Noname manuscript No.
(will be inserted by the editor)

Threat Modeling for Electronic Health Record Systems

Ahmad Almulhem

Received: date / Accepted: date

Abstract The security of electronic health record (EHR) systems is crucial for their growing acceptance. There is a need for assurance that these records are securely protected from attacks. For a system as complex as an EHR system, the number of possible attacks is potentially very large. In this paper, a threat modeling methodology, known as *attack tree*, is employed to analyze attacks affecting EHR systems. The analysis is based on a proposed generic client-server model of EHR systems. The developed attack tree is discussed along with some system properties that enable quantitative and qualitative analysis. A list of suggested countermeasures are also highlighted.

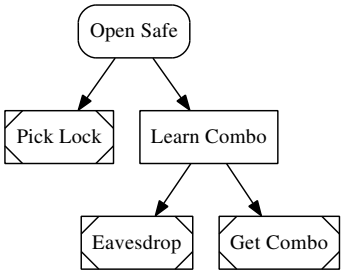
Keywords Electronic Health Record · EHR · Electronic Medical Record · EMR · Attack Tree · Threat Model

1 Introduction

Nowadays, more healthcare providers are migrating from traditional paper-based medical record systems to electronic health record (EHR) systems [1] [2]. EHR systems are increasingly becoming attractive solutions that utilize new advances in computing technologies. In particular, digital storage media currently allows for significant amount of storage space with a relatively low cost. Also, having medical records in digital form is an enabler for efficient computerized processing and networking that are beyond the capabilities of paper-based systems. More importantly, studies have shown that EHR systems do lead to a better patient care [3].

Security, however, is a major concern when adopting EHR systems [4] [5]. For paper-based medical records, physical security measures are probably adequate to

Ahmad Almulhem
Computer Engineering Department
King Fahd University of Petroleum and Minerals (KFUPM)
Dhahran 31261, Saudi Arabia
Tel.: +966-3860-7554
Fax: +966-3860-3059
E-mail: ahmadsm@kfupm.edu.sa



(a)

GOAL: Open Safe (OR)
1. Pick Lock
2. Learn Combo (AND)
2.1 Eavesdrop
2.2 Get Combo

(b)

Fig. 1 An example of an attack tree. The goal of the attack is to open a safe. In (a), the attack tree is depicted graphically in which an OR node is shown as a rectangle with rounded corners, an AND node is shown as a rectangle, and a leaf node is shown as a rectangle with diagonals at the corners. In (b), the same attack tree is listed in textual form.

secure them. Electronic health records, however, are more portable and accessible which make them susceptible for unauthorized accesses and modifications [6].

Threat modeling is a useful tool to analyze the security requirements of a computer system. In this paper, a threat modeling methodology, known as *attack tree*, is used to analyze threats facing EHR systems. The goals of this paper are twofold. The first goal is to identify possible attacks on a typical EHR system, and to identify suitable countermeasures. The second goal is to show that the attack tree methodology is an effective tool when designing and deploying such systems.

The rest of the paper is organized as follows. In section 2, a brief description of attack trees is presented. In section 3, EHR systems are reviewed and a generic model of EHR systems is proposed. Then, an attack tree of EHR systems is presented and discussed in section 4. Next, some suggested countermeasures are highlighted in section 5.

2 Attack Trees

An *attack tree* is a conceptual tree that represents possible attacks on a system [7] [8] [9]. Developing the tree provides a systematic methodology to enumerate possible attacks on a given system, and captures inter-dependencies between these attacks. In this methodology, attacks are depicted as a tree structure where the root represents the goal of the attack, and the children nodes represent means to achieve this goal. The tree may be depicted either graphically or in textual form as shown in Figure 1.

In an attack tree, a node represents an attack that succeeds when the node's direct children are true. Specifically, node's children form preconditions for the attack to happen. These preconditions (children) are combined using two logical operators: *OR* and *AND*. When combined with an OR operator, an attack succeeds if any of the preconditions is true. When combined with an AND operator, an attack succeeds if all of the preconditions are true.

The example in Figure 1 is a stripped-down version from [8]. In this example, the goal of the attack is to open a safe. For the graphical representation, the following symbols are used to distinguish OR, AND and leaf nodes:

– An OR node is shown as a rectangle with rounded corners
– An AND node is shown as a rectangle
– A leaf node is shown as a rectangle with diagonals at the corners

For this example, an attacker can either pick the lock OR learn the combination to open the safe. Subsequently, to learn the combination, an attacker has to eavesdrop a related conversation AND get the combination. Note that the node "Open Safe" is an OR node, which indicates that "Open Safe" succeeds if any of its children nodes is true. Similarly, note that "Learn Combo" is an AND node, which indicates that "Learn Combo" succeeds when all its children nodes are true.

Attack trees are quite flexible in representing attacks at different levels of abstractions and scales. They can also be used to perform quantitative and qualitative analysis of attacks. To this end, leaf nodes are assigned values which can be categorical, ordinal and/or numerical. These values are then combined at OR and AND nodes according to defined operations. Effectively, values are accumulated bottom-up propagating from leaf nodes to the root node. For instance, in the above example, assume each node has a value representing the cost in dollar to carry an attack. One, then, can accumulate the costs to find out the total cost of the main goal, i.e. "opening the safe".

3 Electronic Health Record Systems

Health records (paper-based and electronic-based) are inherently complex and diverse [10]. The International Organization for Standardization (ISO) defines an electronic health record as a "repository of information regarding the health status of a subject of care, in computer processable form, stored and transmitted securely and accessible by multiple authorized users, having a standardized or commonly agreed logical information model that is independent of EHR systems and whose primary purpose is the support of continuing, efficient and quality integrated health care" [11]. ISO's definition makes a clear distinction between EHR and an EHR system; i.e. the content of the EHR and its architecture. This distinction is essential from the viewpoint of standardization in order to ensure semantic interoperability [12]. Other related terminologies and definitions are also quite common in the literature and in practice [3].

The architectures of EHR systems differ from one healthcare provider to another [13]. Generally, an EHR system is composed of several components as shown in Figure 2. These components closely reflect the various services a patient receives from the different departments such as radiology, laboratory, pharmacy, and administration [14]. Additionally, a component is usually implemented as a *client-server* application that employs a request-reply protocol [15]. A client-server architecture facilitates secure access for multiple authorized users. Accordingly, an EHR system is actually made of several client-server applications. These applications may be provided by different vendors. Therefore, it is not uncommon for a clinician to log into different applications in a typical patient visits [14]. For usability purposes, a GUI is usually implemented to provide a uniform integrated access to all of the applications.

In order to apply the attack tree methodology discussed in Section 2, the following generic EHR system model is used. It consists of three components:

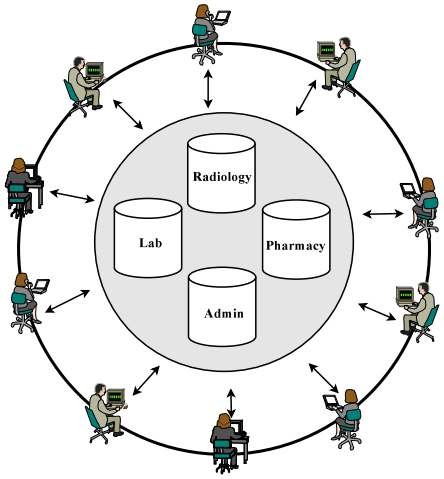


Fig. 2 EHR System Architecture: The system typically consists of several client-server applications and supports secure access for different authorized users.

- **Client(s):** One or more applications that access services made available by servers. The type of access (reading, writing, .. etc.) depends on a user's authorizations.
- **Server(s):** One or more servers that provides services. A server can be a server-side application, a supporting backend server (like databases, web servers) or a physical server.
- **Network:** A network infrastructure for bidirectional communications between clients and servers .

This proposed model is representative of EHR systems used in practice [16].

4 Attacking EHR Systems

4.1 Attacker Goals

As a first step in developing the attack tree, one needs to specify attacker goals. The main goal is designated as "Compromise the EHR System", which is then divided into the following subgoals:

1. Compromise Client
2. Compromise Server
3. Compromise Network

These attacker goals closely follow the EHR model presented in section 3. An attacker most likely would target the visible components of the system; namely the client(s), the server(s) or the network. Compromising any one of them results in compromising the entire system.

```
GOAL : Compromise the System (OR)

1. Compromise Client (OR)
1.1. Shoulder surfing
1.2. Use unattended logged-on client
1.3. Obtain valid username/password (OR)
1.3.1. Social engineering
1.3.2. Network interception
1.3.3. Key-logging
1.3.4. Phishing emails
1.4. Infect with malware (OR)
1.4.1. Deliver malware through Email attachment (e.g. PDF documents)
1.4.2. Lure into visiting a malicious website
1.4.3. Run infected programs (AND)
1.4.3.1. Gain local access
1.4.3.2. Obtain administrator username/password
1.5. Steal client if portable (e.g. smart phones)
1.6. Destroy client

2. Compromise Server (OR)
2.1. Gain remote access (OR)
2.1.1. Use default username/password (e.g. guest)
2.1.2. Use exploit(s) (AND)
2.1.2.1. Find open port(s)
2.1.2.2. Identify working exploit(s)
2.2. Gain local access (AND)
2.2.1. Gain physical access
2.2.2. Obtain administrator username/password
2.3. Make server slow or unavailable (OR)
2.3.1. Flood with traffic
2.3.2. Flood with requests
2.3.3. Destroy or steal server (AND)
2.3.3.1. Gain physical access
2.3.3.2. Use suitable tool(s)

3. Compromise Network (OR)
3.1. Eavesdrop traffic (AND)
3.1.1. Capture packets
3.1.2. Decode traffic
3.2. Modify or inject traffic (OR)
3.2.1. Perform man-in-the-middle attack
3.2.2. Perform replay attack
3.3. Make network unavailable (OR)
3.3.1. Cut network cables
3.3.2. Destroy wireless access points
```

Fig. 3 A proposed attack tree of an EHR system. The main goal of compromising the system is divided into three subgoals; namely compromising clients, compromising servers, and compromising the network.

4.2 Attack Tree

The identified attacker goals are further elaborated resulting in the attack tree shown in Figure 3. The tree comprises 42 attacks where some are technical and some are not. For instance, performing man-in-the-middle attack requires technical knowledge, while social engineering does not. This demonstrates the flexibility of attack trees in representing different types of attacks.

In some respect, the health records flowing in an EHR system are the main assets of the system. Technically, securing them means ensuring their confidentiality, integrity, and availability. These three key principles of information security are implicitly embedded in the proposed attack tree. For instance, compromising the network may result from either eavesdropping traffic, modifying or injecting traffic, or making the network unavailable. These network attacks correspond to breaching confidentiality, integrity, and availability respectively.

In an EHR system, clients are probably the most visible parts of the system. They also play the key role of viewing, entering and modifying health information. Therefore, they are expected to be attacked the most. As listed in the attack tree, some attacks can be as simple as shoulder surfing. Also, as more clients run on portable devices (e.g. smart phones), stealing these devices becomes an attractive attack.

Considering all the attacks, compromising a server is probably the most serious attack. Damages to the system may include exposing, altering and/or destroying health information. From an attacker viewpoint, however, it is probably the most rewarding attack. In particular, gaining a remote access grants a complete control of the entire system.

Compromising the network is yet another attractive goal. Two factors are helping in this regard. One is the increase adoption of wireless technologies by healthcare providers. The second is using the Internet to share health information between collaborating providers. Both factors give an attacker more chances to attempt eavesdropping, modifying or injecting health information [17] [18].

4.3 Quantitative and Qualitative Analysis

Attack tree methodology is not only effective to identify various attack, but it is also useful to perform quantitative and qualitative analysis. While identifying various attacks is useful to implement countermeasures, quantitative and qualitative analysis are used to prioritize the implementation of such countermeasures. For instance, one might identify the attacks with disastrous effects and address them first.

To perform such analysis, two steps are needed to augment an attack tree. First, leaf nodes are assigned certain values that correspond to system proprieties of interest. Second, operations are defined to combine these values at "AND" and "OR" nodes.

The following is a list of three system proprieties along with their operations:

- **Attack is Possible?:** A Boolean value indicating whether the attack is possible or not.
  - Type: Boolean
  - Domain: True, False
  - AND: Possible if all children nodes are True; impossible otherwise
  - OR: Possible if any children node is True; impossible otherwise
- **Cost of Attack:** An aggregate estimate of the cost incurred if the attack is successful.
  - Type: Numerical
  - Domain: 1 - 10 (cheap — expensive),

  - AND: Sum of children nodes
  - OR: Minimum of children node
- **Probability of Detection:** An estimate of how difficult (easy) the attack can be detected.
  - Type: Numerical
  - Domain: 0 - 1 (difficult — easy),
  - AND: Product of children nodes
  - OR: Minimum of children node

Using the defined properties, one may answer several interesting queries about the system. Such queries can help to prioritize the attacks, and hence provides a practical approach to address these attacks. The following is a sample of queries that can be answered by the proposed attack tree:

- Which attacks are possible?
- What is the costliest attack?
- Which attacks cost more than 7 on the defined scale?
- What is the most difficult attack to detect?

5 Countermeasures

The following is a list of suggested countermeasures to address the attacks identified earlier.

- User Authentication: Users must be authenticated to use a client. Good practices should be employed such as strong password policy, biometrics and multi-factor authentication.
- User Authorization: Role-based access policies should be developed employing principles such as least privileges and separation of duties. Authorized users should be assigned to roles with minimum privileges.
- Auditing: All activities must be recorded and attributed.
- Malware Detection: Clients can be infected with different types of malwares. Therefore, malware detection tools should be installed and kept up to date.
- Timeout Policy: When a client is inactive for a certain time, it should be locked or logged out.
- Access Revoke Policy: Portable clients can easily be stolen. Therefore, there should be a mechanism to revoke their accesses to the system.
- Servers Security: Servers should be secured using multi-layer countermeasures including physical security, firewall, intrusion detection systems.
- Encryption: Many network attacks can be significantly reduced if all communications are encrypted using cryptographic protocols such as TLS/SSL [19].
- Security Awareness: All authorized users should complete an awareness program in which they learn about attacks, consequences and good practices.

6 Conclusion

Clinicians as well as patients need assurance that their electronic health records are protected from current and future attacks. Such concerns will likely continue to

grow as more healthcare providers adopt EHR systems. Additionally, the promising benefits of adopting these systems will be greatly affected should their security is compromised. A tool like attack tree can prove effective in enumerating such attacks (technical or non-technical). It can be used to account for different types of attacks that threaten complex systems such EHR systems. Early attacks analysis would help in planning for countermeasures, and would greatly reduce the impacts of these attacks.

References

1. Byron Hamilton. *Electronic Health Records*. McGraw-Hill, 2nd edition, 2010.
2. HIMSS Analytics. The emr adoption model. http://www.himssanalytics.org/, June 2011.
3. Kristiina Hyrinen, Kaija Saranto, and Pirkko Nyknen. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International Journal of Medical Informatics*, 77(5):291–304, May 2008.
4. Ross Anderson. A security policy model for clinical information systems. In *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, pages 30–43, 1996.
5. Ross Anderson. Clinical system security: interim guidelines. *British Medical Journal*, 312(7023):109–111, January 1996.
6. Randolph C. Barrows and Paul D. Clayton. Privacy, Confidentiality, and Electronic Medical Records. *Journal of the American Medical Informatics Association*, 3(2):139–148, March 1996.
7. Edward Amoroso. *Fundamentals of Computer Security Technology*. Prentice Hall, 1994.
8. Bruce Schneier. Attack trees - modeling security threats. *Dr. Dobb's Journal*, 24(12):21–29, December 1999.
9. A. P. Moore, R. J. Ellison, and R. C. Linger. Attack modeling for information security and survivability. *Software Engineering Institute, Carnegie Mellon University*, Technical Note: CMU/SEI-2001-TN-001, 2001.
10. Thomas Beale. The health record why is it so hard? *IMIA Yearbook of Medical Informatics*, 2005:301–304, 2005.
11. ISO/TR 20514. *Health informatics Electronic health record Definition, scope and context*. ISO, Geneva, Switzerland, 2005.
12. Marco Eichelberg, Thomas Aden, Jörg Riesmeier, Asuman Dogac, and Gokce B. Laleci. A survey and analysis of electronic healthcare record standards. *ACM Computing Surveys*, 37:277–315, December 2005.
13. Takeharu Sonoda. Evolution of electronic medical record solutions. *Fujitsu Scientific & Technical Journal*, 47(1):19–27, January 2011.
14. MITRE Corporation. Electronic health records overview. Technical report, National Institutes of Health National Center for Research Resources, April 2006.
15. H.K. Huang. *PACS and Imaging Informatics: Basic Principles and Applications*. John Wiley & Sons, 2nd edition, 2010.
16. Cecily Morrison, Adona Iosif, and Miklos Danka. Report on existing open-source electronic medical records. Technical Report UCAM-CL-TR-768, University of Cambridge, February 2010.
17. Wu Liu, Ping Ren, Yong Zhang, and Hai xin Duan. Ssl-dp: A rootkit of network based ssl and tls traffic decryptor. In *Cybercrime and Trustworthy Computing Workshop (CTC), 2010 Second*, pages 29 –33, july 2010.
18. Kurt Seifried. Attacks against ssl. *Linux Magazine*, 112:60–61, March 2010.
19. T. Dierks and E. Rescorla. The transport layer security (tls) protocol version 1.2. RFC 5246, Internet Engineering Task Force, August 2008.