Enhancing Security and Privacy of Healthcare Data using XML Schema

Ruchika Asija
School of Computer and Information Sciences (SOCIS)
Indira Gandhi National Open University (IGNOU)
New Delhi, India

ABSTRACT

Information security and privacy in the health care sector is an issue of growing importance. Widespread use of digital data in health care industry has provided potentially immeasurable benefits by instant access to patient information practically from anywhere in the world. Connecting HIS to the network and making EHR available over the Internet put the data vulnerable to security threats and risk the privacy of patients. For the storage and exchange of health care data, Extensible Markup Language (XML) is widely used. This paper presents a survey on security and privacy of health care data and details about the paper-less hospitals, risks to the privacy and security of health data due to digitization. Further the paper presents XML and its security technologies for the security and privacy of health care data. But these are having many challenges. Dealing with those challenges, the paper presents how XML Schema exploiting the capabilities of XML can be used to enhance the security and privacy of health care data.

General Terms:

Security and Privacy in health care

Keywords:

XML, XML Schema, XML Security, Paper-less hospitals

1. INTRODUCTION

Often voluminous, heterogeneous, unstructured, lacking standardized or canonical form, and incomplete, as well as, surrounded by ethical considerations and legal constraints, the characteristics of patient healthcare records make them "messy". Because they originate primarily as a consequence of direct patient care with the presumption of benefit for the patient, their use for research or administrative purposes must happen with care to ensure no harm to the patient. Radiologic images, lab test results, medications, allergies and other clinical information are stored and viewed on computers. Inappropriate disclosure, loss of data integrity or unavailability may each cause harm. Sometimes the information needs to be accessed for physicians to be able to make the best decisions about patient care. To provide effective healthcare services in an efficient manner, Health Information Systems (HIS) [16] are helping healthcare organizations by quickly adopting the enormous advances in information and communication technologies (ICT). Anywhere access to health care data using devices with wired and wireless connections and sharing it across heterogeneous networks have helped in the rapid delivery of healthcare services. HIS and Electronic patient records have changed the face of healthcare being very effective in local and global efforts to collect, process, and use healthcare data to influence policies and decision making, programme action, individual and public health outcomes and research. Paper-based records are being converted to electronic format in healthcare organizations. A report published by Price Waterhouse Coopers [11] shows that electronic patient records could improve care and allow health professional to spend more time with patients and save billions of dollars. Many hospitals are converting to paperless form, which involves substitution of clipboards at the end of patients' bed with bedside observation recorded on handheld devices like smart phones and tablets by nurses. Doctors use mobile computers for sending the prescription to the pharmacy (E-prescription), clinicians order and review test electronically, monitor vital signs digitally while automatically recording information. But with these automations, protection of security and privacy of healthcare data has become a big challenge. This calls for a comprehensive security and privacy framework for health information systems. For the exchange and storage of healthcare data, Extensible Markup Language (XML) is widely used. XML is the de facto standard for electronic data exchange across the world. Health care data is no exception and is used for describing the structure of health data and content on the Internet. Some of the well-recognized benefits of using XML include its richness of the data structure, simplicity and excellent handling of international characters. But, XML documents have to travel across untrusted networks and has the potential of being manipulated by external systems, which bring unique challenges for health care data security. However, there are cases where data must be protected from possible threats since the data may contain confidential information or to prevent repudiation. Therefore, it is necessary to combine data representation like XML used in health care data with various security features. XML is enhanced with a sophisticated access control mechanism that allows not only to securely browse healthcare XML documents but also to securely update each document element. Several technologies can be used to achieve XML data security. These technologies include XML encryption [7] for confidentiality of health data, XML Digital Signature [18] for integrity and signing solutions, XML Key Manage-

ment (XKMS) [8] for public key registration, location and validation, Security Assertion Markup Language (SAML) [6] for conveying authorization, authentication and attribute assertions, XML Access Control Markup Language (XACML) [22] for defining access control rules, and platform for Privacy Preference (P3P) for defining privacy policies and preferences. Major use cases include securing web services (WS-Security) and Digital Rights Management (eXtensible Rights Markup Language). The growing digitization, risks due to these digitizations, and use of XML Schema to address these risks are discussed in this paper.
The rest of the paper is organized as follows: The switching from paper to paperless technology is useful in lowering costs and laid the foundation of legal health record as described in Section 2, but it has increased the probability of increasing data breaches. The digital data can be in any form as detailed in Section 3. XML and its security technologies can be integrated with health care data to enhance the security and privacy of health care data given in Section 4 but these technologies are having many challenges while their implementation is done. To deal with such issues, XML Schema is presented using its attributes to enhance the security and privacy of health care data as described in Section 5 and conclusions are given in Section 6.

2. PAPERLESS HOSPITALS - A DIGITIZED APPROACH

Several hospitals have started the process of moving their paper based medical records to EHR [14] thus becoming digitized. EHR is a repository of various electronic medical records containing records like patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports. This incorporation of technology into healthcare system specifically through the use of electronic health records is optimizing the management of medical information. Increased Medicare and Medicaid reimbursements for medical providers to establish ”meaningful use” of electronic health records have further spurred the conversion from paper to electronic records by healthcare practices and facilities. With the sophistication of modern technology, the switch from paper to paperless environment is useful in lowering costs and hours wasted searching through paper documents for patient records and alleviate human oversight and error. The EHR consists of many components that work together to create the foundation of the legal health record. These components may include software applications such as computerized physician order entry, integration with laboratory, radiology and cardiology systems or an electronic document management system. Main benefits of paperless health records are-

(1) Improve quality of Patient care - Document scanning and management systems have a search function which allow users to sort through documents in seconds. Instead of flipping medical charts we can find what we are looking for instantly which leads to less time spent on pulling and filing medical charts and more time can be spent interacting with and treating the patients.
(2) Increase convenience and office efficiency - If all patient charts and documents are retrieved in seconds, it will improve the convenience and efficiency of the healthcare office. Doctors, nurses, clerks and other staff can have quick access to records even from remote locations.
(3) Improve care coordination - With electronic health records, we can quickly transfer access to files for other departments removing the time and hassle it takes to send a copy of the physical medical chart or documents thereby increasing the inetroperability between multiple hospitals or offices.
(4) Increased space in growing office - Having all the medical records stored digitally cuts down on the storage required for filing cabinets or carts.

The first paperless hospital in India Seven Hills Group is having two hospitals located in Visakhapatnam and Mumbai. In Mumbai, it is a 1500 bedded multi-specialty tertiary care facility providing technology for diagnostic and therapeutic procedures. The hospital has integrated almost all processes right from the front office registration to doctor consultation, investigations, the complete inpatient treatment, including the collection management. To facilitate the progress towards paperless technology in healthcare, Amrita Hospitals in India, a charitable hospital with 2500 outpatient visits a day and 1200 inpatients, created Amrita Hospital Information System (AHIS) [1], a paperless electronic clinical documentation system to reduce administrative waste and inefficiency. In U.S., hospitals are investing in health information technology to improve the quality, safety and effectiveness of patient care, according to an analysis of HIMSS Analytics' Electronic Medical Record Adoption Model (EMRAM) [4]. EMRAM identifies seven stages of electronic medical record (EMR) capabilities ranging from limited ancillary department systems through a paperless EMR environment. The stages are:

(1) Stage 0 - No ancillaries installed (Lab, Radiology, Pharmacy data output).
(2) Stage 1 - All ancillaries installed (Lab, Radiology, Pharmacy data output online from external service providers).
(3) Stage 2
    (a) Electronic Patient Record
    (b) Controlled medical vocabulary
    (c) Controlled clinical decision support system
    (d) Document imaging and health information exchange (HIE) capability.
(4) Stage 3
    (a) Clinical documentation
    (b) Clinical decision support system
    (c) Picture Archiving and Communications System (PACS) available outside radiology
(5) Stage 4
    (a) Computerized Physician Order Entry (CPOE) in atleast one clinical service area
    (b) May have clinical decision support system based on clinical protocols
(6) Stage 5
    (a) Full compliance of PACS replaces all film based images.
(7) Stage 6
    (a) Clinical decision support system (CDSS) available with stored templates related to clinical protocols triggering variance and compliance alerts.
    (b) Closed loop medication administration
(8) Stage 7
    (a) Complete EMR
    (b) Continuity of Care Document (CCD) transition to share data
    (c) Data warehousing feeding outcome reports
    (d) Data continuity with ED
    (e) Quality assurance
    (f) Business intelligence

In UK, East Kent Hospitals University NHS Foundation Trust [21] has signed a pilot agreement for 500 licenses and working towards a clinical portal solution to create a paperless and mobile clinical environment. It has highlighted few hospitals like the Royal National Orthopedic Hospital, which has conducted a trial of a system that asks spinal surgery patients to record their progress on an iPad while in hospital and then through an online system at home after being discharged. In 2011, more than 80% of the U.S. acute-care hospitals have achieved Stage 5 or Stage 6 of EMRAM and 63% have achieved Stage 7. But, still less than two percent of hospitals in U.S. are paperless. The majority of hospitals, 41.3%, fall in Stage 3. Other paperless hospitals include Mediterranean Institute for Transplantation and Advanced Specialized Therapies (ISMETT) [20]. It was found that 95% of patients preferred the new online process. At King's College Hospital [19], vast majority of staff manage about 80% of their processes through electronic systems. Using iPads and other mobile technologies help organizations to access information whenever and wherever they need it.

3. RISKS TO HEALTH CARE DATA

The automations have led to improved patient care workflow and reduced costs, but it has raised healthcare data to increased probability of security and privacy breaches. Facing increased challenges for the protection of personal health information including growing volumes of electronic health records, new government regulations and a more complex IT security landscape there is a growing need to ensure knowledgeable and credentialed security and privacy practitioners to be in place to protect this sensitive information. Earlier, personal details of people were stored in telephone books. But, now the Internet has opened new avenues, which store almost each and every person's details. Personally identifiable information (PII) is now required by various organizations like health care, schools and universities. The emergence of Internet technology healthcare providers has adopted electronic health records. EHR store all information about patient electronically. This information includes PII. The PII is defined as "any information that can be used to uniquely identify, contact or locate an individual, or can be used with other sources to uniquely identify a person". PII is based on few identifiers like - name, date of birth, email address, social security number (SSN), medical record numbers, health insurance beneficiary numbers, account numbers, biometric identifiers (finger prints, retinal and voice prints), full face photographic images, prescriptions, diagnostic information, X-ray, MRI, CT scan and test reports. But, this adoption of EHRs among health organizations has exposed them to various data breaches. A risk is a function of the probability of a threat agent exploiting vulnerability, and the resulting business impact. As more sensitive information is consolidated in health record systems, the need to avoid breaches through unauthorized access to EHRs becomes increasingly critical. Types of electronic health record breaches range from loss or theft of electronic devices, unauthorized access and the use of unsecured email with sensitive information. In an analysis done by Health Information Trust Alliance (HITRUST) for U.S. healthcare data breaches [15], has reported theft as the most often cause of breaches in healthcare. In 2009, interim final breach notification rule was issued as a part of the HITECH Act. Since 2009, a total of 538 large breaches of protected health information affecting over 21.4 million patient records have been reported [3]. In 2013, Ponemon Institute [9] has conducted a study in nine countries (Australia, Brazil, France, Germany, India, Italy, Japan, UK, U.S.) for data breach. According to its report, it has found that the cause of data breach in German companies were malicious or criminal attack followed by Australia and Japan. Brazilian companies were most likely to experience breaches caused by human error and Indian companies were most likely to experience breaches caused by a system glitch or business process failure. Redspin [3] has reported that during 2010-2011, there was a 97% increase in total records breached and more of the breaches were due to the loss or theft of electronic devices like laptops and backup tapes. In 2012, 146 breaches affecting 2,413,397 individuals were reported to Health and Human Services (HHS).

3.1 States of Healthcare Data

In general, digital data is found to be in one of the following three states:

(1) Data In Motion
(2) Data in Use
(3) Data at Rest

3.1.1 Data in Motion. Data In Motion is data that is traversing a network or temporarily residing in computer memory to be read or updated. Health data moves across public or "untrusted" networks such as Internet as well as in private networks or corporate LANs. It includes the health data stored in cloud that can be inadvertently exposed or targeted by attackers. There are various cloud applications available to store health data on cloud like Box and Drop box but these are also not secure from cyber attacks. According to Google - "Data in motion" - an estimated 5 exabytes of information moves through the internet every two days. Personalized social data, video, and information created by a plethora of new devices. All this data is moving at tremendous velocity and its value is fleeting unless it can be captured, analyzed and responded to in real time. For some industries, such as financial services organizations, even a few milliseconds can have multi-dollar implications. For example, YouTube, Instagram, Gdrive, Drop box. Few years ago, viruses and Trojans were the top chief information security officer concern for endpoint security, but the current data suggests today the threat of data loss through endpoints, whether inadvertent or international, is now the top concern. In 2011, 20,000 patient health records of Stanford Hospital were exposed on a public website accidentally [12]. In 2013, Oregon Health and Science University (OHSU) notified 3,044 patients that their health data was stored on Internet based email or document storage service (also known as cloud computing). The Internet service provider was not an OHSU business associates and hence was unable to confirm about the exposure of data to any attack [5]. Due to the adoption of electronic health records, the sensitive information is becoming more accessible. Many of these records are stored in databases called health information exchanges (HIE), which are linked online making them more accessible. One more application of cloud is the social networking sites like Face book, LinkedIn and Google+, which are the tools for sharing information and to discuss ideas and issues. But these sites have caused various harms to healthcare organizations by exposing these organizations to privacy, security and ethics breaches. The increase in these breaches are due to the gaps in federal privacy regulations, lack of enforcement of existing legislation and the potential for widespread monetization of personal health information by unauthorized users.

3.1.2 Data in Use. Data in Use is active data under constant change stored physically in databases, data warehouses, spreadsheets, files, etc. In healthcare, data in use is the data in active use by HIS and other application. With the ubiquitous use of laptops, tablets and other mobile devices in health sector, the mobility and

remote access to patient information has increased a lot. New technologies such as Web browsing, social media, e-mail and personal applications on personal mobile devices offer many ways to access corporate data. In turn, moving data access across different devices and networks is increasing security risks to the corporate network and opens sensitive corporate data to leaks and attacks. This is because employee-owned mobile devices are beyond the scope of control of internal tech teams, and the risks are compounded by the growth in mobile malware. When employees download and install mobile apps for their personal use, they allow unregulated third-party access to other sensitive, corporate information stored on their devices. These apps may be pre-infected with malware, which might be instructed by hackers' command and control servers to steal information from the mobile device without alerting the users. When the employees' handsets connect to open Wi-Fi networks, the corporate data stored on their devices gets exposed. BYOD (bring your own device) [17] offers healthcare professionals using their personal mobile devices to access applications whenever and wherever it is needed. This whenever and wherever access to sensitive data has lead to various security and privacy risks. Android and its Google operating system that make up a sizeable portion of all the smart phones have attracted the attention of cyber criminals. Blue Coat Security Labs [2] reported an increase in Android malware in the recent past. In July-September 2012 quarter alone, Blue Coat Security Lab saw a 600% increase in Android malware. Additionally, as the cameras and video recorders have become a standard for mobile devices, the risk for data breach has increased. PwC Health Research Institute [10] has recently given a report, which shows that 69% patients were concerned about the privacy of their records as healthcare providers utilize mobile devices for access to them.

3.1.3 Data at Rest. Data at Rest is inactive data stored physically in databases, data warehouses, spreadsheets, archives, tapes, off-site backups (Business Continuity Planning (BCP) and Disaster Recovery (DR) sites), etc. Health data archival refers to the long-term accumulation of patient records and health records into an archive in hospital or healthcare unit. This data is the result of completed activities and is not subjected to change, therefore also known as fixed data. The process of health records archiving ensures that these are well stored in electronic storage like magnetic tapes, disks, CDs, USBs etc. This archived data may include photographs, audio, video recordings and can be used for future database access, any healthcare disaster recovery and emergency management continuity. Data at rest is of increasing concern to business, government agencies and other institutions. Mobile devices are often subject to specific security protocols to protect data at rest from unauthorized access when lost or stolen and there is an increasing recognition that database management systems and file servers should also be considered as at risk; the longer data is left unused in storage, the more likely it might be retrieved by unauthorized individuals outside the network. In 2010, South Shore Hospital in Chicago [23] lost 473 unencrypted back-up computer tapes, which contained personal and protected health information of 800,000 individuals. The information lost included names, social security numbers, financial account numbers and medical diagnoses. Another hospital in Canada [13] lost an unencrypted USB drive containing health data of more than 25000 patients. It contained patient's name, summary data on the type of the service provided, the date of service and health service provider code.

4. XML TECHNOLOGIES FOR SECURITY AND PRIVACY OF HEALTH DATA

To meet the security requirements for privacy, integrity and confidentiality is essential in healthcare. With the growing acceptance of XML technologies, it is logical that security should be integrated with XML solutions. The XML security standards define XML vocabularies and processing rules in order to meet security requirements. These standards use XML security techniques in healthcare to improve the quality, reliability and security of the healthcare services. The XML based security technologies are discussed below.

(1) XML Digital Signature XML Digital Signature is used to lock and seal the contents of a document. It takes the concept of paper-based signature into the digital realm, by adding a digital "fingerprint" as a signature to a document [18]. It allows the signing of a whole or specific section of a document, which provides integrity, message authentication as well as authentication for the signer of the document.

(2) XML Encryption XML Encryption is a specification governed by W3C recommendation that defines how to encrypt the contents of an XML element[7]. Its advantage is that only a specific portion of an XML document can be encrypted rather than the complete document.

(3) XML Key Management Specification (XKMS) XKMS is one of the specifications of XML security which defines the protocol for distributing and registering public keys for verifying digital signatures and enciphering e-documents of e-commerce applications with various functions[8]. The W3C defines it as a set of "protocols for distributing and registering public keys, suitable for use in conjunction with the standard for XML Signatures and companion standard for XML encryption".

(4) Security Assertion Markup Language (SAML) SAML is an XML-based framework for communicating user authentication, entitlement and attribute information developed by the Security Services Technical Committee of OASIS. It allows secure web domains to exchange user authentication and authorization data. In 2009, The OASIS International Consortium announced SAML as an information standard that give hospitals, insurers and others in healthcare community much needed mechanism for exchanging privacy policies, evaluating consent directives and determining authorization.

(5) Extensible Access Control Markup Language (XACML) XACML is an open standard XML-based language that defines the rules needed to make authorization decisions and to express security policies and access rights to information for web services, digital rights management and enterprise security applications [22].

4.1 Issues

XML security is about removing complexities and remedying performance degradation introduced by hefty authentication methods, but with the emergence of cloud computing, mobile computing and social identity technologies, data security has become a challenge. With respect to the management of individual authentication, authorization and privileges in IT systems several standards have been developed which are discussed above. The applicability of these standards for the cloud is challenging. A major challenge in applying these standards and technologies in cloud is scalability. A cloud could have hundreds of nodes and could carry extensive computations and yet ensure that the clients get the responses in a timely manner. Therefore, performance is a major issue.

Fig. 1. XML Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema>
<xmlns:xs="http://www.w3.org/2001/XMLSchema"
element FormDefault="qualified">
<xsd:element name="Patient">
<xsd:complexType>
<xsd:choice minOccurs = "1" maxOccurs = "unbounded">
<xsd:sequence>
<xsd:element ref ="ds:Name"/>
<xsd:element ref ="ds:Address"/>
<xsd:element ref ="ds:DOB"/>
<xsd:element ref="ds:date"/>
<xsd:element ref="ds:blood_group"/>
<xsd:element ref="ds:Gender"/>
<xsd:element ref="ds:marital_status"/>
<xsd:element ref ="ds:ContactNo"/>
<xsd:element ref ="ds:Report"/>
<xsd:element ref ="ds:nurse_progress_report"/>
<xsd:element ref = "ds:doctor_progress_notes"/>
<xsd:element ref = "ds:username"/>
<xsd:element ref = "ds:password"/>
</xsd:sequence>
</xsd:choice>
<xsd:attribute name ="sec_level" type ="xsd:integer"
 use ="required"/>
<xsd:attribute name = "PID" type="xsd:string"
 use = "required"/>
</xsd:complexType>
</xsd:element>
</xsd:schema>
```

5. XML SCHEMA WITH ENHANCED SECURITY AND PRIVACY OF HEALTH DATA

XML Schema can be used to offer a granular access control in XML documents enhancing the security and privacy of health care data. XML Schema is namespace aware because XML Schema is written in XML, therefore, XML Schemas can be programatically processed just like any XML document. Another advantage of XML Schema is its ability to implement strong typing. An XML Schema can define the data type of certain elements, and even constrain it to within specific lengths or values. This ability ensures that the data stored in the XML document is accurate. XML Schema has a wealth of derived and built-in data types to validate content. By introducing security levels for each element inside an XML schema thereby within the XML data itself, as well as assigning security levels to each and every user accessing the data ensures the full security of health records.

Here, in Figure 1, an XML Schema is defined for health care data. In the above listing, "Patient" is the name of the element. PID is the patient number, sec_level is the security level attached with the element for a particular patient. The security level defined here can be given different value depending upon the sensitivity of the user and its data. For example, five users are there which includes- patient, doctor, nurse, admin and pharmacist. Out of all these, patient data is the most sensitive data which should be given highest level of sensitivity and the data for the element patient will also be assigned a security level. Similarly, the security levels are assigned as an attribute for other users and their data such that only an authorized user having a particular security level can access only that portion of data for which he is having access therefore enhancing the security and privacy of health care data leading to an efficient healthcare.

6. CONCLUSIONS

The emergence of digitized techniques has provided many benefits to improve efficient delivery of healthcare services, disease trend analysis and policy formulation by healthcare agencies and service providers. But, these technologies and the connected devices expose healthcare data to increased security and privacy risks. Various XML security technologies are present to deal with these risks which are discussed in the paper. But these techniques have the various issues due to which the vulnerabilities for the security and privacy of health care data have not ended. To deal with these issues, the paper has presented XML Schema using its attributes to provide the security and privacy of health data leading to a secured and efficient health care.

7. REFERENCES

[1] Amrita medical solutions. amritamedical. com/healthcare.html/.

[2] Blue coat systems 2013 mobile malware report. Technical report. www.bluecoat.com/sites/default/files/documents/files/.

[3] Breach report 2011, protected health information. Technical report. www.redspin.com/docs/.

[4] HIMSS. www.himssanalytics.eu/emram/.

[5] OHSU notifies patients of cloud health information storage. www.ohsu.edu/xd/about/new-events/news/2013/.

[6] SAML single sign-on (SSO) service for google applications. http://developers.google.com/google-applications/SSO/.

[7] XML encryption. en.wikipedia.org/wiki/XML-Encryption.

[8] XML key management specification (XKMS). 2001. http://www.w3.org/TR/2001/NOTE-xkms-20010330/.

[9] Cost of data breach study: Global analysis. Technical report, 2013. https://www4.symantec.com/mktginfo/whitepaper/.

[10] PWCs Health Research Institute identifies the top ten health industry issues to watch in 2013. 2013.

[11] Study on the impact of digital technology in health and social care, 2013. https://www.gov.uk/government/publications/.

[12] Jason Dearen. The rising risk of electronic medical records. 2012. www.smartplanet.com/blog/pure-genius/the-rising-ofelectronic- medical-records/8293/.

[13] Jeff Goldman. Montfort Hospital suffers security breach.

[14] Bichilien Hoang and Ashley Caudill. pages 20062012. Published byIEEE Emerging Technology.

[15] Chris Hourihan and Bryan Cline (HITRUST). A look back: U.S healthcare data breach trends. 2012. Published by HITRUST.

[16] Theo Lippeveld, Rainer Sauerborn, and Claude Bodart. Design and implementation of health information systems, WHO. 2000.

[17] P. Rubens. Four steps to securing mobile devices and applications in the workplace. 2012.

[18] Ed Simon, Paul Madsen, and Carlisle Adams. An introduction to XML digital signature. 2001. www.xml.com/pub/a/2001/08/08/xmldsig.html/.

[19] Colin Sweeney. Moving towards a paperless NHS, Kings College Hospital. 2013. Published by Guardian Professional.

[20] Piazza T. and Vizzini G. ISMETT: A paperless hospital. pages 19, 2010. Published by eChallenges.

[21] Jamie Thompson and Web Produces. Healthcare IT News. 2013. www.healthcareitnews.com.

[22] Manish Verma. XML security: control information access with XACML. 2004. www.ibm.com/developerworks/library/x-xacml/.

[23] Weymouth and Ma. Health care IT News, 2012.