well as the maintainability of the services, which definitely vary significantly, have also to be accounted for and possibly verified. The authors mentioned developed a privacy preservation approach named 'Prevent Digital Data Loss in the cloud (PPM-DDLM)'. Moreover, the developed model assists the Cloud Requesters (CR) in developing trust as to their proprietary data stored in the cloud.

Zhou et al. (2015a) presented a key management system for both time- and location-based mobile attacks with the cooperation of mobile patients in the same social group, the system having distributed and hierarchical character. In the framework of this approach, by utilizing the blinding technique

and implanting the human body's symmetric structure into Blom's symmetric key mechanism through modified proactive secret sharing patient's identity privacy, location privacy, and sensor deployment privacy were protected. Finally, the simulation results show a better result when compared with the previously known approaches in terms of computation and communication overhead, resisting mobile attacks, and storage requirements.

Zhou et al. (2015b) developed white-box traceable and revocable multi-authority encryption system, called TR-MABE to proficiently accomplish multi-level privacy preservation without initiating additional special signatures. Also, the presented scheme was able to prevent secondary physicians from knowing the patient's identity. Moreover, the system could effectively track the physicians who would leak the secret keys used to protect PHI and patients' identities. Finally, systematic security evidence and comprehensive simulations demonstrated the efficacy and practicality of the developed TR-MABE in e-health cloud computing systems.

Wang et al. (2019) presented a scheme based on the homomorphism concept for data processing in eHealth domain and privacy protection. The scheme proposed limits the arbitrary actions of patients and doctors.

2.2. Review

Let us now address some of the features and challenges, related to the approaches reported in the literature we surveyed. Thus, the PPDM algorithm (Zhou et al., 2015a) minimizes the communication cost along with the computational cost. Nevertheless, the application of the model was quite difficult. The naïve Bayesian classifier (Liu et al., 2016) can resist the attack called collision attack and attains efficient privacy preservation, but, since it is somewhat more complex, the communication, as well as computational overheads, are higher. The greedy algorithm from Wang, Chen and Zhang (2015) exhibited modest running time of the process. The PHDA approach (Zhang et al., 2014) can secure the individuals' identity and data privacy, but its application turned out to be a complex task. Further, the model must be lightened for attaining better outcomes in correspondence with its efficiency. The quasi-identifier index (Zhang et al., 2013) provided high portability. Nevertheless, anonymized dataset scheduling was yet more challenging. The Privacy-Preserving Model to Prevent Digital Data Loss in the Cloud (PPM–DDLC) from Chandramohan et al. (2017) can solve both portability and privacy issues, related to the use of the cloud, but the model is not so trustworthy, because the method works only when there is an agreement on that cloud requesters or end users assure that every data have their privacy policy even when they use various Cloud service providers. Thus, to solve all the security challenges in the cloud, it is essential to have a truly effective privacy-preserving approach.

3. The proposed methodology for privacy preservation of medical data

The proposed medical data hiding and restoration process is illustrated in Fig. 1. Two phases compose the sanitization process, by which the sensitive data are preserved. The two phases are Data Hiding and Data Restoration. Data hiding is realised by generating the optimal key, and the preserved data are treated as sanitized data, which are to be sent to the receiver. On the receiver side, the receiver can view the original data only if the inverse of the same key is given, and this process is referred to as the restoration process. The received original data can be used for clinical diagnosis analysis, and the data from the report can be sent to appropriate patients.

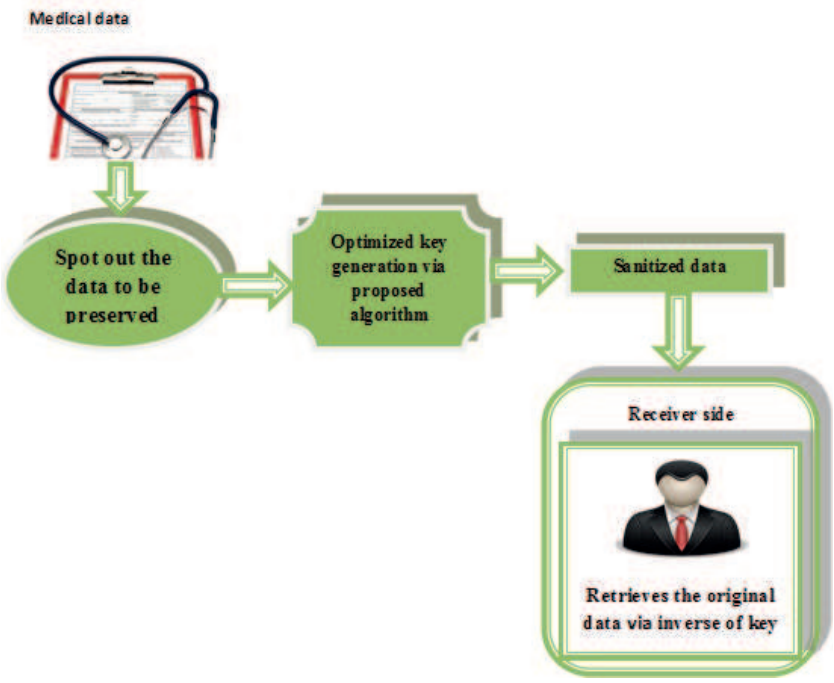

Figure 1: Architecture diagram of the proposed medical data preservation and restoration process

4. The proposed hybrid algorithm for key extraction

4.1. The hybrid GSO-ABC model

In 2005, the GSO algorithm (see Zhou et al., 2013) was proposed by Krishnanand and Ghose, this algorithm being an enhanced form of the Ant Colony Optimization (ACO) procedure. It is fully based on the glowworm metaphor and has been

applied in the context of collective robotics. Basically, in the GSO algorithm (Wu et al., 2012), a swarm that is here interpreted as a group of glowworms, is arbitrarily dispersed in the search space of the objective function. According to their position in the search space, the particular glowworms feature a certain quantity, in this case referring to the luminescent substance of luciferin, radiated by the particular glowworm individuals. The glowworms take their decisions on their behavior according to their decision domain $K_f^l \left(0 < K_f^l \leq K_v\right)$. Let l be the index of a glowworm and k the index of another glow worm, being the neighbour of the former if k is within the range of neighborhood of l and the luciferin level of k is higher than that of l. In such a case, the neighborhood is stated as a local-decision domain that has a K_f^l variable neighborhood range bounded by a radial (sensor) range. A distinct logic rules the communication between glowworms: they release light, which is directly proportional to the associated luciferin quantity and this happens within a flexible neighborhood. The intensity of glowworm's luciferin is related to the fitness (objective function value) of its respective current positions. When the intensity of luciferin is higher, it means that a better location of a glowworm has been found, corresponding to a better objective function value.

Consider a glowworm indexed g that treats another glowworm, indexed n, as its neighbor glowworm only if n is located in its neighborhood range. In particular, the neighborhood is defined as a local decision domain along the K_f^l, the neighborhood variable that is delimited by a value range named radial sensor range, $K_v \left(0 < K_f^l \leq K_v\right)$. The selection of glowworm takes place in a probabilistic setup, in which the neighbor glowworm with higher luciferin value than the currently considered glowworm moves toward it. Furthermore, the size of the neighborhood range of all glowworms is influenced by the glowworm's quantity within the range of neighborhoods. The glowworm's neighborhood range is proportional to the density of its neighbors. If the neighborhood range covers little density of glowworms, it will be enhanced, otherwise it will get decreased.

The GSO algorithm includes four phases, namely: (a) Initialization; (b) Luciferin update; (c) Movement; (d) Neighbourhood range update.

Initialization: In the initial phase, the glowworms are randomly dispersed in the objective function search space. Initially, all the glowworms contain an equal quantity of luciferin, LE_0.

Luciferin update: Luciferin intensity of glowworms is related to the fitness of their current positions. The locations of glowworms change in each iteration and the values of luciferin intensities are updated automatically.

The position (location) of the g^{th} glowworm at time ti is $Y_g(ti)$ and the corresponding value of the objective function is $N(Y_g(ti))$. In the next step, we

use the value of $N(Y_g(ti))$ in calculating $LE_g(ti)$, where $LE_g(ti)$ denotes the luciferin level of the g^{th} glowworm at time ti, according to Eq. (1):

$$LE_g(ti) = (1 - \upsilon)LE(ti - 1) + \gamma(N(Y_g(ti))) \tag{1}$$

where υ denotes the luciferin decay constant ($0<\upsilon<1$), and γ specifies the luciferin improvement coefficient.

***Movement*:** In this stage, all glowworms select their neighbors and then move toward them with a distinct probability. The selected g's neighbor glowworm must fulfill two conditions: first, this glowworm must lie within the decision domain of the g^{th} glowworm; and the second is that the value of luciferin of the selected glowworm must be higher than that of the g^{th} glowworm. The g^{th} glowworm moves towards its selected neighbor n that arrives from $Z_g(ti)$ with some probability $\Pr_{gn}(ti)$ and the respective formula is given in Eq. (2).

$$\Pr_{gn}(ti) = \frac{LE_n(ti) - LE_g(ti)}{\sum_{j \in Z_g(ti)} LE_j(ti) - LE_g(ti)}. \tag{2}$$

The position is updated after the g^{th} glowworm's movement, and the position update is performed according to Eq. (3), in which $SIZE$ refers to the step size.

$$Y_g(ti+1) = Y_g(ti) + SIZE * \left(\frac{Y_n(ti) - Y_g(ti)}{||Y_n(ti) - Y_g(ti)||} \right). \tag{3}$$

***Neighbourhood range update*:** Let K_0 be the initial neighborhood range of each glowworm. The neighbourhood range update follows the location update of the glowworm. If the neighborhood range covers lower density of glowworms, then the range will get extended, otherwise, the neighborhood range gets condensed or reduced. The updating formula is provided in Eq. (4), where β refers to a constant parameter, while n_p is a distinct parameter, meant for controlling the number of neighbors. The pseudo-code of the GWOSEB is given in Algorithm 1.

$$K_f^g(ti+1) = \min\left\{K_v, \max\left\{0, K_f^l(ti) + \beta(n_p - |Z_g(ti)|)\right\}\right\}. \tag{4}$$

The Artificial Bee Colony (ABC) algorithm, see Karaboga and Basturk (2008), tries to mimic the foraging behavior of honey bee colonies. In the ABC algorithm, three kinds of behaviour are distinguished: employed bee, onlooker bee, and scout bee. All three are involved in identifying the optimum food source. In this paper, only the employed bee aspect is considered, meant to be hybridized with GSO for identifying the optimal food source (here, optimal key). Initially, the food source is randomly generated, $F^S = 1, \ldots, S^P$, where S^Pdenotes the population size of food sources. The updating rule of the employed bee is given in Eq. (5), where $F_{\hat{k}}^S$denotes the solution within the neighborhood of $F_{\hat{i}}^S$and $\varphi_{\hat{i}}$ is a random number from the range [-1,1]:

$$\overline{F_{\hat{i}}^S} = F_{\hat{i}}^S + \varphi_{\hat{i}}\left(F_{\hat{i}}^S - F_{\hat{k}}^S\right). \tag{5}$$

Algorithm 1: Hybrid GWOSEB based key generation

Set the number of dimensions as n^d

Set the number of glowworms as n^G

Initialize LE_0 and K_0

Consider the step size $SIZE$

Let $Y_g(ti)$ denote the position or location of glowworm g at time ti

Randomly arrange the agents (glowwworms) in the search space

For $g = 1$ to n^G do $LE_g(0) = LE_0$

$K_f^g(0) = K_0$

Set MAX^{it} i.e. maximum iteration number

Set $ti = 1$

While $ti < MAX^{it}$ do

{

- For each glowworm g, determine $LE_g(ti)$ through Eq. (1)
- For each glowworm g do
- {
  - $Z_g(ti) = \left\{ n : e_{gn}(ti) < K_f^g(ti) : LE_g(ti) < LE_n(ti) \right\}$;
  - For each $g \in Z_g(ti)$ do
    - Calculate $\mathrm{Pr}_{gn}(ti)$ using Eq. (2)
  - n = choose glowworm (Pr)
- Find the probability $\overline{\mathrm{Pr}}$ using Eq. (6) for all n^G solutions based on the fitness function
- Choose a random number ra
- If $ra < \overline{\mathrm{Pr}}$ then
  - Update using Eq. (5)
- else
  - Calculate $Y_g(ti+1)$ using Eq. (3)
  - Calculate $K_f^g(ti+1)$ using Eq. (4)
- }end for
- $ti = ti + 1$

}

In the above scheme, $e_{gn}(ti)$ is the Euclidean distance between the glowworms g and n at time ti.

It is observed that the conventional GSO achieves less accuracy and suffers from slow convergence, while, at the same time, it is observed that the ABC algorithm is more flexible, robust, and can be easily calculated using fewer parameters. Hence, this paper intends to propose a new hybrid algorithm by making use of the advantages of both ABC and GSO to get the optimal key in the developed model. The proposed algorithm is given in Algorithm 1. The probability $\overline{\Pr}$ is calculated from Eq. (6), given below, for all the given solutions (i.e., for up to n^G glowworms). Then, a random value ra is chosen and it is checked whether the chosen value of ra is smaller than $\overline{\Pr}$, and, if so, then updating of the status of the employed bee is done using Eq. (5), else, updating is done using Eq. (3).

$$\overline{\Pr} = 0.9 \times \frac{Fitness}{\max(Fitness)} + 0.1. \tag{6}$$

In this manner, the proposed hybrid algorithm is meant to generate the optimum key for the data hiding process.

4.2. The data hiding process

The place and the general course of the data hiding process are illustrated in Fig. 2. This is the phase of data preservation, by which the sensitive (medical) data are preserved using the generated optimal key. To hide the data, the generated optimal key is converted into binary value and the data to be protected are multiplied by the created binary value, following which it is referred to as sanitized data.

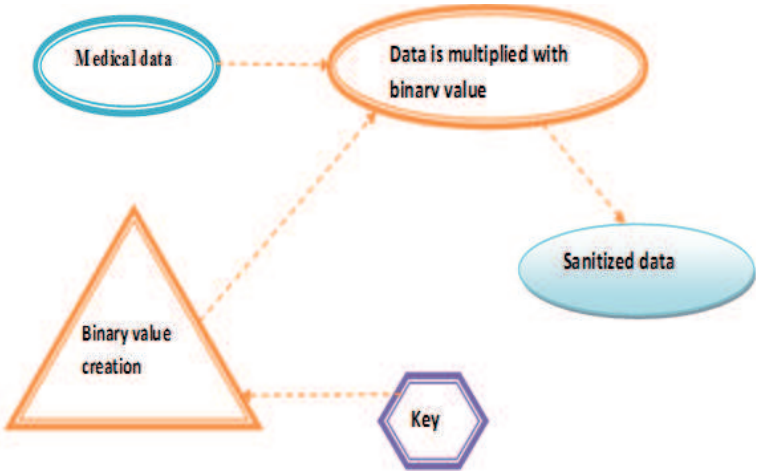

Figure 2: The illustration for the data hiding aspect of the process

The binary data creation scheme is as follows: Consider the data of size $D_1 \times D_2$ (for example, 200×4), where D_1 denotes the number of records and

D_2 denotes the number of fields. Let the size of the optimal key be 20×1, this being multiplied by the original data to create the sanitized data. The converted binary value must have the same length as that of the original data. For this, the elements in the key (20×1) are separated into five subsets, where each subset comprises 4 elements, and each element in the key is converted into 40 binary bits, by which each subset obtains 40×4 data elements. Thus, the partitioned set has five (40×4) data items, this being the generated binary data. The complete five (40×4) data items are concatenated to make the cumulative binary data of size 200×4. The obtained binary data are multiplied by the original medical data to obtain the sanitized data.

4.3. The data restoration process

The diagrammatic representation of the data restoration process is shown in Fig. 3. In the restoration process, the original data is restored from the sanitized data. This is achieved using the inverse of the generated optimal key. The inverse of the generated optimal key includes two pieces of information: the index and the sensitive data. First, the vector of the same length as the sanitized data is generated for sensitive data and then it is multiplied by the index of the key. The multiplied value is added afterwards to the sanitized data, which yields the restored data.

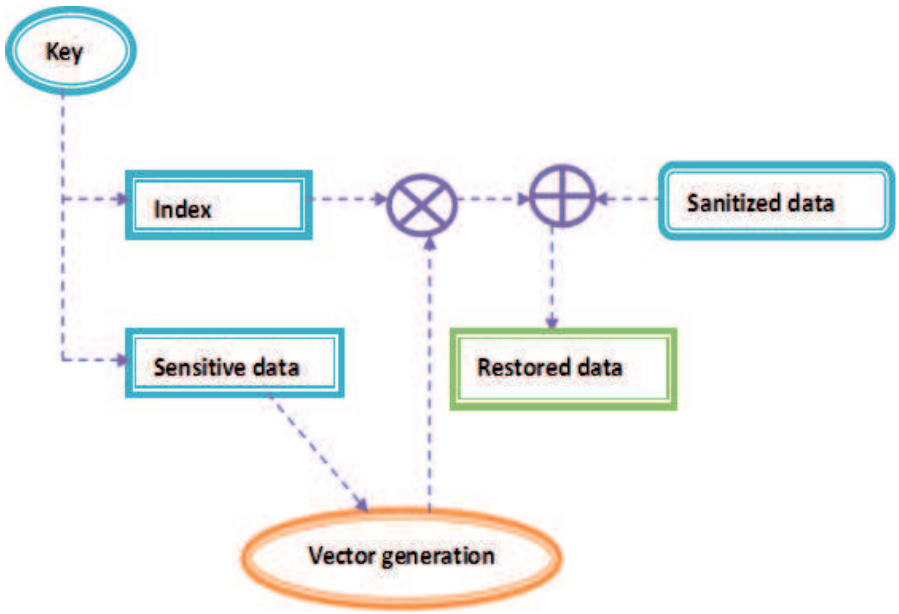

Figure 3: Scheme of the data restoration process

4.4. The objective function

The purpose of this paper is to present the method, in which the optimal key is obtained using the proposed GWOSEB algorithm. Here, the input solution is 'the key.' The length of the chromosome is $(D_1/40) \times D_2$. The minimum bound on the solution length is 1, and the maximum bound on the solution length is $2^b - 1$, where b is the number of bits, for instance, $b = 40$. Let the original data be d^{or}, the data to be preserved be d^{pr}, the sanitized data be d^{sa}; N^{data} denoting the number of data elements. Equations (8) and (7) establish the task to be solved and the corresponding objective function.

$$E = min(F) \tag{7}$$

$$F = \frac{\sum_{i=1}^{N^{ata}} d^{sa}}{\sum_{i=1}^{N} d_i^{or}} - \left[\frac{\sum_{i=1}^{N^{data}} d_i^{or} - \sum_{i=1}^{N^{data}} d^{pr}}{\sum_{i=1}^{N^{data}} d_i^{or}} \right]. \tag{8}$$

5. Results and discussions

5.1. Experimental procedure

The developed data sanitization and restoration model here presented was subject to experiments in MATLAB 2015. We shall discuss here the simulation outcomes. The experimental implementation was performed using heart disease data. The data set is of the dimensions [200×4], i.e., 200 records and 4 fields. The synthetic data were generated from the original data. The respective data, subject to experimentation, were varied by 10%, 20%, and 30%, respectively, corresponding to three test cases, the test cases 1 through 3. Random data were generated for each variation, for 10% variation random data were generated in the range of (-10% to + 10%), for 20% variation, data were generated in the range of (-20% to +20%), and for 30% variation, data were generated in the range of (-30 % to +30%), either by adding or subtracting the values. For each test case, ten synthetic data sets were generated. The performance of the proposed algorithm was compared with that of other existing algorithms, such as GA (Mc Call, 2005), ABC (Karaboga and Basturk, 2008), PSO (Tanweer, Suresh and Sundararajan, 2015), FF, i.e. firefly algorithm (Fister et al., 2013), GSO (Wu et al., 2012) and GMGW, the genetically modified glowworm swarm algorithm (Alphonsa and Amudhavalli, 2018).

5.2. Statistical analysis

Naturally, the meta-heuristic algorithms function according to a definite stochastic behavior, and hence they do not, in general, obtain an accurate optimum result. Thus, it is necessary to execute the algorithms several times – in our case

five times – for taking measures like best, worst, mean, median, and standard deviation. As already mentioned, the performance of the proposed model is compared to that of some other stochastic algorithms, namely GA, ABC, PSO, FF, GSO and GMGW. Tables 1 through 3 show the comparison mentioned for the three test cases. Particularly, in test case 1, the proposed method performed well in all runs, and for the best-case scenario, it is by 95.47%, 96.95%, 98.65%, 98.80%, 98.78%, and 98.82% better than GMGW, GSO, FF, PSO, ABC, and GA, respectively. Similarly, in all test cases, the analysis has shown the efficiency of the proposed model with better privacy preservation.

Table 1: Comparison of the proposed approach with some of the known ones for test case 1

Measure	GA (Mc Call, 2005)	ABC (Karaboga and Basturk, 2008)	PSO (Tanweer et al., 2015)	FF (Fister et al., 2013)	GSO (Wu et al., 2012)	GMGW (Alphonsa and Amudhavalli, 2018)	GWO-SEB
Best	1.4789	1.4391	1.4603	1.2933	0.573	0.3852	0.0174
Worst	1.5771	1.4796	1.5156	1.4028	0.754	0.7147	0.5546
Mean	1.5286	1.4548	1.4874	1.3594	0.6592	0.5866	0.3949
Median	1.5370	1.4479	1.4887	1.3895	0.6471	0.6525	0.4627
Standard deviation	0.0370	0.0180	0.0198	0.0494	0.0866	0.1390	0.2154

5.3. Attacks

In this section, attacks like Known Plain Text Attacks (KPA) and Cipher Plain Text Attacks (CPA) are analysed, and the results shown in Tables 4 and 5. The analysis regarding KPA is evaluated by correlating one original data with all original data and one sanitized data with all sanitized data. Similarly, the CPA analysis is carried out by calculating correlations of each sanitized data set with the corresponding restored data.

5.4. Convergence analysis

In general, the optimal key that is generated for the sanitization process is considered to be the best key if it has a distinct character: the cost function must decrease as the number of iterations increases. The analysis of convergence of the proposed model compared to some of the known methods for all three test cases is illustrated in Fig. 4. The analysis is carried out by varying the

Table 2: Comparison of the proposed method with some of the known ones for test case 2

Measure	GA (Mc Call, 2005)	ABC (Karaboga and Basturk, 2008)	PSO (Tanweer et al., 2015)	FF (Fister et al., 2013)	GSO (Wu et al., 2012)	GMGW (Alphonsa and Amudhavalli, 2018)	GWO-SEB
Best	1.5865	1.4961	1.5214	1.3556	0.6843	0.5941	0.5315
Worst	1.6835	1.5463	1.6402	1.4995	0.7904	1.1734	0.7599
Mean	1.6282	1.5186	1.5822	1.4032	0.7276	0.8181	0.6585
Median	1.6191	1.5138	1.5790	1.3944	0.7089	0.7060	0.6702
Standard deviation	0.0433	0.0208	0.0427	0.0570	0.0462	0.2454	0.1002

Table 3: Comparison of the proposed method with some of the known ones for test case 3

Measure	GA (Mc Call, 2005)	ABC (Karaboga and Basturk, 2008)	PSO (Tanweer et al., 2015)	FF (Fister et al., 2013)	GSO (Wu et al., 2012)	GMGW (Alphonsa and Amudhavalli, 2018)	GWO-SEB
Best	1.5511	1.582	1.625	1.4392	0.6760	0.3969	0.3919
Worst	1.7611	1.658	1.6963	1.5907	0.8080	0.9462	0.7491
Mean	1.6586	1.6200	1.6662	1.5300	0.7487	0.7569	0.5697
Median	1.6738	1.6178	1.6854	1.5322	0.7431	0.8308	0.5906
Standard deviation	0.0783	0.0267	0.0364	0.0578	0.0505	0.2319	0.1411

Table 4: CPA-related analysis of the proposed method compared with some known methods for test cases 1, 2 and 3

Method	**Test case 1**	**Test case 2**	**Test case 3**
GA (Mc Call, 2005)	0.97819	0.96071	0.99611
ABC (Karaboga & Basturk, 2008)	0.98771	0.95458	0.99037
PSO (Tanweer et al., 2015)	0.94195	0.99145	0.99625
FF (Fister et al., 2013)	0.98963	0.99784	0.99618
GSO (Wu et al., 2012)	0.95649	0.94364	0.99496
GMGW (Alphonsa & Amudhavalli, 2018)	0.97878	0.94094	0.96676
GWOSEB	0.97137	0.91578	0.95779

Table 5: KPA-related analysis of the proposed method compared with some known methods for test cases 1, 2 and 3

Method	**Test case 1**	**Test case 2**	**Test case 3**
GA (Mc Call, 2005)	0.99656	0.979	0.98123
ABC (Karaboga & Basturk, 2008)	0.99544	0.98084	0.9861
PSO (Tanweer et al., 2015)	0.99883	0.99028	0.9791
FF (Fister et al., 2013)	0.9866	0.99459	0.97926
GSO (Wu et al., 2012)	0.98936	0.97555	0.99051
GMGW (Alphonsa & Amudhavalli, 2018)	0.9629	0.97844	0.93071
GWOSEB	0.95423	0.97268	0.94779

number of iterations (20, 40, 60, 80, and 100) over different cost functions. Figure 4 shows for the proposed method the gradually decreasing cost function values, with the minimum values attained, within the experiment reported, at the 100^{th} iteration. A similar outcome is observed for test cases 2 and 3. This demonstrates that the proposed approach can effectively reduce the cost function value and that the efficacy of the developed approach is with this respect better than for the compared other known methods.

5.5. Key sensitivity

The sensitivity of the generated optimal key is investigated by varying key size in the proportions of 10%, 30%, 40%, 50%, and 70%, respectively. The correlation between the encrypted data using the original key and the key with variation must be low. Figure 5 shows the values of the correlation coefficient for the compared methods for different variation degrees of the key in the three here considered test cases. In almost all of the situations (test cases and variation levels) the proposed method is better than the other ones considered. Fig 5 (a) shows, for instance, that the proposed GWOSEB method for 30% variation is by 2.53%, 3.75%, 2.53%, 4.93%, 1.28% and 4.93% better than the methods like GMGW, GSO, FF, PSO, ABC and GA, respectively. The conclusion from this series of experiments is that the proposed method generally featured lower correlation compared to other methods, which suggests the efficiency of the proposed model regarding data privacy preservation.

5.6. Restoration effectiveness

Regarding the effectiveness of restoration, it is the correlation between the original data and the restored data that is analyzed. In these terms, the method here proposed is also better than the other ones considered. The analysis is again performed for all the three test cases, with 10 experiments performed for each case, as this is illustrated in Fig. 6. In test case 1, for experiment 10, the proposed model has attained high correlation, by 7.60%, 6.45%, 8.79%, 10%, 11.23%, and 12.5% better than GMGW, GSO, FF, PSO, ABC, and GA, respectively. For experiment 1, the proposed model is by 2.08%, 5.37%, 8.88%, 7.69%, 6.52% and 7.10% better than GMGW, GSO, FF, PSO, ABC, and GA, respectively. From the graph of Fig. 6, it is clear that the proposed method improves the restoration process over the other methods considered.

5.7. Sanitization effectiveness

Here, Fig. 7 shows the comparative study of the sanitization effectiveness of the proposed method, compared to the other ones, accounted for here. The proposed method has attained more effective minimization of the objective function over other methods here considered. In test case 2, the proposed model for experiment 1 is by 66.66%, 74.69%, 85.31%, 86.18%, 86.09%, and 86.79% better than GMGW, GSO, FF, PSO, ABC, and GA, respectively. Altogether, in all

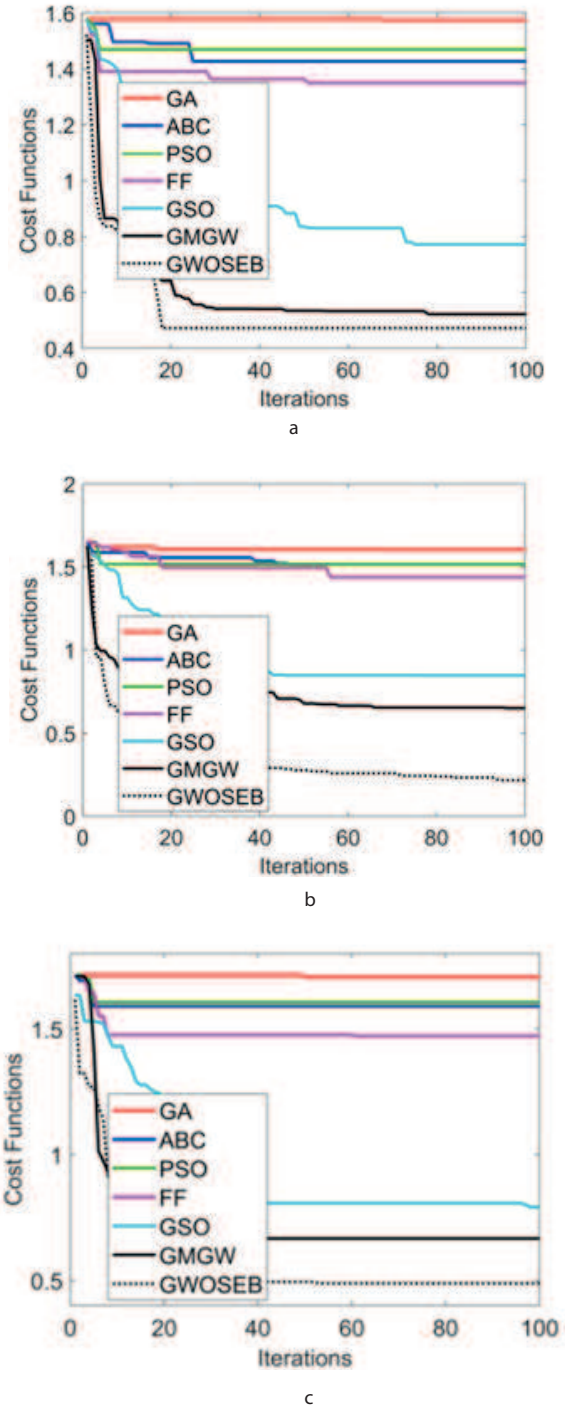

Figure 4: Convergence analysis of the proposed method and some of the known methods for (a) Test case 1 (b) Test case 2 (c) Test case 3]

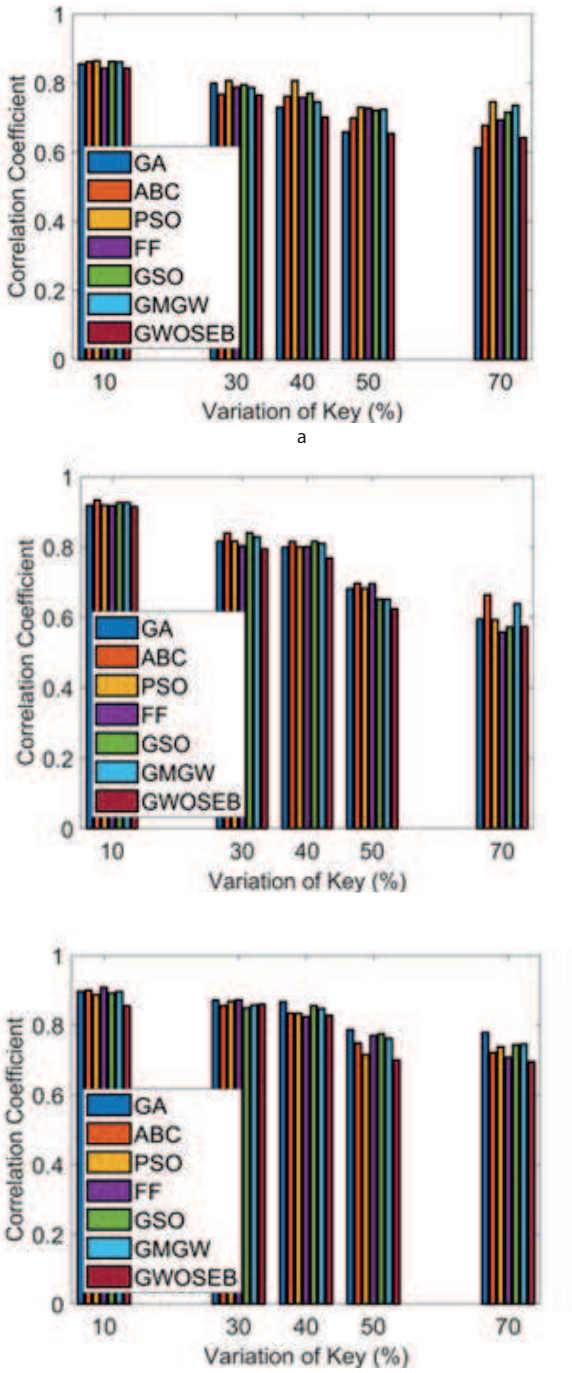

Figure 5: Key sensitivity in terms of correlation coefficient values for the proposed method and some of the known other methods for (a) Test case 1 (b) Test case 2 (c) Test case 3

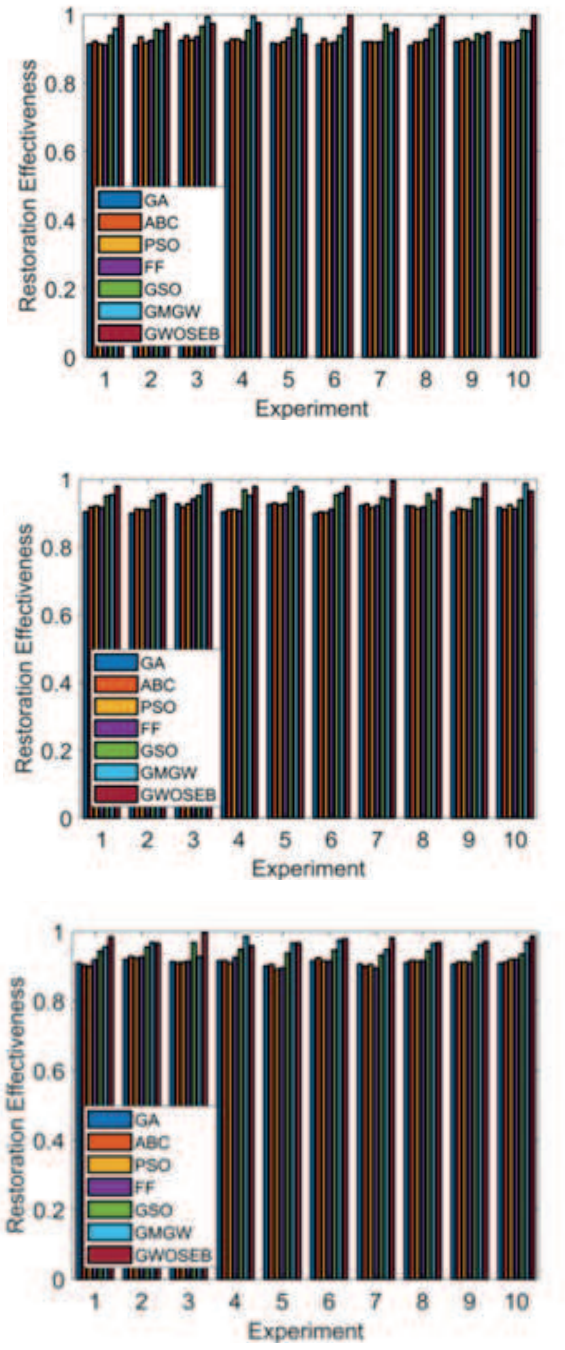

Figure 6: Restoration effectiveness of the proposed and the compared methods for (a) Test case 1 (b) Test case 2 (c) Test case 3

test cases, it was found that the developed approach is more efficient than the other ones considered, as it can effectively reduce the objective function value.

5.8. Sanitization and restoration effectiveness for varying population size

The effectiveness of the sanitization and restoration process of the developed approach is analyzed in this section for the different swarm population sizes, namely 10, 20, 30 40, and 50, respectively. The calculations were performed again for all three test cases, as this is shown in Fig. 8.

6. Conclusion

This paper proposed a hybrid GWOSEB algorithm for both data sanitization and restoration processes. This algorithm was meant to ensure improved privacy preservation of sensitive data. The new algorithm is the hybrid of the ABC and the GSO algorithms. This GWOSEB algorithm generates the optimal key used for both sanitization and restoration processes. The new algorithm was compared with some of the known methods, namely with GA, ABC, PSO, FF, GSO, and GMGW in terms of various aspects, such as sanitization and restoration effectiveness, convergence analysis, key sensitivity analysis, etc. The results demonstrated clearly the superiority of the proposed method regarding the effective privacy preservation of data. Thus, in particular, it was concluded that the proposed method attained minimum correlation (correlation among encrypted data using the original key and the key with variation) when compared with other methods. Therefore, by using the proposed GWOSEB algorithm the privacy of the medical data of a patient can be effectively improved.

References

Alphonsa, M.A. and Amudhavalli, P. (2018) Genetically modified glowworm swarm optimization based privacy preservation in cloud computing for healthcare sector. *Evolutionary Intelligence,* **11**(1-2): 101-116.

Azadeh, A., Fam, I.M., Khoshnoud, M. and Nikafrouz, M. (2008) Design and implementation of a fuzzy expert system for performance assessment of an integrated health, safety, environment (HSE) and ergonomics system: The case of a gas refinery. *Information Sciences,* **178**(22): 4280-4300.

Barua, M., Liang, X., Lu, R. and Shen, X. (2011) ESPAC: Enabling security and patient-centric access control for ehealth in cloud computing. *International Journal of Security and Networks,* **6** (2-3): 67–76.

Bianchini, V., Cecilia, M.R., Roncone, R. and Cofini, V. (2017) Prevalence and factors associated with problematic internet use: an Italian survey among L'Aquila students. *Riv. Psichiatr,* **52**(2): 90-93.

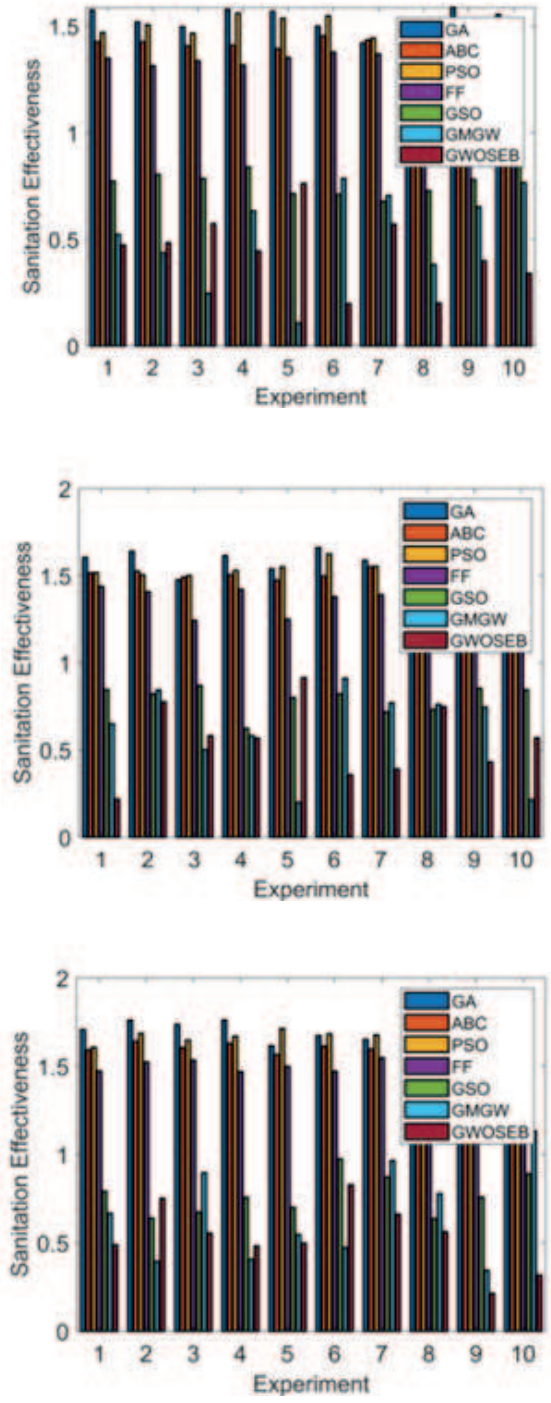

Figure 7: Sanitization effectiveness of the proposed and the compared methods for (a) Test case 1 (b) Test case 2 (c) Test case 3

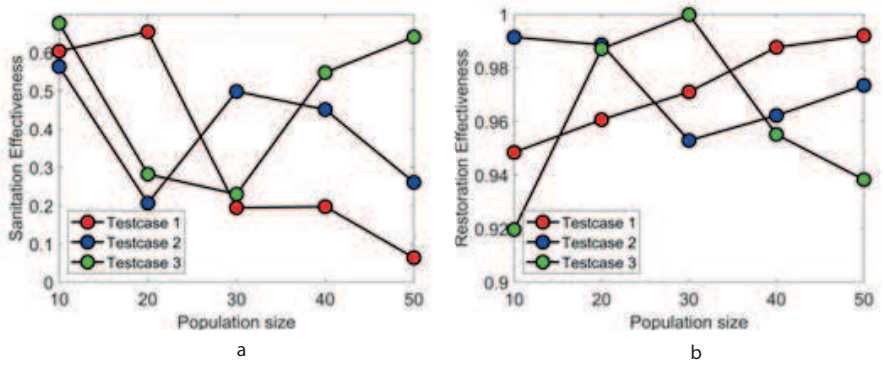

Figure 8: Effectiveness of sanitization and restoration process of the proposed approach for all the three test cases for different population sizes; (a) data hiding process (b) data restoration process

Bossolasco, M. and Fenoglio, L.M. (2018) Yet another PECS usage: A continuous PECS block for anterior shoulder surgery. *Journal of Anaesthesiology. Clinical pharmacology,* **34**(4): 569.

Chandramohan, D., Vengattaraman, T. and Dhavachelvan, P. (2017) A secure data privacy preservation for on-demand cloud service. *Journal of King Saud University-Engineering Sciences*, **29**(2), 144-150.

Fister, I., Fister, I., Yang, X.-S. and Brest, J. (2013) A comprehensive review of firefly algorithms. *Swarm and Evolutionary Computation,* 13: 34-46.

Gatzoulis, L. and Iakovidis, I. (2007) Wearable and Portable eHealth Systems. *IEEE Engineering in Medicine and Biology Magazine,* **26**(5): 51-56.

George, A. and Rajakumar, B.R. (2013) On Hybridizing Fuzzy Min Max Neural Network and Firefly Algorithm for Automated Heart Disease Diagnosis. In: *Fourth International Conference on Computing,Communications and Networking Technologies, Tiruchengode, India* (ICCCNT), 1-5. IEEE.

Gracco, A., Luca, L., Cozzani, M. and Siciliani, G. (2007) Assessment of palatal bone thickness in adults with cone beam computerised tomography. *Australian Orthodontic Journal,* **23**(2): 109.

Grobauer, B., Walloschek, T. and Stocker, E. (2011) Understanding Cloud Computing Vulnerabilities. *IEEE Security & Privacy,* **9**(2): 50-57.

Iakovidis, I. (1998) Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe. *International Journal of Medical Informatics,* **52**(1-3): 105-115.

Karaboga, D. and Basturk, B. (2008) On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing,* **8**(1): 687-697.

Lee, S.H., Song, J.H. and Kim, I.K. (2016) CDA Generation and Integration for Health Information Exchange Based on Cloud Computing System. *IEEE Transactions on Services Computing,* **9**(2): 241-249.

Li, H., Xiong, L., Ohno-Machado, L. and Jiang, X. (2014) Privacy preserving RBF kernel support vector machine. *BioMed Research International*. Article 827371.

Liu, X., Lu, R., Ma, J., Chen, L and Qin, B. (2016) Privacy-Preserving Patient-Centric Clinical Decision Support System on Naïve Bayesian Classification. *IEEE Journal of Biomedical and Health Informatics,* **20**(2): 655-668.

Lombardo, L., Scuzzo, G., Arreghini, A., Gorgun, Ö., Ortan, Y.Ö. and Siciliani, G. (2014) 3D FEM comparison of lingual and labial orthodontics in en masse retraction. *Progress in Orthodontics,* **15**(1): 38.

Lu, R., Liang, X., Li, X., Lin, X. and Shen, X. (2012) EPPA: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications. *IEEE Transactions on Parallel and Distributed Systems,* **23**(9): 1621-1631.

Lu, R., Lin, X. and Shen, X. (2010) SPRING: A Social-based Privacy-preserving Packet Forwarding Protocol for Vehicular Delay Tolerant Networks. In: *2010 Proceedings IEEE Infocom,* 1-9.

Manassero, A., Bossolasco, M., Ugues, S. and Bailo, C. (2014) An atypical case of two instances of mepivacaine toxicity. *Journal of Anaesthesiology. Clinical Pharmacology*, **30**(4): 582.

Manfredini, D., Segù, M., Arveda, N., Lombardo, L., Siciliani, G., Rossi, A. and Guarda-Nardini, L. (2016) Temporomandibular joint disorders in patients with different facial morphology. A systematic review of the literature. *Journal of Oral and Maxillofacial Surgery*, **74**(1): 29-46.

Mc Call, J. (2005) Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics,* **184**(1): 205-222.

Monkaresi, H., Calvo, R.A. and Yan, H. (2014) A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam. *IEEE Journal of Biomedical and Health Informatics,* **18**(4): 1153-1160.

Moreira, M.G., Ferraz, G.A.S., Barbosa, B.D.S., Iwasaki, E.M., Ferraz, P.F.P., Damasceno, F.A. and Rossi, G. (2019) Design and construction of a low-cost remotely piloted aircraft for precision agriculture applications. *Agronomy Research* **17**(5), 1984–1992.

Nallakumar, R., Sengottaiyan, N. and Arif, M.M. (2014) Cloud Computing and Methods for Privacy Preservation: A Survey. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* **3**(11).

Sahi, A., Lai, D. and Li, Y. (2016) Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. *Computers in Biology and Medicine,* 78, 1-8.

Shi, E., Chan, T., Rieffel, E., Chow, R. and Song, D. (2011) Privacy-preserving aggregation of time-series data. *Proc. NDSS* **2**, 1-17.

Shi, J., Zhang, R., Liu, Y. and Zhang, Y. (2010) PriSense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems. In: *2010 Proceedings IEEE INFOCOM, San Diego, CA,* 1-9.

Takabi, H., Joshi, J.B.D. and Ahn, G.J. (2010) Security and Privacy Challenges in Cloud Computing Environments. *IEEE Security & Privacy,* **8**(6): 24-31.

Tanweer, M.R., Suresh, S. and Sundararajan, N. (2015) Self regulating particle swarm optimization algorithm. *Information Sciences,* 294: 182-202.

Viswanathan, H., Chen, B. and Pompili, D. (2012) Research challenges in computation, communication, and context awareness for ubiquitous healthcare. *IEEE Communications Magazine,* **50**(5): 92-99.

Wang, W., Chen, L and Zhang, Q. (2015) Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation. *Computer Networks,* 88, 136-148.

Wang, X., Bai, L., Yang, Q., Wang, L. and Jiang, F. (2019) A dual privacy-preservation scheme for cloud-based eHealth systems. *Journal of Information Security and Applications,* 47: 132-138.

Waqar, A., Raz, A., Abbas, H. and Khan, M.K. (2013) A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata. *Journal of Network and Computer Applications,* 36, 235–248.

Wu, B., Qian, C., Ni, W. and Fan, S. (2012) The improvement of glowworm swarm optimization for continuous optimization problems. *Expert Systems with Applications,* **39**(7): 6335-6342.

Zhang, X., Liu, C., Nepal, S. and Chen, I. (2013) An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud. *Journal of Computer and System Sciences,* **79**(5): 542-555.

Zhang, K., Liang, X., Baura, M., Lu, R. and Shen, X. (2014a) PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs. *Information Sciences,* 284, 130-141.

Zhang, K., Liang, X., Shen, X. and Lu, R. (2014b). Exploiting multimedia services in mobile social networks from security and privacy perspectives. *IEEE Communications Magazine*, **52**(3), 58-65.

Zhang, X., Liu, C., Nepal, S. and Chen, I. (2013) An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud. *Journal of Computer and System Sciences,* **79**(5): 542-555.

Zhou, J., Cao, Z., Dong, X. and Lin, X. (2015a) PPDM: A Privacy-Preserving Protocol for Cloud-Assisted e-Healthcare Systems. *IEEE Journal of Selected Topics in Signal Processing,* **9**(7): 1332-1344.

Zhou, J., Cao, Z., Dong, X. and Lin, X. (2015b) TR-MABE: White-box traceable and revocable multi-authority attribute-based encryption and its applications to multi-level privacy-preserving e-healthcare cloud computing systems. In: *2015 IEEE Conference on Computer Communications*

(INFOCOM). IEEE, 2398-2406.

Zhou, J., Cao, Z., Dong, X., Xiong, N. and Vasilakos, A.V. (2015c) 4S: A secure and privacy-preserving key management scheme for cloud-assisted wireless body area network in m-healthcare social networks. *Information Sciences,* 314: 255-276.

Zhou, Y., Zhou, G., Wang, Y and Zhao, G. (2013) A Glowworm Swarm Optimization Algorithm Based Tribes. *Applied Mathematics and Information Sciences*, **7**(2): 537-541.

Zissis, D. and Lekkas, D. (2012) Addressing cloud computing security issues. *Future Generation Computer Systems,* **28**(3): 583-592.