# Confidentiality Preserving Audits of Electronic Medical Record Access

**Bradley Malin**[a], **Edoardo Airoldi**[b]

[a]*Department of Biomedical Informatics, School of Medicine, Vanderbilt University, United States*
[b]*Lewis-Sigler Institute for Integrative Genomics & Department of Computer Science, Princeton University, United States*

**Abstract**

*Failure to supply a care provider with timely access to a patient's medical record can lead to patient harm or death. As such, healthcare organizations often endow care providers with broad access privileges to electronic medical record (EMR) systems. In doing so, however, care providers may access a patient's record without legitimate purpose and violate patient privacy. Healthcare privacy officials use EMR access logs to investigate potential violations. The typical log is limited in its information, so that it is often necessary to merge access logs with other information systems. The problem with this practice is that sensitive information about patients and care providers may be disclosed in the process.*

*In this paper, we present a privacy preserving technique that enables linkage of disparate health information systems without revealing sensitive information. The technique permits any number of vested parties to contribute to audit investigations without learning information about those being investigated. We motivate the protocol in a real world medical center and then generalize the protocol for implementation in existing healthcare environments.*

***Keywords:*** *Privacy, Confidentiality, Computer Security, Electronic Medical Records Systems, Medical Record Linkage*

## Introduction

In the mid-1990's the National Research Council of the United States concluded that electronic medical record (EMR) systems increase the potential for the inappropriate disclosure of a patient's health information to parties both inside and outside of healthcare organizations. [1] The commission recommended that healthcare organizations design and adopt policies, as well as technologies, to prevent and address intrusions. Following recommendations from the NRC and other studies, such as [2], state and federal regulations were enacted in the United States to regulate the transmission, privacy, and security, of electronic personal health information. [3, 4] Regulations safeguarding medical information have been enacted by many other countries as well, including member states of the European Union [5], Japan [6], and Australia [7].

Technological protections for personal medical information have lagged behind policy ratification, which is due in part to the complexity of the healthcare environment. Many off-the-shelf EMR systems are equipped with role-based access control (RBAC) [8], a common policy requirement. However, RBAC is rarely enforced at point-of-care because a lack of medical record availability can cause patient harm or death. Rather, hospitals tend to use a "break the glass" model: they endow care providers with broad access privileges, but stress that harsh punishments will result for system misuse. [9] Nonetheless, unauthorized accesses occur. For instance, since 2002, the University of California, Davis has fired at least six employees, demoted one, suspended one without pay, and retrained eighty for inappropriate accesses. [10]

System misuse is discovered through audits of medical records access logs. A recent survey found that 28 of 28 EMR systems incorporate audit capability. [8] Yet, only 10 of the systems alert healthcare administrators of potential violations. A principle challenge is that EMR systems do not always contain the necessary information to characterize violations. Oftentimes, it is necessary to gather information from other information systems. However, these systems can be controlled by different facets of the organization with diverse privileges and proprietary knowledge.

### An EMR Audit Paradox

Consider the following example. The Vanderbilt University Medical Center (VUMC) is a large healthcare organization with a centralized EMR system. The EMR access log documents each medical record viewed, including a timestamp, the login of the care provider, and the medical record number (MRN) accessed. It is the role of the VUMC Privacy Office to audit the EMR access logs for inappropriate accesses.

A particular type of access privacy officials look for is pre-existing relationships between care providers and patients; e.g., "Are the care provider and patient coworkers at the university?" Such information is not in the EMR, but is in the university's human resources (HR) knowledgebase. Yet, no primary key exists between the EMR and HR databases. Thus, the common attributes are personal identifiers, such as personal names, demographics, or Social Security Number must be used to link the systems. However, the execution of a database join on patient's identifiers, as shown in Figure 1, would

reveal patient-specific information to human resources administrators, including information on patients that are not university employees. This investigation, though performed for surveillance purposes to detect wrongs committed against patients can violate patient privacy.

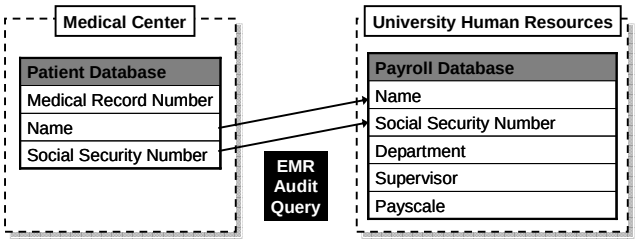


*Figure 1- A query across organizational boundaries can violate patient privacy.*

How can medical privacy officials determine if a patient is an employee without revealing patient identifiers?

### Contributions of this Research

In this paper, we propose a novel protocol called CAMRA (Confidential Audits of Medical Record Access) that allows an auditor to access information from non-EMR systems without revealing the identity of those being investigated. CAMRA leverages a secure protocol in which information is encrypted and stored at a third party. The EMR auditor sends queries of encrypted information to the third party to gather information from disparate systems for its investigation. The end result is that the identities of those being investigated are never provided beyond the EMR system. CAMRA simultaneously enhances confidentiality (i.e., no disclosure of person identifiers) and security (i.e., improved audit ability) in EMRs.

## Methods

### Record Linkage and Identity Protection

Record linkage is an activity that is common to a range of biomedical environments. Record linkage within an organization is often performed via person-specific identifiers, such as name, address, or a unique identifying number (e.g. the Social Security Number is a feature frequently used in the United States). However, disclosing identifiers across organizational boundaries is often legally prohibited.

To satisfy legal constraints, biomedical researchers have applied various techniques to link records on entities without revealing identifiers. Secure one-way hash functions were developed and successfully applied for epidemiological follow-up studies [11]. Hash functions convert an individual's identifiers into a pseudonym; e.g., the name *John* becomes 0sad01a. To link information, each record is hashed and matching hashes are linked. A limitation of the one-way hash; however, is that all organizations linking data must share a hash key. The existence of a common key is a security risk because the hashed identifiers are susceptible to a "dictionary attack". For instance, an HR system administrator can hash identifiers until it matches a received hash from the medical center.

This is deemed acceptable risk in France, where the methods have been standardized for use at the Securite des Systemes d'information [12, 13], but in other countries this model has been contested.

To overcome the use of a common key, alternative methods have been proposed. For instance, Berman proposed a commutative random string technique [14]. In this method, each organization generates a string of random characters. The organizations exchange the strings and add them together to generate a set of common strings. Next, the organizations add identifiers (e.g., names, dates, etc.) to the random strings and compare the results.

One-way hashing renders it impossible to recover the original identities. Once records are linked, the organization must sacrifice the knowledge of the corresponding identities. This is adequate when identity is not needed. However, this model is not acceptable for EMR audits. To complete a privacy audit the EMR administrator must append knowledge to the identities of the individuals under investigation.

### Commutative Cryptography for Record Linkage

In this research, we adopt a commutative method. Specifically, we leverage a cryptographic technique called quasi-commutative encryption [15]. Each organization encrypts and decrypts identifiers to satisfy the following commutative property:

$$H(H(John\_Doe, \varepsilon_1), \varepsilon_2) = H(H(John\_Doe, \varepsilon_2), \varepsilon_1)$$

for any ordering of keys $\varepsilon_1, \ldots, \varepsilon_n$ and a function $H$.

A crucial distinction between the proposed method and prior models is that the function we apply can be converted into an asymmetric keyed cryptosystem. In other words, key $\varepsilon_i$ can be paired with a key $\kappa_i$ so that the original identifying value can be recovered using the same function, but different keys. So, organizations can encrypt and decrypt identifiers, such that

$$H(H(H(John\_Doe, \varepsilon_1), \varepsilon_2), \kappa_1), \kappa_2) = John\_Doe.$$

Fortunately, a variant of RSA cryptography satisfies these properties, such that $H(x, y) = x \bmod (n)^y$. RSA is an accepted standard for secure messaging in healthcare systems, so the CAMRA protocol can be built on top of existing healthcare information technology infrastructure. Unlike how RSA is used in practice; however, we define a private-key system, such that no organization discloses any keys. Note, the switch from a public to a private-key system does not require changing infrastructure.

### CAMRA Overview

The CAMRA protocol is designed so that disparate organizations can share encrypted versions of their identifiers with a "semi-trusted" third party. The third party is trusted to correctly execute a record linkage function, but the third party is not trusted to view the original identifiers for which it is performing the linkage. The third party analyzes queries of en-

crypted identifiers from the EMR auditor and responds with either encrypted identifiers linked to relevant information for another organizations information system or "No Link Made". The feedback that is provided to the EMR auditor is encrypted information from which only the auditor, not even the third party, can learn the identifiers.

Let us walk through a basic, high-level, implementation of the CAMRA protocol. In this version, there are two organizations, such as the VUMC and the HR divisions in the example above. Figure 2 provides an illustration of the process.

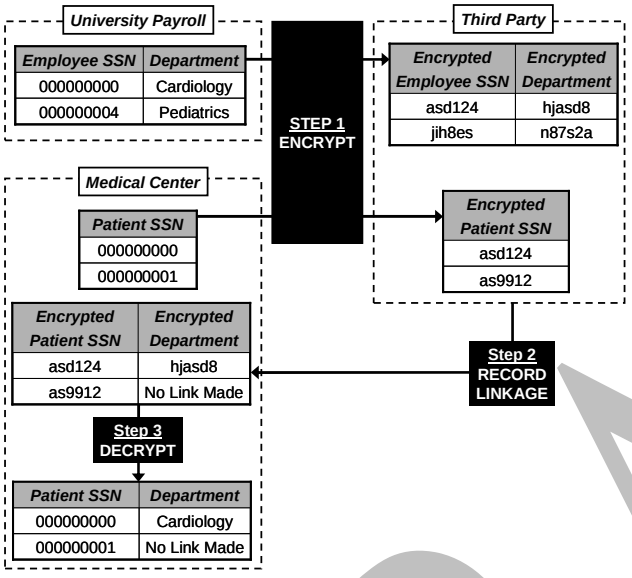


*Figure 2- Enrichment of access logs via the CAMRA protocol.*

**Step 1: Encrypt.** The two organizations commutatively encrypt each other's datasets. Upon completion of this process, the organizations send their encrypted records to the third party. At no point throughout the encryption process can either participant achieve a dictionary attack because the records are not encrypted with the same key. After completion of the encryption phase, all records are comparable because they are encrypted with both of the participants' encryption keys. [16]

**Step 2: Link.** Upon reception of the commutatively encrypted records, the semi-trusted third party performs record linkage on the encrypted identifiers. A list of matches, if any, is produced and sent back to the EMR auditor.

**Step 3: Decrypt**. The EMR auditor initiates a commutative decryption of the linked records. In the end, this provides the auditor with a list of decrypted identifiers that are linked to relevant information from outside of the EMR.

The above overview describes salient features of the proposed methodology. This example illustrates the architecture of the CAMRA protocol, but it obscures the exact ordering and manner by encryption and decryption process.

**The CAMRA Protocol**

We now describe the CAMRA protocol in more detail, as well as show how the protocol generalizes to settings with an arbitrary number of participating organizations.

To implement the commutative encryption and decryption required for CAMRA, each of the participating organizations must encrypt/decrypt each organization's sets of identifiers. To achieve this goal, two technical aspects must be in place. The first is a protocol for how to use maintain and use encryption/decryption keys at each organization. The second is a routing scheme is that specifies the order according to which organizations exchange sets of identifiers to eventually provide the semi-trusted third party with a set of encrypted, comparable records.

**Protocol keys.** Let $D_V$ and $D_R$ be the sets of patient and employee identifiers held by the VUMC and HR, respectively. Similarly, let $\langle \varepsilon_V, \kappa_V \rangle$ and $\langle \varepsilon_R, \kappa_R \rangle$ be the keys for the VUMC and HR, respectively. After commutative encryption the HR has $H(H(D_R, \varepsilon_R), \varepsilon_V)$. Notice, however, that HR is also in possession of $H(D_V, \varepsilon_V)$, which it received from the VUMC. Now, since HR can decrypt commutatively, it can generate $H(H(H(D_R, \varepsilon_R), \varepsilon_V), \kappa_R)$, which results in $H(D_R, \varepsilon_V)$. The consequence is that now the HR can compare its records and the VUMC records as if they were singly hashed by the VUMC: $H(D_R, \varepsilon_V)$ compared to $H(D_V, \varepsilon_V)$. This problem was recognized in [16] and so, to prevent such leaks, it is necessary to use "blinding".

More formally, let $K$ be the set of organizations that share information with the third party. Each organization ($k=1,...,K$) will maintain two pairs of ⟨encryption, decryption⟩ keys,

$$\langle \varepsilon^b_k, \kappa^b_k \rangle \quad \text{and} \quad \langle \varepsilon^m_k, \kappa^m_k \rangle,$$

for an agreed upon quasi-commutative hash function $H$ as defined above. The function $H$ is made public, however, all keys are kept private to each organization. The first key pair is used for "blinding" purposes (superscript $b$) by organization $k$ with its own set of records, akin to the blind signature process defined in the original description of untraceable payment systems [17]. Blinding the records prevents leaks of information as illustrated above. The second key pair corresponds to "multi-party" keys, which each organization uses to encrypt/decrypt each organization's set of records.

**Record Routing.** Alternative routing schemes share the property that each participating organization *starts* and *ends* each round of commutative encryption (initial and final) with its own set of records. In other words, each organization will blind its records before sending them according to a prespecified routing scheme for commutative encryption. The organization will then remove the blinding from its records before sending them to the third party after all other organizations have encrypted it with their multi-party key.

Figure 3 depicts two possible routing schemes. The left panel shows a circular routing procedure based on commutative protocols. This procedure is efficient in terms of communication costs, but it requires that a peer-to-peer architecture be set in place. In contrast, the centralized routing model, shown in the

right panel, are more computationally expensive, but can take advantage of existing network architectures that are designed for high-speed environments with dedicated and trusted communication channels

The best routing option will depend on the features of the organizations involved in the audit. Regardless, the security of the CAMRA protocol does not depend on the specific routing scheme that is implemented.

The complete CAMRA protocol is executed as follows.

**Step 1: Blinding to Encrypt.** Each organization *k* creates a dataset of *dummy* records and adds them to their own set of EMRs, and encrypts the resulting set of records $D_k$ using $\varepsilon^b_k$. After this initial encryption, a blinded dataset $H(D_k, \varepsilon^b_k)$ exists for, and is in the sole possession of, each organization.

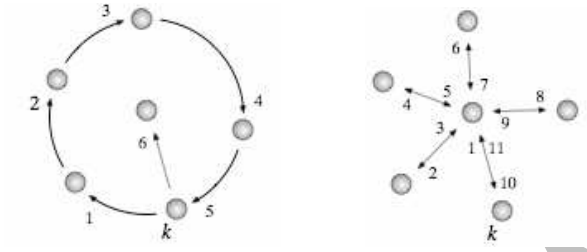


*Figure 3- Two alternative routing schemes from organization k to the third party (in the middle). Nodes are different organizations and numbered edges are routing steps.*

**Step 2: Commutative Encryption.** Each organization encrypts its blinded records with its multi-party key and sends the result to organizations following the pre-specified routing scheme. Each organization encrypts the received sets of records its multi-party key, shuffles the order of the records, and continues to send the records around according to the routing scheme. At the end of this process, each organization *k* is in the possession of

$$H(H(H(\ldots H(H(D_k, \varepsilon^m_1), \varepsilon^m_2) \ldots \varepsilon^m_{K-1}), \varepsilon^m_K), \varepsilon^b_k).$$

Notice that the blinding key $\varepsilon^b_K$ must be removed if the third party is to perform record linkage. This is where commutative decryption is handy. Organization *k* can remove the blinding simply by decrypting with $\kappa^b_k$.

**Step 3: Encrypted Record Linkage.** Each participating organization sends the resulting dataset to the semi-trusted third party, who performs record linkage over the set of encrypted EMRs. The resulting list is sent to the EMR auditor.

**Step 4: Blinding for Decrypt.** Upon reception, the auditor selects a new pair of blinding keys and blinds the list.

**Step 5: Full Decryption.** As in Step 2, the auditor, sends the encrypted list of matches to each organization according to the routing scheme. Now, each organization decrypts the list and sends it back to the auditor. Once every organization has decrypted the list, the auditor decrypts to remove the remove the blinding key.

## Results

The CAMRA protocol is based upon our previous work on collusion resistant protocols [16] and inherits some of its desirable properties. We list them below.

**Collusion-Resistant.** It is possible to show that no set of organizations can successfully collude to learn the identifiers in the encryptions sent by the EMR auditor. Similarly, the EMR auditor can not collude with any set of organizations to learn the contents of another organization's set of records without proceeding through the commutative decryption process.

It is important to note that collusion resistance in CAMRA extends to prevent the third party from colluding with organizations against the EMR auditor. For instance, imagine the third party passes the VUMC records to HR. HR will learn which records the EMR and HR have in common, but will not know which records they correspond to. This is because 1) HR never decrypts records in CAMRA and 2) the VUMC shuffled the encrypted HR records. In doing so, HR can not determine which of its encrypted records corresponds with any of its unencrypted records. The fact that HR can not complete decryption without the assistance of the VUMC is the benefit of using a multiparty authentication mechanism with a single organization that is capable of decryption.

**Detection of Malicious Actions.** Collusion is a primary concern when organizations act according to the specified protocol. However, beyond collusion, there are actions that an organization can take to prevent record linkage from being correctly executed at the third party. For instance, an organization may use a faulty encryption key during the encryption process and the following comparison could be made by the third party:

$$H(H(John, \varepsilon^m_{VUMC}), \varepsilon^m_{HR}) \neq H(H(John, \varepsilon^{bad}_{HR}), \varepsilon^m_{VUMC})$$

In this case, HR correctly used $\varepsilon^m_{HR}$ with the name John from the first set of records (on the left), but inserted a "bad" key into the commutative encryption process to encrypt John in the second set of records (on the right). As a result, the names are not properly linked.

Fortunately, the CAMRA protocol can be extended to detect such malicious behaviors both during the encryption process and during the decryption process. These extensions are described in [16] and are important when trust between organizations is low.

**Scalable.** The CAMRA architecture is extendible in several ways. First, the security protocol is not limited by the number of organizations that are involved in the investigation. To increase the number of organizations that contribute to an audit, we only need an additional set of encryption/decryption keys. Second, CAMRA can be augmented so that each participating organization is provided with differential responses. To do so, the third party can send a response list to each organization. Each organization can use its own set of blinding keys to perform decryption. Yet, this extension must be performed

with caution as it violates collusion resistance property (i.e., the third party can collude with any organization).

## Discussion

The CAMRA protocol illustrates that privacy does not have to be sacrificed in healthcare operations. The protocol prevents the disclosure of identifiers during the audit process. Nonetheless, the protocol has certain limitations that we now address.

### Towards a Fault Tolerant Model

Typographical errors and variation of personal information are a part of healthcare. For instance, an entity's name may be "Jon" in the EMR, but "Jonathon" at HR. Traditional record linkage methods account for such issues through string comparators and probabilistic matching algorithms [18]. Unfortunately, hashes and encrypted and encrypted versions of identifiers do not retain their similarities. As a consequence, linking encrypted identifiers can trigger false matches and non-matches.

Recently, methods have been proposed to measure the similarity between strings in an encrypted environment [19]. However, existing methods do not scale for use with CAMRA. This is because these methods are based on protocols that are designed for use between two organizations. Moreover, due to the cryptographic basis of these methods, they do not scale beyond two organizations. Yet, the CAMRA protocol is applicable to environments in which two or more organizations are involved. The development of scalable record linkage algorithms for encrypted data will limit the false non-linkage rate in CAMRA.

### Third Parties in Healthcare

In the realm of healthcare, third parties are leveraged for a variety of purposes, including data warehousing, data aggregation, and brokering. Yet, from a data security perspective, the less number of organizations that handle data the better. As such, a more attractive possibility is to remove the third party from EMR auditing. Research in computational theory has shown that third parties can be removed from cryptographic protocols without sacrificing the level of security. However, the application of such theory in the real world is limited because resulting protocols are computationally burdensome and inefficient for daily practice. Nonetheless, minimal information sharing is the key to maintaining privacy in healthcare systems. We intend on developing confidential EMR audits models that limit the involvement of third parties.

### Acknowledgments

The authors would like to thank Professors Dario Giuse and Dan Masys of the Department of Biomedical Informatics, and Gaye Smith at the Privacy Office of the Vanderbilt University Medical Center, for helpful discussions during this research.

## References

[1] National Research Council. For the record: protecting electronic health information. Washington, DC: National Academy Press. 1997.

[2] Safran C, Rind D, Citroen M, Bakker AR, Slack WV, and Bleich HL. Protection of confidentiality in the computer-based patient record. MD Computing 1995: 12: 187-192.

[3] U.S. Department of Health and Human Services, Office for Civil Rights. Standards for protection of electronic health information; Final Rule. Federal Register, 2003 Feb 20; 45 CFR: Part 164.

[4] U.S. Department of Health and Human Services. Standards for privacy of individually identifiable health information; Final Rule. Federal Register, 2002 Aug 12; 45 CFR: Parts 160-164.

[5] European Commission. Directive 95/46/EC: the protection of individuals with regard to the processing of personal data and on the free movement of such data. 24 Oct 1995.

[6] Diet of Japan. Law on the protection of personal information. 23 May 2003.

[7] Office of the Federal Privacy Commissioner, Commonwealth of Australia. The commonwealth privacy amendment (private sector) act. 2001 Dec 21.

[8] Rehm S and Kraft S. Electronic medical records: the FPM vendor survey. Fam Pract Manag. 2001: 8(1): 45–54.

[9] Ferreira A, Cruz-Correia R, Antunes L, Farinha P, Oliveira-Palhares E, Chadwick DW, and Costa-Pereira A. How to break access control in a controlled manner. Proc IEEE CBMS 2006: pp. 847-54.

[10] Youngstrom N. Nosy employees are a risk, require a wide range of remedies. Report on Patient Privacy, Atlantic Information Services. August 2005: 5(8).

[11] Bouzelat H, Quantin C, and Dusserre L. Extraction and anonymity protocol of medical file. Proc AMIA Symp. 1996: pp.323-7.

[12] Quantin C, Kerkri E, Allaert FA, Bouzelat H, and Dusserre L. Security aspects of medical file regrouping for the epidemiological follow-up. Medinfo 1998: 9 (Pt 2): 1135-7.

[13] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, and Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. Meth Inf Med 1998: 37(3): 271-7.

[14] Berman JJ. Zero-check: a zero-knowledge protocol for reconciling patient identities across institutions. Arch Pathol Lab Med 2004; 128(3): 344-6.

[15] Benaloh J and deMare M. One-way accumulators: a decentralized alternative to digital signatures. LNCS 765: Proc Eurocrypt '93. 1994: 274-86.

[16]Malin B, Airoldi E, Edoho-Eket S, and Li Y. Configurable security protocols for multi-party data analysis with malicious participants. Proc IEEE ICDE 2005: pp. 533-44.

[17]Chaum D. Blind signatures for untraceable payments. Advances in Cryptography, Crypto 1982. Plenum Press. 1983: pp. 199-203.

[18]Grannis S, Overhage M, and McDonald C. Real world performance of approximate string comparators for use in patient matching. Medinfo 2004: 11 (Pt 1): p. 43-7.

[19]Churches T and Christen P. Some methods for blindfolded record linkage. BMC Med Inform Decis Mak 2004: 4: 9.

**Address for correspondence**

Bradley Malin, Vanderbilt University, Department of Biomedical Informatics, Eskind Biomedical Library, Fourth Floor, 2209 Garland Avenue, Nashville, TN 37232 USA.