careevolution
HEALTHCARE TECHNOLOGY

Security & Privacy Mandates for RHIO Architectures

A candid discussion of the RHIO security and privacy needs vis-à-vis current technology options.

RHIO Security Infrastructure

Security and Privacy: CareEvolution's Design Mandate

Exchange of medical information is an extremely sensitive undertaking. While the potential value proposition of a large RHIO is very compelling, the risks assumed by its participants can also be significant if security and privacy considerations are not addressed at every stage of RHIO development. Choosing the right technology platform is the essential first step to ensuring your RHIO offers the most secure solution to protect patient privacy.

The CareEvolution RHIO Technology Platform implements advanced security features in all aspects of its design. The following security-focused design principles drive CareEvolution's architecture:

- Aggregated records managed in a centralized store create a critical security risk.
- CareEvolution believes that a centralized index of *patient demographic information* is essential to provide performant record location services. However, this centralized index must be protected with a cryptographic hash scheme to ensure that patient information cannot compromised by an attack on this central store.
- To eliminate the security and privacy risks of a centralized store of *medical records*, all clinical data should be stored at servers located at participating hospitals and clinics. There must be no centralized database of clinical data.
- Clinical data must be transferred directly from a source clinic or hospital to the site the requires the clinical data. Clinical data must never pass through a centralized entity.
- Communications between entities must be strongly secured with technologies to protect against several potential threats. A secure communication infrastructure built on X.509 Public Key Infrastructure must provide:
  - Identity verification of both the sending and receiving entities
  - Message encryption to ensure data confidentiality.
  - Message signing verifies data origin.
  - Protection against malformed or malicious messages.

Secure Distributed Record Management: The CareEvolution Record Locator Service

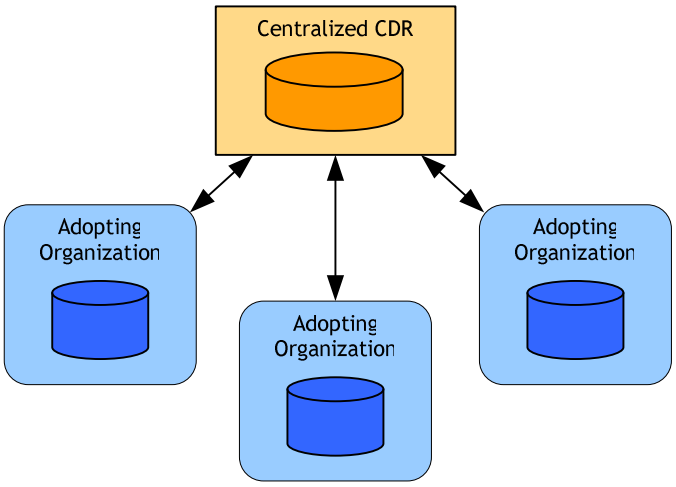
While there are many important security considerations in a RHIO network, determining the appropriate system architecture for record location is the first and most crucial step in building an effective and secure healthcare data exchange solution. A Record Location Service (RLS) is the actor in an RHIO network that determines which patients have records at multiple institutions

and therefore where clinical information may be located. An RLS has all of the complexities of an MPI solution: it must first ensure that a given institution's patient database is free of duplicates or to identify those duplicates. It must then communicate with the network to find matching records from every other institution. Similar to an MPI, the RLS record linking must be optimized for near-100% specificity to prevent the possibility of incorrectly associating clinical data with patients. An RLS also has all the complexities of a general purpose record linking product in that it must have acceptably high sensitivity to find the appropriate records at each site, and to automate the linking process to keep system maintenance costs low. Finally, Record Location Services must be secure as the information flowing though them is highly sensitive.

Proposed RHIO Architectures

Because record location is such a fundamental consideration in RHIO security, any proposed architecture will necessarily be influenced by the requirements for record location. Information must be made available so that the RLS can compare records for linkage. However, this information must be strongly secured to protect against security breaches during the record linking process.

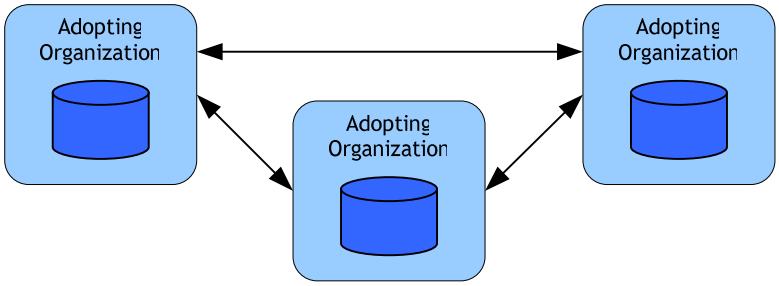
Centralized CDR



An attractive architecture, because of its simplicity, is to create a central repository where institutions replicate their records. This central actor could identify linked records and share the applicable clinical information. While the centralized repository model can achieve the data sharing requirements of an RHIO, it is seriously flawed from a security and privacy perspective. A large scale aggregated store of clinical and demographic information would be a privacy liability, a primary target for malicious hackers, and an opportunity for disgruntled employees. Even if the information was not aggregated centrally, the "everyone trusts the hub" paradigm means that an attacker that can

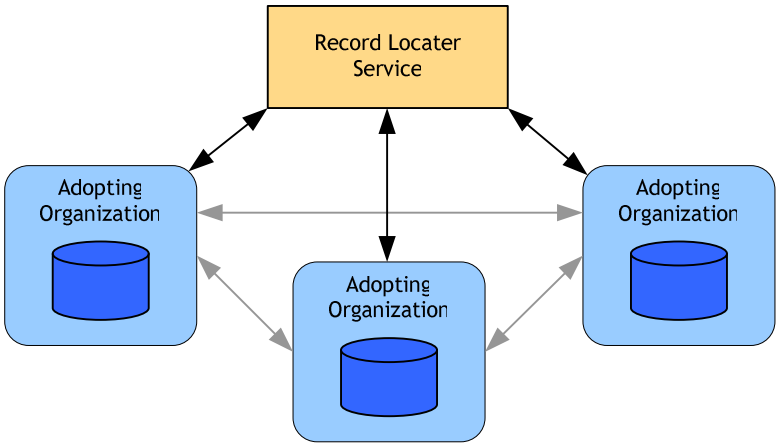
compromise the hub will have unfettered access to every institution's information. Performance in this scheme could also potentially suffer. A system that centrally aggregates each institution's record database will grow in size quickly. As the number of participating institutions grows, not only will the database grow in size, but the number of transactions per second will increase as well. These factors would make the management of such a system very costly in a large RHIO. A similar but improved implementation would create references to documents at each site rather than replicating them centrally. However, this pointer based scheme still suffers from the same security concerns because of the trust relationships.

Peer-To-Peer



At the other extreme is a completely peer-to-peer network where the RLS function is distributed across each institution in the system. Each institution makes its demographic information available and participates in record linking. This approach suffers from adverse performance / scalability implications, as well as from reliability and security considerations. As the number of institutions in the network grows so does the number of queries that must be issued to determine record links. Since each node must contact every other node to establish links, each new node that comes online will exponentially increase the amount of network traffic. Reliability is also an issue: if a node goes offline for any reason, links from that node will be missed. As the number of nodes increases, the likelihood of one requiring downtime increases and so does the possibility of missed links.

The NHIN Model



Not surprisingly, due to system performance the industry has largely rejected the above architectures in favor of an architecture that centralizes demographic information and record locations but that shares clinical information in a peer-to-peer fashion. The nonprofit consortium Connecting For Health has suggested such a model in their response to the ONCHIT RFI for a National Information Health Network (NHIN) [1]. Performance issues are lessened in this model because aggregating demographic information is less costly than aggregating clinical data and having a central demographic store prevents the exponential growth in network traffic. The system is also more secure because institutions never share clinical information with the central RLS and therefore compromising the RLS becomes less lucrative. The network can be designed to only allow clinical information to flow between institutions for which the records have been properly linked by the RLS. Thus, compromising an institution only allows access to patient records that have already been linked from other sites, giving the attacker only marginally more clinical information than is possible without any RHIO infrastructure.

A Better Model is Required

Even with these improvements in security, having a centralized store of demographic information (and potentially encounter information) is still an enormous privacy liability. A centralized, potentially national database containing names, addresses, phone numbers, SSNs, DOBs, MRNs and health record location information is simply too large a target to be a viable long term solution.

Securing Identity Management - The Privacy Mandate

RLS - The Traditional Weak Link in the RHIO Security Chain

As we have discussed earlier, we believe that the centralized store of accessible demographic information employed by most RHIO implementations creates an unacceptable security risk for any RHIO. Data aggregation accentuates three critical risk factors that increase the potential that sensitive information will be improperly disclosed:

- First, data aggregation increases the value of the centralized store creating a lucrative target for potential attackers.
- Second, it increases the number of entities that legitimately should have access to the central store; this in turn increases the number of avenues that can be compromised by attackers.
- Third, a centralized store of sensitive data can become a valuable resource that may be susceptible to political pressure for legalized access by interests claiming a need to know. A concerted effort by the government to obtain data from the large Internet search engines is a compelling example of this third risk factor.

Blinded Record Linkage – The Solution

Methods must be deployed that can strongly secure this centralized data store. The CareEvolution RHIO Technology Platform provides a solution for this challenge. The CareEvolution RTP achieves a secure, performant solution to record linkage in the distributed system by using a **blinded directory** for centralized demographic data used in record location. A set of techniques are implemented to cryptographically (one-way) hash any demographic data that will be aggregated centrally. This ensures that patient demographic data stored in the centralized index is unrecoverable. There are two direct results of hashing the centralized index :

- **World Class Security -** From any plaintext string (i.e. “Smith”), a one-way hashing algorithm can quickly produce a long sequence of numbers (a “hash”), which represents the string “Smith”. However, to take this hash and reverse the algorithm to arrive at “Smith” would require years of computation, hence the term “one-way” hash.

- **Record Linking Challenge** – Since hashes of similar strings, such as “Smith” and “Smit” yield drastically different number sequences, the very process of hashing renders the traditional preferred probabilistic record linking techniques inoperable. As a result, all contemporary providers of MPI or Identity management solutions have avoided the formidable technical challenges post by a crypto-hashed central directory. While this may have been acceptable when such solutions were intended to be implemented behind the security firewalls within an institution, **we believe that extending a non-hashed centralized repository of demographic**

information across a region, let alone the country, poses an unprecedented and unwarranted privacy risk.

There is a solution, though it is not technically straightforward. Sophisticated string processing techniques are available that allow for both the security of one-way hashing and effective probabilistic matching. Approximate matching in this scheme is accomplished using a technique called bigramming. Bigramming breaks up the source string into many derived strings. Each derived string is given a similarity score that indicates how similar it is to the source. Two strings that have been bigrammed can then be compared by determining if they share a derived string. If so, the two derived similarity scores can be used to compute an overall "dice score." Using a bigramming technique to generated derived strings, and then hashing derived strings allows for approximate, blinded identifier matching. This is the technique employed by the CareEvolution Crpto-Record Locator Service (Crypto-RLS)

Crypto-RLS – The CareEvolution Implementation

In summary, using the CareEvolution Crypto-RLS provides an unprecedented privacy guarantee than any two-way encryption scheme because the underlying data cannot be decoded – even by the operators of the RHIO itself. Fortunately, once the underlying store of patient demographic information is no longer at risk, a centralized model for record location can actually **increase** security of the system. While the data itself cannot be compromised, record location requests from around the network can be monitored and audited for suspicious patterns of requests or other inappropriate activity.

Securing the Network

Blinding records before sending them to the Record Locator Service cryptographically protects the transmitted information. However, clinical information on the network, sent peer-to-peer, cannot use a one-way hash; the plaintext must be decipherable to the recipient. Thus, a general encryption mechanism is required. Indeed, even for blinded information a comprehensive security framework is required that will establish trust relationships and provide authentication services.

Design Principles

Broadly, security considerations can be grouped into message level protection and service level protection. They include:

- Message level protection
  - Identity verification
  - Data confidentiality
  - Data origin
  - Message integrity

- Service level protection
  - Protection from malformed or malicious messages
  - Shielding exceptions from revealing sensitive implementation details
  - Replay protection
  - Audit logging

The CareEvolution RHIO Technology Platform's messaging and security framework adhere to Web Service (WS-*) standards. The WS-* specifications are developed through industry collaboration and published by the standards body OASIS to provide secure, platform independent messaging. These standards are used as building blocks to achieve the security objectives outlined above.

Message Level Protection

Because of the dynamic, peer-to-peer requirements of the network, the platform implements a brokered authentication mechanism using Public Key Infrastructure (PKI). This trust brokerage service, also known as a Security Token Service (STS), allows messages to carry identity information. Using PKI and X.509 certificates, it's also possible to encrypt and sign messages, providing data origin, data confidentiality, and message integrity.

Service Level Protection

As evidenced by the increasing trend toward software update services and security hot-fixes, securing services from all possible attack is challenging. While deep knowledge of the nature of network attacks can help guide security implementation choices, the best approach is simply to use well understood industry infrastructure rather than homegrown solutions. The CareEvolution RHIO Technology Platform leverages well supported security libraries to provide replay protection, exception shielding, and malicious message protection. Then, as a final step, audit logging is implemented at each node in the system to preserve appropriate authentication, authorization, and data flow information. With this information system security can be verified on an on going basis.

Summary

A viable RHIO platform requires that many complex systems participate in order to accurately and securely share information. A primary requirement to enable inter-hospital data sharing is a Record Locator Service. The NHIN has proposed an architecture that makes reasonable performance and reliability tradeoffs. However the central aggregation of plaintext demographic information and record locations presents a privacy and security risk. The CareEvolution RHIO Technology Platform addresses this security shortcoming by

only centralizing hashed identifiers and using blinded record linkage techniques to identify matching records. For peer-to-peer clinical information flow, industry standard X.509 signing and encryption are employed to enable institutions to safely exchange data over the public internet. With a sensible architecture, advanced security features, and a state-of-the-art record linking pipeline, the CareEvolution RHIO Technology Platform is the right platform to enable the seamless flow of information between institutions, regions, and the nation.

About CareEvolution, Inc.

CareEvolution is a leading provider of secure interoperability solutions. Our RHIO Technology Platform is a robust Service Oriented Architecture (SOA) to enable RHIOs' heterogeneous underlying EMRs to "share" clinical information in a secure, reliable, and incremental manner. Distinct component such as Identity Management, Record Location, Clinical Data Integration, Audit & Log, Data Persistence, Visualization, Terminology, and Data Mining may be adopted piecemeal or as a comprehensive technology platform.

For More information please visit us at www.careevolution.com