Non-repudiation and privacy-preserving sharing of electronic health records

Mohamed Sharaf$^{1,2,*}$

Abstract: In this work, we present a framework that facilitates the sharing of EHRs among the community of health-care providers (HCPs). However, the sharing might be obstructed by patients’ privacy and the controlling legislation. Nevertheless, our sharing scheme of EHRs strives to respect patients’ privacy and comply with relevant legislator guidelines, e.g., HIPPA. The proposed work introduces two services while sharing the EHRs: privacy and non-repudiation. To this end, we introduce the partners and the role of each during the course of exchanging of EHRs. The principle of sharing EHRs among HCPs has to be reinforced to save patients’ lives while cryptographic primitives have to be employed to serve this purpose. In this paper, we are motivated to introduce the notion of non-repudiation private membership test (NR-PMT). In NR-PMT, we help patients receive medical services with great flexibility while maintaining their privacy and thwarting all possible threats to disclose their identities. In addition, a formal security analysis based on Kailar’s accountability framework has been used to analyze the proposed framework. Also, a complexity analysis has been conducted.

Subjects: Computing & IT Security; Computer Science (General); Information & Communication Technology (ICT)

ABOUT THE AUTHOR

Mohamed A. Sharaf received his BCs in computer engineering from Al-Azhar University in Egypt (1998), M.Sc. in computer engineering from Cairo University in Egypt (2005) and Ph.D. in computer engineering from the University of South Carolina (2014). He is currently an assistant professor at Computer Science and Engineering Dept., Al-Azhar University in Cairo, Egypt. His research areas of interest are computer network, artificial intelligence, and information security. He has been awarded a fellowship (2009–2012) from the Egyptian Cultural and Educational Bureau in Washington, D.C. to the University of South Carolina, where he worked on his Ph.D. In 2019, Dr. Sharaf has been granted a postdoctoral fellowship from the USAID. He spent his fellowship at High Performance Computing Lab, School of Engineering, George Washington University (GWU). During his fellowship, he worked on the convergence of 5G, AI and IoT. Also, Dr. Sharaf is a visiting assistant professor at Computer and Information Sciences College, Jouf University.

PUBLIC INTEREST STATEMENT

Free-Space Optical (FSO) Communication has fundamentally altered the way people communicate. It allows for efficient voice, video, and data transmission over a medium such as air as an alternative to wire communication systems. The primary benefits of FSO include high speed, cost savings, low power, energy efficiency, maximum transfer capacity, and applicability. However, atmospheric turbulence such as fog, rain, haze is one of the main problems that seriously limits the transmitted data rate in FSO systems. Atmospheric turbulence causes signal distortion during transmission that leads to data loss at the receiver side totally or partially. As a result, this study attempts to mitigate adverse effects of atmospheric turbulence, including clear weather, medium haze, medium rain, and medium fog during transmission of the data through the FSO link and increases the capacity of data transmitted and the distance under diverse atmospheric turbulence. Increasing the data rate transmitted through the FSO link enables the use of FSO in various applications essential to our daily lives.
Keywords: Privacy; electronic health record; non-repudiation; Bloom filter; accountability; authentication

1. Introduction

The main focus of this work is on Electronic Health Record (EHR), which is the aggregate of patient’s health-related information. EHR includes the patient’s medical history, laboratory data, radiology images, progress report, immunization, and medications. EHR has been adopted by policymakers to replace the paper-based system and to enhance the service provided at hospitals and doctors’ offices through the elimination of manual handling of patients’ data. In addition, EHR could help in preventing dishonest claims and achieving better coordination between health-care service providers. It seems that EHR is a promising new technology but with an expected financial cost and potential concerns as well. From one hand, the cost of deploying EHRs to the cloud has been already estimated and could fit or marginally exceed the budget, but eventually it will be within the acceptable limits. On the other hand, sharing EHRs among caregivers has significant concerns, such as privacy and authentication. Therefore, a great body of research focuses on achieving the aforementioned security. To this end, this work aims at promoting a privacy-preserving, accountable, and non-repudiation collaboration between health-care providers. However, such collaboration may incur a leakage of sensitive personal information that violates patients’ privacy.

To determine what may explicitly identify users, we follow the guidelines that have been enacted by the officials, i.e., personally identifiable information (PII; McCallister et al., 2010). PII can be used to uniquely identify an individual’s identity. PII is considered highly sensitive. Examples of attributes and assets that we consider PII: name, SSN, phone number, email address, IP address, MAC address, biometric data, face image, etc. The security risk of sharing PII information with other parties is the possibility of unauthorized access to data and thus identity theft, fraud, and misconduct.

However, contingencies sometimes mandate that patients need immediate health service, so access to their EHRs could be a factor in their survival. For example, a person travels abroad in a business trip and becomes in need for a procedure or a check-up. Care providers may be unwilling to embark a procedure before knowing about the medical history of the patient or instead waiting for a lab work that reveals whether a patient is allergic or intolerant to specific medications. Our model assumes that data trustees (custodians) are responsible for preserving patients’ records according to the relevant legislation and guidelines, i.e., HIPPA in the USA (United States, 1996), GDPR in the European Union (Union, 2016), etc. In medical sector, patient’s sensitive data in transit or at the recipient side cannot be used to identify the patient.

However, a custodian cannot share a patient’s EHR, because this is considered a violation of the patient’s privacy. Our proposed model assumes that the sharing process follows an anonymity approach that hinders privacy violations. Also, we suggest the presence of a trusted third party (TTP) that works as a library to facilitate the mapping/linking of patients to thier custodians or health-care providers. The difference between the scenario that we propose here and the well-known Privacy-preserving record linkage (PPRL; Fellegi & Sunter, 1969)–(Churches & Christen, 2004) is that the linkage key that is produced by the participant service providers is not necessarily unified among all participants and the check for membership is conducted at the TTP.

Contribution of this work is as follows: proposing a framework to allow secure sharing of EHRs that facilitates immediate interventions of care providers (CPs) in times of pandemics. As those CPs do not own patients EHRs, a collaboration framework with the guardians of the EHRs is inevitable to save lives. Having a contract or umbrella under which all participants have to deliver responsibly is mandatory. This work targets the sharing of EHRs without compromising the patients’ privacy or
leading to any kind of records linkability whatsoever. Also, at a point of cooperation, the framework guarantees that each participating CP has committed with its part undeniably.

2. Related work
Data hiding and the attainment of secure communication have been recently the main concern to attain secret sharing of sensitive data (Parah et al., 2017), (Sarosh et al., 2021). Next, we cover bloom filter with its variation and applications as we use it for membership check.

2.1. Matching under uncertainty
Approximate matching to test for a similarity between two sets A and B has been presented in several works; for more details check the survey by Vatsalan et al. (Vatsalan et al., 2013) and that of Dong et al. (Dong et al., 2013) where scalability is an issue.

In this work, we focus our model on approximate matching and test for membership with a level of confidence. In our settings of private set membership (PSM), we are given a group of n sets and a record, i. We may consider it as a variation of the privacy-preserving record linkage (PPRL) by assuming that the given record is a set of cardinality 1 and then look for the set with which an intersection exceeds a given confidence value, i.e., threshold. From practical standing, we adopt 2−gram decomposition, i.e., bigram. Let \( w \in \Sigma^* \) be a string and \( \Sigma \) denote the alphabet. Then, the 2−gram of \( w = \text{hassan} \) is expressed as follows:

\[ 2-\text{gram(hassan)} = \{ h, a, s, s, s, a, a, n, n, \} , \]

where \( _n \notin \Sigma \).

2.2. Similarity measure
Approximate matching in literature has been relied mostly on either Jaccard index or Dice coefficient with a scale that is preferably inclined more toward Dice coefficient. They only differ in the branch lengths. However, their dendrogram topologies are the same (Grzebala & Cheatham, 2016).—Given two sets \( A = 2-\text{gram (hassan)} \) and \( B = 2-\text{gram (hasan)} \), one can compute the Dice coefficient using the following equation:

\[
\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|}
\]  

\( (1) \)

For example, \( A = \{ h, a, s, s, s, a, a, n, n, \} \) and \( B = \{ h, a, s, s, a, a, n, n, \} \) have a Dice coefficient computed as follows:

\[
\text{Dice}(A, B) = \frac{2 \times 5}{6 + 7} \approx 0.77
\]

Definition 1 Threshold Dice coefficient (TDC). Given a record \( a \), a set \( A \), and a predetermined match threshold, \( k \), s.t., \( 0 \leq k \leq 1 \), TDC is defined as ThDice and is calculated by the following equation:

\[
\text{ThDice}(a, A, k) = \begin{cases} 
1 & \text{Dice}(a, A_i) \geq k \\
0 & \text{otherwise}
\end{cases}
\]  

\( (2) \)

For specific domains, even the size of items is considered sensitive information (Ateniese et al., 2011). For example, finding members of a flight who belong to the list that is banned from traveling to the USA has to be done with discretion as the Department of Homeland Security (DHS) cannot give away the size of Terror Watch List (TWL). Padding is a possible solution that is
adapted from literature–computer networking. However, padding incurs a sacrifice of computational resources and is considered a naive solution.

2.3. A membership test

Membership test between an entity and a given party A is shown in Algorithm 1.

Algorithm 1 Threshold-based membership test

Input: \((x, k, \{a_1, a_2, \ldots, a_n\})\), where \([a_1, a_2, \ldots, a_n]\) is A’s database of size \(n\), \(x\) is an entity and \(k\) is a threshold

Output: \(d = 1\) if \(x \in A\) or \(d = 0\) otherwise

Let temp \(\leftarrow 2 - \text{gram}(x)\);

for \(i = 1\) to \(n\) do

Let \(a_i \leftarrow 2 - \text{gram}(a_i)\);

Let \(d \leftarrow \text{ThDice}(a_i, \text{temp})\);

if \(d \geq k\) then

\(d \leftarrow 1\);

return \(d\);

end

end

d \leftarrow 0;

return \(d\);

3. Preliminaries

In our pursuit for space-efficient data structure, we consider Bloom filters (BF), devised by Bloom (Bloom, 1970). An alternative to BF could a cuckoo filter (CF), which is a space-efficient probabilistic data structure. One merit of cuckoo filter over BF is that the former supports the potential of deleting existing elements while the later does not support deletion of existing elements. Also, CF has lower space overhead than BF.

Several usages of BF and extensions have been proposed to add features such as scalability (Almeida et al., 2007). Also, BF has been applied in numerous disciplines, e.g., network applications such as resource routing, collaboration in overlay and peer-to-peer networks, packet routing, to name a few, see Broder and Mitzenmacher for more details (Broder & Mitzenmacher, 2004). In our study for efficient database representation and privacy-preserving, we adopt the strategy followed by Schnell et al. (Schnell et al., 2009).

A BF is an array of \(m\) bits with all bits initially set to zero. To map an object to a BF, the object is passed to a group of \(k\) independent hash functions, \(H = \{h_1, h_2, \ldots, h_k\}\).

The work by Bose et al. (Bose et al., 2008) proves that the false-positive rate (FPR) computed in Eq. 3 represents a lower bound of FPR.

\[
\text{FPR} = \left[1 - (1 - \epsilon)^n\right]^k \approx (1 - e^{-kn})^k, \tag{3}
\]

where \(m\) denotes the size of BF, \(k\) denotes the independent hash functions and \(n\) denotes number of objects mapped to the BF.

From Eq. 3, it is obvious that the value of \(k\) that minimize the FPR is computed as follows:
Table 1. Bloom filter and FPR

| FPR = 0.001 & \( k = 10 \) |
|---|---|
| \( n \) | \( m \) |
| 100 | 1438 |
| 1000 | 14378 |
| 10000 | 143776 |

\[ k_{opt} = \frac{m}{n}\ln2. \]

However, this is still an optimization to a lower bound FPR. In Table 1, we set FPR = 0.001, \( k \) (independent hash functions) = \( -\frac{\ln(FPR)}{m} \) and let \( n \), number of stored items, to take a value \( \in \{100, 1000, 10000\} \) and compute the Bloom filter length \( m = \frac{n\ln(FPR)}{(\ln(2))^2} \).

A possible solution to mitigate the effect of false positive is to let each custodian share non-discriminant (non PII) but enlightening information about the patient, e.g., 50 ≤ age ≤ 60 || telephone area code as last four digits of patient’s SSN, e.g., ***—9783, is considered PII. The returned information helps the enquirer to eliminate the majority of the ‘false’ custodians if not all of them except the right one.

Figure 1 shows approximate matching of two inputs, i.e., \( A = \) "HASAN", and \( B = \) "HASAN" by bigram Bloom filter with two hash functions. Computing the Dice coefficient, \( ThDice = \frac{2 \times 12}{19 + 19} = 0.96 \), yields that most likely there is a match, whereas exact matching returns null.

Another problem with custom Bloom filter is frequency attack. As we increase the number of hash functions among q-grams more positions will be shared and set to 1. Some techniques have been proposed to extenuate this as balanced Bloom filter presented by Schnell and Borgs (Schnell & Borgs, 2016) where each Bloom filter of length \( l \) is concatenated with its negation and then permuting the resulting 2l. Also, Rivest’s chaffing and winnowing (Bleumer, 2011) could be applied for the same purpose. To overcome frequency attacks, the main effort is revolving around masking the frequency distribution of the original filter. Figure 2 shows Bloom filter false positive rates based on different parametrization.

4. The proposed framework

Private set membership (PSM) is a tweak to the well-known private set intersection (PSI), a secure multiparty computation cryptographic technique. Figure 3 shows the interactions between an HCP and the TTP server. Initially, the HCP knows only its client, \( C \) and looks for the custodian of this \( C \)’s EHR. While TTP initially knows the populated sets of the subscribed HCPs, \( A = \{A_1, A_2, \ldots, A_n\} \).
Eventually, an HCP knows \( A_i \iff \exists A_i, \ s.t. \ C \in A_i \ or \ nil \) otherwise. Next, the HCP corresponds with \( A_i \) to retrieve the sought EHR that belongs to \( C \).

4.1. Publishing patients pseudonyms at a TTP

Next, we propose the usage of patient's ID to look for the health-care provider to which the patient’s EHR belongs. However, the search for membership has to respect patients’ privacy. Namely, the process is called private set membership (PSM; Meskanen et al., 2015)–(Tamrakar et al., 2017).

Let \( B = \{b_1, b_2, ..., b_n\} \) be the set of subscribed health-care providers (SHCP), s.t., \( \forall \text{PatientID} \in B, \forall i \in [1,n] \text{and} j \in [1,m] \), where \( n \) and \( m \) are number of SHCP and patients, respectively.

Figure 4 shows the framework and the interactions between the three participating parties, i.e., custodian, \( A \), requester, \( B \) and the trusted linkage library, \( C \). The first step begins with obtaining \( A \)'s consent to publish its records keys that are unidentifiable and cannot disclose its patients true
identity in the trusted library, C. Then, a patient who is obliged to have a medical service away from its HER’s custodian at healthcare provider B, step 2. So, B on behalf of the doctor will issue a request to linkage library asking for the patient’s EHR, step 3. By its turn, the trusted party, C, upon successfully identifying the custodian of the patient, forwards the enquiry to the corresponding custodian, step 4. Upon the reception of the enquiry from the trusted party, custodian A inspects its records and looks for a match if any, step 5. If there is a match, A sends B the patient’s sanitized EHR, step 6, otherwise A refrains or sends a decline to B, step 6.

However, sending patients’ medical records has to go through a sanitization process (Bier et al., 2009) in which sensitive data or generally any PII information is removed and replaced with meaningless information. The aim is to prevent the exposure of patients’ identities. Figure 5 shows typical patients’ data categorized into two overlapping sets, i.e., sensitive and medical data. Any item that falls in the sensitive class is wiped out before sharing non-linkable medical data. As we mentioned earlier, medical images could compromise patients’ identities. Yet, diagnosis, drug allergy and prescriptions may be considered – in our perspectives – as of less or zero sensitivity level and could be a life saving information in contingencies.

The linkage key that we seek to match has to be registered within a trusted third party. It is widely accepted that a patient may be authenticated using one factor or more of the three-factor authentication: 1) a factor that the patient knows, e.g., passwords, 2) a factor that the patient has, e.g., token, and 3) a factor that the patient is, e.g., bio-metric.
4.2. Medical data exchange

Algorithm 2: Query TTP for patient P’s custodian

Input: (The visiting patient Bloom filter);
Output: A list of potential candidates, L, which might own P’s HER;

\[ \forall \text{ custodian } i \in L \]

for \( i = 1 \) to \( n \) do

Query \( i \) for \( P \);
Response \( L[i] \) sends P’s EHR but non PII, however limiting, e.g., age \( \in [50 - 60] \);
Filter Inquirer overlooks (due to probabilistic nature of BF) seemingly non-conformant responses;
Ensure utility Inquirer may negotiate with \( L[i] \) more tightenings approaches without compromising patient’s privacy, e.g., affirming telephone area code (k-anonymity).
end

Algorithm 3: Transfer of an anonymized patient’s EHR

Input: \( M = K^{−1}_{\text{ID}} (B_{\text{ID}} + H(P_{\text{ID}}) + C_{\text{ID}}) \);
Output: HER, or Null;

\( \forall \text{ A asks C, trusted third party, to authenticate B}^* \)
authen = authenticate \( (C, B_{\text{ID}}) \);
auther = authorize \( (C, B_{\text{ID}}) \);
if (authen and auther) = True then
send anonymized HER ;
else
send decline;
end

Figure 6 shows a depiction of the proposed non-repudiation protocol in Alice/Bob notation. The aim is to warrant that no participating party denies receipt of the requested and delivered information. In addition, the protocol assumes that the trusted third party (TTP) is honest and will not collude with other parties. However, even if the TTP colludes, the collusion effect is dissipated by masking (anonymizing) patients’ sensitive data that might identify them directly or indirectly by linking with other datasets.

Table 2 shows the used symbols along with definitions.

The first two messages are between the requestor and the TTP. The aim is to figure out the custodian of that patient’s EHR. Once Alice, A, figures out the custodian, A forwards the request to Bob, B.

To ask trustee B for a patient’s EHR, party A sends a message that is composed of the participating parties identities along with \( \text{Sign}_{K^1_A} (X||\text{PatientID}) \) to vindicate the origin of the request. However, the signed part is encrypted with a shared secret key, \( K_1 \), that is shared between A and C. As the trustee cannot approve the request without the aid of a TTP, C, B, once agreed to process the request, signs \( \{X||Y\} \) using its private key and then forwards the message to C.

When C receives the message sent by B, C’s role is to check whether the request is valid and to prepare a response to B. Firstly, C authenticates the sender using its public key. Secondly, C decrypts \( \{\text{Enc}_C (K_1)\} \) using its private key. Now, C has an access to \( K_1 \) that enables it to proceed to the next step. Thirdly, once C verified the passed request from B and acquired \( K_1 \), C decrypts
Alice (A)                Bob (B)                Charlie (C)

1. Who is the owner of this patient’s EHR?
2. B is the owner; respond with:
   - $X \leftarrow A || B || C$
   - $Y \leftarrow \text{Enc}_K_2[H(PatientID_i) || N_1]$
   - $Z \leftarrow \text{Enc}_N_2(K_1)$
   - $T \leftarrow \text{Sign}_K^{-1}(X || H(PatientID_i))$
3. $X || Y || Z || \text{Enc}_N(T)$

4. $Y \leftarrow \text{Sign}_K^{-1}(X || H(PatientID_i))$
5. $X || Y || Z || \text{Enc}_N(T)$
6. $W \leftarrow \text{Enc}_N_2[H(PatientID_i) || N_1 || N_2 || N_3]$
   - $\text{PoREQO} \leftarrow \text{Sign}_K^{-1}(T)$
   - $\text{PoREQD} \leftarrow \text{Sign}_K^{-1}(Y || \text{PatientID_i})$
7. $X || Y || Z || \text{Enc}_N(T)$
8. $Y \leftarrow \text{Sign}_K^{-1}(X || H(EHR_1) || N_1 || N_2 || N_3)$
9. $\text{PoRESO} \leftarrow \text{Sign}_K^{-1}(T)$
10. $X || \text{Enc}_N_2(K_2) || \text{Enc}_N_2(\text{PoRESO})$
11. $\text{PoRESD} \leftarrow \text{Sign}_K^{-1}(Y || \text{EHR_1})$
12. $X || \text{Enc}_N_2(\text{PoRESD})$

$\text{[Enc}_K_1(\text{Sign}_K^{-1}_2[X || B || C || \text{PatientID_i}])$ using $K_1$, that $C$ has just accessed it, as shown in the previous step. Fourthly, $C$ authenticates the validity of the decrypted message using $A$’s public key, computes $H(PatientID_i)$ and compare it with the one that has been retrieved earlier. If the received hash and the computed one match, $C$ commences preparing a response to $B$ to enable it to reply to $A$’s request of $PatientID$, health record.

Next, $C$ has to send two messages: first message acts as a proof of request origin (PoREQO) and it is directed to $B$ and second message indicates a delivery of request to the recipient and acts as a proof of request delivery (PoREQO).

Then $C$ computes message 5 which contains POO and forward it to $B$. In message 5, PoREQO is computed as $T \leftarrow \text{Sign}_K^{-1}[\text{Sign}_K^{-1}(X || H(EHR_1))]$. Afterwards, $C$ computes message 6 which comprises POD and sends it to $A$. In message 6, POD is computed as follows:
Table 2. Symbol definitions

| Symbol   | Definition                                      |
|----------|-------------------------------------------------|
| PatientID | An ID that uniquely identify patient’s EHR     |
| MSP      | Medical Service Provider                       |
| MDC      | Medical Data Custodian                         |
| DP       | Differential Privacy                           |
| TTP      | Trusted Third Party                            |
| A, B, and C | Identity of party A, B and TTP, C           |
| Enc_{K_A} | Encryption using B’s public key                 |
| A        | Sign with B’s private key                      |
| H        | One way hash function                          |
| EHR, i   | Patient’s health record                        |
| PoREQO   | Proof of Request origin                        |
| PoREQD   | Proof of Request Delivery                      |
| PoRESO   | Proof of Response Origin                       |
| PoRESD   | Proof of Response Delivery                     |
| B → C : M | Party B sends message M to party C            |
| K_{1} and K_{2} | Shared secrets                             |
| N_{1}, N_{2}, and N_{3} | Nonces                                   |

\( M \leftarrow \text{Enc}_{K_{1}}(\text{Sign}_{K_{2}}(\text{Sign}_{K_{3}}(X|\text{PatientID}))) \). For reliability issue, A and B may acknowledge receipt of messages (PoREQO and PoREQD) and C may retransmit if timers expire before receiving acknowledgements.

Next, as B approves the request, B starts preparing the response by decrypting \( \text{Enc}_{K_{1}}(K_{1}) \) using its private key. Once B acquires the key, \( K_{1} \), it decrypts the part \( \text{Enc}_{K_{1}}(T) \) to obtain \( T \leftarrow \text{Sign}_{K_{1}}(\text{Sign}_{K_{2}}(X|\text{PatientID})) \). Then B uses C’s and A’s public keys successively to obtain \( X|\text{PatientID} \).

At that point, B has an access to the patient’s ID which will be used to retrieve and send the patient’s anonymized record to the requestor. First, B sign \( A|B|C|EHR_{i} \) with its private key and encrypt the result with a key, \( K_{2} \), that the requestor has to figure out by consulting C to obtain EHR.

Upon reception of the response from B, A knows it has to forward a request to C to help it to know \( K_{2} \). When C receives the request from A, C authenticates the origin of the message by using A’s public key. Then, C uses its private key to decrypt \( \text{Enc}_{K_{2}}(K_{2}) \) and obtain \( K_{2} \). Once \( K_{2} \) is acquired, C uses it to decrypt \( \text{Enc}_{K_{2}}(\text{Sign}_{K_{3}}(X|EHR_{i})) \) and sign the decrypted message using its private key. The previous step results in \( \text{Sign}_{K_{3}}(\text{Sign}_{K_{2}}(X|EHR_{i})) \) which is the PoRESO.

Finally, C has to send two more messages, each contains a proof ticket: first message carries a proof of response origin and second message holds a proof of delivery response.

5. Security analysis

Secure protocols may suffer one or more of the following attacks: Dictionary (guessing) attack, cryptanalysis attack, frequency attack, impersonation attack, denial of service (DoS) attack, replay attack, and other possible attacks, like collusion attack if the protocols count on third-parties. Table 3 presents a comparison between the proposed protocol and other related secure protocols.
Next, we present a detailed analysis of the proposed protocol accountability and non-repudiation.

5.1. Accountability and non-repudiation analysis

In addition, we utilize a familiar accountability framework that has been introduced by Kailar (Kailar, 1996) to prove our protocol accountability. The basic principle is that whether we have a proof of a statement validity or a mere belief of the statement validity; we look for the former.

According to Kailar (Kailar, 1996), principals are denoted by uppercase letters and statements are denoted by lowercase letters. Kailar introduces two proofs: strong proof and weak proof. In strong proof, a principal Y can prove a statement t. In this sense, Y, without disclosing secrets, follows a procedure and eventually convinces X of t. Our goals are to show that:

**G1**: CSP Can Prove (MSP Sent M) (Req.Sent)

**G2**: MSP Can Prove (CSP ReceivedM) (Req.Deliverd)

**G3**: MSP Can Prove (CSP SentM) (Res.Sent)

**G4**: CSP Can Prove (MSP ReceivedM) (Res.Delivered)

According to the construct, several assumptions are required:

A1. CSP, MSP CanProve ($K_{TTP}$ Authenticates TTP)

A2. CSP, TTP CanProve ($K_{MSP}$ Authenticates MSP)

A3. TTP, MSP CanProve ($K_{CSP}$ Authenticates CSP)

A4. CSP, MSP CanProve (TTP IsTrustedOn (TTP Says))

A5. (MSP Says M) $\Rightarrow$ (MSP Sent M)

A6. (CSP Says H(M)) $\Rightarrow$ (CSP Received H(M))

A7. (TTP Says (MSP, EHR)) $\Rightarrow$ (M SentTo MSP)

---

Table 3. Comparison of related secure protocols

|                | Randall et al. (Randall et al., 2014) | Jiang (Jiang et al., 2018) | Ramezanian et al. (Ramezanian et al., 2019) | Schnell and Borgs (Schnell et al., 2009) | Santiago and Vigneron (Santiago & Vigneron, 2007) | Hathaliya et al. (Hathaliya et al., 2019) | Proposed Framework |
|----------------|----------------------------------------|-----------------------------|---------------------------------------------|------------------------------------------|---------------------------------------------|------------------------------------------|-------------------|
| Privacy-preserving linking | ✓                                      | ✓                           | ✓                                           | ✓                                        | X                                           | X                                        | ✓                 |
| Non-repudiation   | X                                      | X                           | X                                           | X                                        | ✓                                           | X                                        | ✓                 |
| Trusted third party (TTP) | X                                      | ✓                           | ✓                                           | ✓                                        | X                                           | X                                        | ✓                 |
| Approximate matching | X                                      | ✓                           | ✓                                           | ✓                                        | X                                           | ✓                                        | ✓                 |
| Sanitization      | ✓                                      | ✓                           | ✓                                           | X                                        | X                                           | X                                        | ✓                 |
A8. \((MSP\text{Received}H(M)) \land (M\text{SentToMSP}) \Rightarrow MSP\text{Received}M\)

A9. \((CSP\text{Received}H(M)) \land (M\text{SentToCSP}) \Rightarrow CSP\text{Received}M\)

First, protocol messages interpretation as shown in Figure 6:

**Message 4**: This is interpreted as:

4.1 TTP, Receives \(H(PatientID)\) SignedWith \(K_{MSP}^{-1}\);

4.2 TTP Receives \(H(PatientID)\) SignedWith \(K_{CSP}^{-1}\)

From 4.1 and utilizing the signature postulate, we can reach the following statement:

TTP Can Prove (CSP Says H (Patient ID))

Moreover, applying assumption A6, we can infer:

TTP Can Prove (CSP Received H (Patient ID))

The above statement indicates that, undeniably, CSP has received a request. Next, from 4.2 and following the same construct we used in 4.1, we can reach the following statements:

TTP Can Prove (MSP Says Patient ID).

Moreover, applying assumption A6, we can infer:

TTP Can Prove (MSP Sent Patient ID).

Also, the above inference indicates that, certainly, MSP has requested the EHR of that patient.

**Message 5**: Upon reception of message 5, CSP can use TTP's public key and based on the association between \((K_{TTP} \text{ and } K_{TTP}^{-1})\), i.e., assumption A1, one can interpret message 5 using the signature postulate as:

CSP Can Prove (TTP Says (Patient ID Signed With \(K_{MSP}^{-1}\)))

Next, applying assumption A4 which indicates that TTP is an authority on whatever it says, one can derive

CSP Can Prove (Patient ID Signed With \(K_{MSP}^{-1}\)).

It follows that CSP can use assumption A2 and prove that \(K_{MSP}\) authenticates MSP, i.e., messages signed with MSP's private key. Hence, applying A2 and the signature postulate, we get

CSP Can Prove (MSP Says Patient ID)

We can further prune the above statement by applying assumption A5 to infer

CSPCanProve(MSPSentPatientID) \hspace{1cm} (G1)
It is apparent that G1 is a vivid proof of origin (POO), i.e., non-repudiation of origin while G2 is a proof of delivery (POD).

**Message 6:** This message is used a confirmation from TTP to MSP that CSP has received its request. Therefore, this message is a proof of request delivery (PoREQD). This message is interpreted as follows:

1. MSP Receives \(((\text{PatientID}) \text{ SignedWith } K_{CSP}) \text{ Signed With } K_{TTP}^{1}\)
2. MSP Receives (CSP, Patient ID) Signed With \(K_{TTP}^{1}\)

Upon reception of information shown in 6.1 and applying assumptions A1, A4 and A6, we can infer

\[
\text{MSP Can Prove (CSP Received H (Patient ID)).}
\]

Also, after attaining 6.2, we can deduce that

\[
\text{MSP Can Prove (TTP Says (Patient ID, CSP))}.
\]

Utilizing the above statement and applying assumption A7, one may infer

\[
\text{MSPCanProve(PatientIDSentToCSP). (iii)}
\]

By merging statements, (i) and (ii), we reach the following

\[
\text{MSP Can Prove ((Patient ID Sent To CSP) \land (CSP Received H (Patient ID))).}
\]

We can prune the above statement using assumption A9 to deduce

\[
\text{MSP Can Prove (CSP Received H (Patient ID)). (G2)}
\]

**Message 8:** This is interpreted as:

1. TTP, Receives H(EHR) SignedWith \(K_{MSP}^{1}\);
2. TTP Receives EHR SignedWith \(K_{CSP}^{1}\)

From 8.1 and utilizing the signature postulate, we can reach the following statement:

\[
\text{TTP Can Prove (MSP Says H (EHR)).}
\]

Moreover, applying assumption A6, we can infer

\[
\text{TTP CanProve (MSP Received H (EHR)).}
\]

The above statement indicates that, undeniably, CSP has replied to MSP’s request. Next, from 8.2 and following the same construct we used in 8.1, we can reach the following statements:
TTP Can Prove (CSP Says EHR).

Moreover, applying assumption A5, we can infer:

TTP Can Prove (CSP Sent EHR).

Also, the above inference indicates that, certainly, CSP has sent the patient's EHR.

**Message 9:** Upon reception of message 9, MSP can use TTP's public key and based on the association between (K_{TTP} and K_{TTP}^{-1}), i.e., assumption A1, one can interpret message 9 using the signature postulate as:

MSP Can Prove (TTP Says (EHR Signed With K_{CSP}^{-1})).

Next, applying assumption A4 which indicates that TTP is an authority on whatever it says, one can derive

MSP Can Prove (EHR Signed With K_{CSP}^{-1}).

It follows that MSP can use assumption A2 and prove that K_{CSP} authenticates CSP, i.e., messages signed with CSP’s private key. Hence, applying A2 and the signature postulate, we get

MSP Can Prove (CSP Says EHR)

We can further prune the above statement by applying assumption A5 to infer

MSP Can Prove (CSP Sent EHR)  \hspace{1cm} (G3)

It is apparent that G1 is a vivid proof of response origin (PoRESO), i.e., non-repudiation of origin while G2 is a proof of delivery (POD).

**Message 10:** This is interpreted as:

10.1. CSP, Receives (H(EHR) Signed With K_{MSP}^{-1}) Signed With K_{TTP}^{-1}.

10.2. CSP Receives (MSP, EHR) Signed With K_{TTP}^{-1}

From 10.1, applying the same approach we followed in message 8 and by utilizing assumptions A1, A4 and A6, we can infer

CSP Can Prove (MSP Received H (EHR)) \hspace{1cm} (iii)

Upon reception of 10.2, CSP get to know the identity of both MSP and TTP as both are signed by TTP. Upon applying the signature postulate, we infer

CSP Can Prove (TTP Says (MSP, EHR)).

From the above statement and applying assumption A7, we can deduce

CSP Can Prove (EHR Sent To MSP). \hspace{1cm} (iv)

Combining (iii), (iv), we get
CSP Can Prove \((\text{EHR Sent To MSP}) \land \) 

\((\text{MSP Received } \text{H} \ (\text{EHR})).\)

From the above statement and applying assumption A8, we can infer

CSP Can Prove (MSP Receivced EHR). \hspace{1cm} (G4)

When CSP receives message 10 from TTP, CSP can use it as an evidence of delivering the requested medical data to MSP so that MSP can provide the emergency medical service to the patient. Message 10 is considered a proof of response delivery (PoRESD). So message 10 stands as a non-repudiation of delivery.

6. Performance analysis and limitations

There are occasions in which things do not follow the predefined (legitimate) course of actions. This is what we call malicious. A malicious behavior could be, but not limited to, any of the following: a party or more are unwilling to participate, they are manipulating their feedback (inputs) and they abruptly abort or cease participating. A protocol design should be flexible to handle this. One measure is to set a time frame for each response and expect loss and/or corruption of messages. Along with keeping a log of misbehaving parties to exclude them from the multi-party computation model as they are devious.

Input size:

- Client’s set has 1 item only
- TTP’s has a population of size w that belongs to n HCPs.

Computational complexity:

- The TTP runs \(O(w)\), where \(w\) is the population size, comparisons with its database against client’s C. It is apparent that the computation is linear w.r.t. the input size at the server.

Communication complexity:

- From participating providers: \(O(w)\)
- From client seeking membership: \(O(1)\)
- Total: \(O(w + 1)\)

7. conclusion

For long, officials had to barter information utility for privacy and vice versa. In other words, adopting strict privacy guidelines means a sacrifice of information utility. In this work, we deal with sensitive information that improper handling could compromise it and hence identity theft may befall. However, due to the implications and pandemics, we have to propose models that permit the sharing of patients’ medical records among health-care providers (HCP) without compromising patients’ privacy. The proposed framework is built upon the existence of a trusted party that manages and authenticates all correspondences among participating HCPs. The framework adds two security services, i.e., privacy and non-repudiation. We believe that EHRs contain sensitive information and sharing the right part in a timely fashion without compromising a patient’s privacy could save lives. We adopt Bloom filter (BF) as a secure probabilistic data-structure. Custodians may publish its BF on a trusted third party allowing other HCPs to seek a visiting
patient’s medical data. Future work may rely on the usage of Blockchain and utilize its infrastructure to build a secure health-care chain. Also, a secure framework that can perform without a trusted third party is a potential area of study.

Author details
Mohamed Sharaf 1,2
E-mail: masharaf@ju.edu.sa
ORCID ID: http://orcid.org/0000-0002-0439-5981
1 Computer and Information Sciences, Jofu University, Saudi Arabia.
2 Systems and Computers Eng. Dept, Al-Azhar University, Cairo, Egypt.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
T. This work is supported by Jofu University.

Citation information
Cite this article as: Non-repudiation and privacy-preserving sharing of electronic health records, Mohamed Sharaf, Cogent Engineering (2022), 9: 2034374.

References
Almeida, P. S., Baquero, C., Preguiça, N., & Hutchison, D. (2007). Scalable bloom filters. *Information Processing Letters*, 101(6), 255–261. http://www.sciencedirect.com/science/article/pii/S0020019006001217
Atieniese, G., De Cristofaro, E., & Tsudik, G. (2011). “[ff] size matters: Size-hiding private set intersection.” In *Public Key Cryptography – PKC 2011* (pp. 156173). (156173). (D. Catalano, N. Fazio, R. Gennaro, & A. Nicolosi, Eds.), Springer Berlin Heidelberg.
Bier, E., Chow, R., Golle, P., King, T. H., & Staddon, J. (2009). The rules of redaction: Identify, protect, review (and repeat). *IEEE Security Privacy*, 7(6), 46–53. https://doi.org/10.1109/MSP.2009.183
Bleumer, G. (2011). Chaffing and winnowing (pp. 197–198). Springer US.
Bloom, B. H. (1970). Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7), 422–426. https://doi.org/10.1145/362866.362692
Bose, P., Guo, H., Kranakis, E., Maheshwari, A., Marin, P., Morrison, J., Smid, M., & Tang, Y. (2008). On the false-positive rate of bloom filters. *Information Processing Letters*, 108(4), 210–213. http://www.sciencedirect.com/science/article/pii/S0020019008001579
Broder, A., & Mitzenmacher, M. (2004). Network applications of bloom filters: A survey. *Internet Mathematics*, 1(4), 485–509. https://doi.org/10.1080/15427951.2004.10129096
Churches, T., & Christen, P. (2004). Some methods for blindfolded record linkage. *BMC Medical Informatics and Decision Making*, 4(1) 9 [https://doi.org/10.1186/1472-6947-4-9].
Dong, C., Chen, L., & Wen, Z., “When private set intersection meets big data: An efficient and scalable protocol,” in *In Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. November 4–8, 2013 Berlin, Germany ACM, 2013, pp. 789–800.
Felleki, L., & Sunter, A. (1969). A theory for record linkage. *Journal of the American Statistical Society*, 64(328), 1183–1210. https://doi.org/10.1080/01621459.1969.10501049
Grzebala, P., & Cheatham, M., “Private record linkage: Comparison of selected techniques for name matching,” in *International Semantic Web Conference*. May 2016 9678, Springer, 2016, pp. 593–606 https://doi.org/10.1007/978-3-319-34129-3_36.
Hathalyia, J. J., Tanwar, S., Tyagi, S., & Kumar, N., “Securing electronics healthcare records in healthcare 4.0: A biometric-based approach,” Computers & Electrical Engineering, Elsevier, vol. 76, pp. 398–410, 2019. http://www.sciencedirect.com/science/article/pii/S004579061930062X
Inan, A., Kantarcıoglu, M., Ghinita, G., & Bertino, E., “Private record matching using differential privacy,” in Proc. 13th Int’l Conf. on Extending Database Technology New York, NY, USA (ACM), 2010, pp. 123–134.
Jiang, Y., Wang, C., Wu, Z., Du, X., & Wang, S., “Privacy-preserving biomedical data dissemination via a hybrid approach,” *AMIA Annual Symposium proceedings. AMIA Symposium*, vol. 2018 (American Medical Information Association (AMIA)), pp. 1176–1185, 12 2018.
Kailar, R. (1996). Accountability in electronic commerce protocols. *IEEE Transactions on Software Engineering*, 22(5), 313–328. https://doi.org/10.1109/32.502224
Karacasidís, A., & Verykios, V., “Privacy preserving record linkage using phonetic codes,” in Proc. 4th Balkan Conf. in Informatics, 17-19 Sep. 2009 (4th Balkan Conf. in Informatics) Thessalonoki, Greece, 2009, pp. 101–106.
Kuzu, M., Kantarcıoglu, M., Durham, E., & Malin, B., “A constraint satisfaction crypanalysis of bloom filters in private record linkage,” in Proc. 11th Privacy Enhancing Technologies Symp. July 27–29, 2011 Waterloo, ON, Canada, 2011, pp. 226–245.
McCallister, E., Grance, T., & Scarfone, K. (2010). Guide to protecting the confidentiality of personally identifiable information (PII). The United States Department of Commerce Gaithersburg, MD 20899-8930, NIST Special Publication 800-122 March 2022 https://nvlpubs.nist.gov/nistpubs/Legacy/sp/nistspecialpublication800-122.pdf.
Meskanen, T., Liu, J., Ramezanian, S., & Niemi, P., “Private membership test for bloom filters,” in *Trustcom/ BigDataSE/ISPA*, 2015 IEEE, August 20–22, 2015, pp. 515–522.
Porah, S., Ahad, F., Sheikh, J., Loon, N., & Bhat, G. (2017). A new reversible and high capacity data hiding technique for e-healthcare applications. *Multimedia Tools and Applications*, 76(3), 3943–3975. https://doi.org/10.1007/s11042-016-4196-2.
Ramezanian, S., Meskanen, T., Naderpour, M., Junnila, V., & Niemi, V., “Private membership test protocol with low communication complexity,” *Digital Communications and Networks*, 2019. http://www.sciencedirect.com/science/article/pii/S2352866818302670. https://doi.org/10.1016/j.dcan.2019.05.002
Randall, S. M., Ferrante, A. M., Boyd, J. H., Bauer, J. K., & Semmens, J. B. (2014). Privacy-preserving record linkage on large real world datasets. *Journal of Biomedical Informatics*, 50, 205–212. (Special issue on informatics methods in medical privacy). http://www.sciencedirect.com/science/article/pii/S1532046413001949
Sanjog, J., & Vigneron, L. (2007). Optimistic non-repudiation protocol analysis. In D. Sauveron, K. Markantonakis, A. Bilas, & J.-J. Quisquater (Eds.), *Information security theory and practices. smart cards, mobile and ubiquitous computing systems* (pp. 90–101). Springer Berlin Heidelberg.
Sarosh, P., Parah, S. A., & Bhat, G. M. (2021). Utilization of secret sharing technology for secure communication: A state-of-the-art review. Multimedia Tools and Applications, 80(1), 517–541. https://doi.org/10.1007/s11042-020-09723-7

Schnell, R., Bachteler, T., & Reiher, J. (2009). Privacy-preserving record linkage using bloom filters. BMC Medical Informatics and Decision Making, 9(1), 41. https://doi.org/10.1186/1472-6947-9-41

Schnell, R., & Borgs, C., “Randomized response and balanced bloom filters for privacy preserving record linkage,” 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) Barcelona, Spain (IEEE), pp. 218–224, 2016.

Tamrakar, S., Liu, J., Paverd, A., Ekberg, J.-E., Pinkas, B., & Asokan, N., “The circle game,” Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security April 2-6, 2017 (ACM) Abu Dhabi, United Arab Emirates, April 2017. https://doi.org/10.1145/3052973.3053006

Union, E. (2016) Regulation (EU) 2016/679 of the European Parliament, “General Data Protection Regulation,” 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

United States, “Health insurance portability and accountability act of 1996,” pp. 104–191, 1996, US Statut Large. 1996 Aug 21;110:1936-2103.

Vatsalan, D., Christen, P., & Verykios, V. S. (2013, September). A taxonomy of privacy-preserving record linkage techniques. Information Systems, 6(6), 946–969. https://doi.org/10.1016/jjis.2012.11.005
