Cataloging and Comparing Logging Mechanism Specifications for Electronic Health Record Systems

Jason King and Laurie Williams
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
{jtking, laurie_williams}@ncsu.edu

Abstract – Electronic health record (EHR) systems must log all transactions with protected health information (PHI) to deter unauthorized behavior and prevent users from denying that they created, read, updated, or deleted PHI. However, a plethora of standardization and governing organizations publish documentation (such as standards, suggestions, and requirements) to outline transactions that should be logged and the data that should be captured for each log entry. *The objective of this research is to guide the design of electronic health record systems by cataloging suggested information that should be captured by logging mechanisms from both healthcare and non-healthcare documentation*. In this paper, we focus on three types of information: data transactions, security events, and log entry content. We collect a set of ten healthcare-related and six non- healthcare related documents that contain specifications for logging mechanisms. From these 16 sources, we catalog 11 data transactions, 77 security events, and 22 data elements for log entry content. Overall, we identify 14 security events and 2 data elements for log entry content that are not explicitly addressed by healthcare documents. We found that developers must consider 13 of the 16 documents to extract 100% of the security events and log entry content cataloged.

1 Introduction

In software security, repudiation threats are threats associated with users who deny performing some action within the software system without other parties having any way to prove otherwise [1]. *Nonrepudiation* refers to the ability of the software system to mitigate such repudiation threats. For example, in healthcare, a nurse may view diagnosis data for his or her neighbor without having a legitimate, authorized need to view the neighbor's diagnosis data. When rumors of the diagnosis circulate the neighborhood, the nurse denies viewing the diagnosis data and cannot be disproved. To counter this repudiation threat, the electronic health record (EHR) system needs to provide proof that the nurse did, indeed, view the neighbor's diagnosis data. If the EHR system had no logging mechanism, the nurse cannot be disproved and may avoid punishment for snooping at her neighbor's diagnosis data. Adequate logging mechanisms are one way EHR systems can stimulate nonrepudiation.

Many industrial and governmental organizations publish guidelines, regulations, standards, certification criteria, general suggestions, and requirements for software logging mechanisms. As a result, documents detailing logging mechanism specifications are distributed across governing organizations, standards bodies, and other administrative groups. For example, EHR systems used in the United States should abide by specifications for logging mechanisms included in meaningful use standards published by the Office of the National Coordinator for Health Information Technology (ONC-HIT) [2] [3]. For EHR systems also used in Canada, the federally-funded Canada Health Infoway maintains privacy and security requirements, including logging mechanisms requirements, which EHR systems must satisfy to protect patient privacy and maintain the confidentiality, integrity, and availability of patient data [4]. For software developers unfamiliar with the various standards and requirements for software logging mechanisms, some of these documents may be overlooked, resulting in inadequate logging mechanisms.

The objective of this research is to guide the design of electronic health record systems by cataloging suggested information that should be captured by logging mechanisms from both healthcare and non-healthcare documentation. In this paper, we focus on cataloging three types of information that should be captured by logging mechanisms:

- **Data transactions:** To promote accountability and deter unauthorized user behavior, all data transactions (such as create, read, update, delete) with protected health information (PHI) should be logged by EHR system logging mechanisms.

- **Security events:** General security events should be logged, such as events associated with authentication (such as user logins) and access control mechanisms (such as granting user privileges).
- **Log entry content:** Appropriate contextual data (such as timestamps and user identification) must be captured to provide a comprehensive, accurate, and trustworthy trail for ensuring user accountability.

First, we collect a set of documents that contain suggested data transactions, security events, and log entry content. We collect a total set of 16 documents, including 10 healthcare and 6 non-healthcare documents. Next, we extract a set of 387 individual data transactions, security events, and log entry content from the source documents. Next, we combine similar data transactions, security events, and log entry content and categorize the collected information.

For this study, we define the following research questions:

RQ1: What data transactions should be logged in EHR systems?

RQ2: What security events should be logged in EHR systems?

RQ3: What log entry content should be captured for each log entry in EHR systems?

RQ4: What data transactions, security events, and log entry content are *not* included in healthcare documents, but are included in non-healthcare documents?

RQ5: Which document offers the most detailed specifications for data transactions, security events, and log entry content that should be captured by EHR logging mechanisms?

RQ6: What minimal set of documents covers 100% of the cataloged data transactions, security events, and log entry content?

This paper contributes the following to the current state of secure EHR system logging mechanisms:

- A centralized catalog of data transactions, security events, and data elements for log entry content extracted from multiple cross-domain sources. This catalog provides a more comprehensive set of suggested data that EHR system developers should consider when designing and implementing logging mechanisms.
- Documented traceability between the source documentation and each collected data transaction, security event, and log entry content data element. Traceability helps EHR system developers identify and locate source documents for additional reading and understanding of various standards, suggestions, and requirements, and how each standard, suggestion, or requirement applies to EHR systems.

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 presents our methodology. Section 4 summarizes our results and findings for RQ1, RQ2, and RQ3. Section 5 discusses findings for RQ4, RQ5, RQ6, and general findings for the study as a whole. Section 6 presents limitations. Section 7 discusses future work. Section 8 summarizes this study.

2 Related Work

In prior work [7] [24], we extracted a set of 16 auditable events (data transactions and security events) from 4 sources of logging mechanism specifications:

- Chuvakin and Peterson [8] "How to Do Application Logging Right"
- The Certification Commission for Health Information Technology (CCHIT) [9] appendix of auditable events for EHR systems.
- The SysAdmin, Audit, Network, Security (SANS) Institute [10] checklist of information system audit logging requirements.
- The "IEEE Standard for Information Technology: Hardcopy Device and System Security" [11] best practices for logging and auditability.

No single auditable event extracted was unanimously suggested by all 4 sources.

We then evaluated logging mechanisms of two open-source and one commercial EHR systems (OpenEMR[1], Tolven eCHR[2], and ProprietaryMed[3]) to determine how many of the 16 general auditable events were actually logged within the EHR systems. We identified an overall deficiency in logged events. OpenEMR logged 62.5% of the general auditable events, compared to Tolven eCHR logging 6.5% and ProprietaryMed logging 18.75% of the auditable events.

For the current study, we expand upon the initial set of four sources of logging specifications to 16 total sources. We also compare and contrast the extracted data transactions, security events, and log entry content to draw conclusions about the adequacy and comprehensiveness of current healthcare documentation.

Software security organizations often provide repositories of information for common security issues. For example, the Common Weakness Enumeration (CWE) & SANS Institute [12] publish a collection of common security vulnerabilities in software. If software developers wish to learn more about potential security vulnerabilities, including the most commonly observed

[1] http://www.open-emr.org/
[2] http://www.tolven.org/echr.html
[3] The company requested to remain confidential

vulnerabilities for the recent year, they may rely on the CWE/SANS collection to gain knowledge and understanding of how to mitigate potential threats in their own software systems. CWE-778 [4] specifically discusses insufficient logging: "When a security-critical event occurs, the software either does not record the event or omits important details about the event when logging it".

The Open Web Application Security Project (OWASP) [13] is an international organization focusing on improving security of software. OWASP maintains an online catalog of informational "cheat sheets" that guide developers in understanding how to mitigate threats associated with given vulnerabilities. For example, the OWASP Logging Cheat Sheet features generalized information on designing, implementing, and testing security logging mechanisms. The cheat sheet also includes a list of references for more information about logging frameworks, including The MITRE Corporation's Common Event Expression[5] and the World Wide Web Consortium's Extended Log File Format[6]. The OWASP Logging Cheat Sheet seems to focus on monitoring of system-level security events, with less emphasis on logging transactions with protected data.

The United States National Vulnerability Database (NVD) [14] maintains a centralized repository of publicly known information security vulnerabilities and exposures. This database may help developers understand common vulnerabilities in software, as well as recent trends in vulnerability exploits. For example, CVE-2010-0502 [7] describes a vulnerability in Apple Mac OS X iChat Server where the sending of certain types of messages are not logged. Remote attackers may avoid detection by launching attacks using one of the unlogged message types. Other vulnerabilities related to logging include leakage of sensitive information contained within log files; script injection through functions that drive the logging mechanism; and capturing incomplete or incorrect information (such as incorrect user identification). The NVD also publishes security checklists that provide guidance on security configuration of operating systems and applications.

Neither CWE/SANS, OWASP, nor NVD offer a comprehensive catalog of documented logging standards or specifications with traceability between source documentation and the suggested data transactions, security events, and log entry content. The OWASP Logging Cheat Sheet offers implementation suggestions, but it does not offer cross-domain logging specifications that feature traceability to source documents. Instead, OWASP offers yet another checklist or suggested outline of how a logging mechanism should work.

CWE/SANS and NVD seem reactive in nature. Both publish information based upon security vulnerabilities and exploits that have already occurred in actual software systems. The catalog of data transactions, security events, and log entry content in our current study adopts a proactive approach to gather and publish data transactions, security events, and log entry content for developers to consider when developing software systems. We aim to mitigate repudiation threats by helping developers implement adequate logging mechanisms from the beginning of the software development lifecycle, not as an afterthought or reaction to a successful repudiation threat.

Without well-defined, comprehensive standards and logging specifications by a central governing body, the industry has no single widely-adopted standard for software logging mechanisms [15]. The responsibility of locating and interpreting ambiguous or nondescript data transactions, security events, and log entry content falls upon individual software development teams who may be unprepared, untrained, or unaware of additional documents that equally apply to and govern the software upon which they work.

3 Methodology

We first present our methodology for collecting our source documentation. Next, we present our methodology for categorizing extracted data transactions, security events, and log entry content.

3.1 Collecting source documentation

We begin with the same set of four documents from previous work [7] discussed in Section 2. Using this set of four documents as our base set, we read each document and traced references to identify any additional documents discussed. For example, the CCHIT certification criteria frequently references HL7 standards [6], Healthcare Information Technology Standards (HITSP) [16], and National Institute of Standards and Technology (NIST) Special Publication 800-53: Recommended Security Controls for Federal Information Systems and Organizations [17]. For traceability, we record the relationships and connectivity among the collected documents.

To include a document in our study, we required the following inclusion criteria be met:

- The document must contain a section discussing logging or auditing mechanism specifications
- The full text of the document must be freely available.

Tracing references in our base set of four documents, we found five additional documents:

[4] http://cwe.mitre.org/data/definitions/778.html
[5] http://cee.mitre.org/
[6] http://www.w3.org/TR/WD-logfile.html
[7] http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2010-0502

- Integrating the Healthcare Enterprise: Audit Trail and Node Authentication [22]
- Health Information Technology Standards: Collect and Communicate Security Audit Trail Transaction [16]
- Health Level Seven International: Common Audit Message version 2.x [6]
- National Institute of Standards and Technology SP 800-53: Security and Privacy Controls for Federal Information Systems and Organizations [17]
- The Joint NEMA/COCIR/JIRA Security and Privacy Committee: Security and Privacy Auditing in Health Care Information Technology [23]

Since CCHIT (one of the documents in our base set) is an approved certification body for the United States ONC-HIT Meaningful Use standards, we manually added two documents to our collection:

- Meaningful Use Stage 1 standards [2]
- Meaningful Use Stage 2 standards [3]

After tracing references in the two Meaningful Use documents, we manually added one additional document:

- ASTM E247: Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems [18]

Based on prior knowledge and experience with logging mechanism specifications across domains, we manually added an additional three documents to our collection:

- Canada Health Infoway EHR Privacy and Security Requirements [4]
- Payment Card Industry Data Security Standard [19]
- United States Department of Defense Trusted computer System Evaluation Criteria "Orange Book" [20]

After tracing references in these three documents, we manually added one additional document:

- United States Department of Defense Directive 5200.28: Security Requirements for Automated Information Systems (AISs) [21]

Overall, we collected a total of 16 documents for logging mechanism specifications. Table 1 summarizes key information about the source documents collected for this study, including a short identifier for each document. Figure 1 visually summarizes observed associations (arrows) among the 16 collected documents. For example, as shown in Figure 1, CCHIT is an authorized certification body for verifying that EHR systems meet Meaningful Use standards. Therefore, an association exists between MU1 and CCHIT, and MU2 and CCHIT. Similarly, there are no cross-domain associations between healthcare documents and non-healthcare documents.

3.2 Collecting data transactions, security events, and log entry content

We manually read and processed the text of each document to identify any data transactions, security events, or data elements that should be captured in the log file for each log entry. For data transactions, security events, or log entry content listed as compound statements, such as "user logins/logouts should be recorded", we split and recorded each separately as two concrete items: (a) user logins and (b) user logouts. For each individual data transaction, security events, and log entry content, we recorded a brief description along with the source document name.

3.3 Categorizing data transactions, security events, and log entry content

We grouped duplicates and similar data transactions, security events, and log entry content. However, we maintained each individual description and source document name as part of each grouping. For example, "user login" from Document A, "user login attempts" from Document B, and "successful user authentication" from Document C would all be gathered into an overall "user logins" grouping, which will provide traceability back to the original source Documents A, B, and C. Since we want to maintain traceability between source documents and each data transaction, security event, and log entry content, and since we also want to count how many documents recommended *each* data transaction, security event, and log entry content, we did not eliminate any items during categorization.

3.3.1 Data transactions. Since most major database transactions fall under either create, read, update, or delete (CRUD), we grouped our data transactions into one of the following groupings: create, read, update, or delete. Not all data transactions were atomic, meaning the data transaction did not fit cleanly into only one of the CRUD groupings. For example, "merge patient record" could involve a combination of creating a new record, reading an old record, updating a new record, and/or deleting an old record. We kept the non-atomic data transactions separate, but we also recorded which of the four atomic transactions (CRUD) may apply.

3.3.2 Security events. If a security event description were very general such as "security administration event", which could include anything from authentication to encryption of data being transmitted, we placed the security event into a broad "general security" category. If a security event description were ambiguous, or if we were unsure what the intent of the security event was, we kept the security event separate and did not combine it into a larger grouping.

3.3.3 Log entry content. If a description for a log entry content were ambiguous, or if we were unsure what the intent of the log entry content was, we kept the

Table 1. Summary of included documents
IDs correspond to those used in Figure 1

ID	Year	Organization/Author	Title
ASTM	2013	American Society for Testing and Materials (ASTM)	E2147: Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems [18]
CCHIT	2011	Certification Commission for Health Information Technology (CCHIT)	Ambulatory EHR criteria [9]
CHI	2004	Candian Health Infoway (CHI)	Electronic Health Record Privacy and Security Requirements [4]
DoD	1988	United States Department of Defense (US DoD)	Directive 5200.28: Security Requirements for Automated Information Systems (AISs) [21]
HITSP	2007	Health Information Technology Standards (HITSP)	Collect and Communicate Security Audit Trail Transaction [16]
HL7	2001	Health Level Seven International (HL7)	Common Audit Message version 2.x [6]
HALR	2010	Chuvakin & Peterson	How to Do Application Logging Right [8]
IEEE	2008	Institute of Electrical and Electronics Engineers (IEEE)	IEEE Standard for Information Technology: Hardcopy Device and System Security [11]
IHE	2012	Integrating the Healthcare Enterprise (IHE)	Audit Trail and Node Authentication (ATNA) [22]
MU1	2010	United States Centers for Medicare & Medicaid Services (US CMS)	Meaningful Use Stage 1 [2]
MU2	2012	United States Centers for Medicare & Medicaid Services (US CMS)	Meaningful Use Stage 2 [3]
NEMA	2001	Joint National Electrical Manufacturers Association (NEMA), European Coordination Committee of the Radiological and Electromedical Industry (COCIR), and Japan Industries Association of Radiological Systems (JIRA) Security and Privacy Committee	Security and Privacy Auditing in Health Care Information Technology [23]
NIST	2013	National Institute of Standards and Technology (NIST)	SP 800-53: Security and Privacy Controls for Federal Information Systems and Organizations [17]
OB	1985	United States Department of Defense (US DoD) Standard	Trusted Computer System Evaluation Criteria [20]
PCI	2010	Payment Card Industry (PCI) Security Standards Council	Payment Card Industry Data Security Standards v2.0 [19]
SANS	2007	SysAdmin, Audit, Network, and Security Institute (SANS)	Information System Audit Logging Requirements [10]

log entry content separate and did not combine it into a larger grouping.

4 Results

In this section, we answer RQ1, RQ2, and RQ3 by presenting the data transactions, security events, and log entry content collected.

4.1 Summary

Before combining similar items, we collected a total of 118 data transactions, 179 security events, and 90 log entry content from the 16 source documents. Table 2 summarizes the quantity of data transactions, security events, and log entry content extracted from the source documents.

Table 2. Summary of extracted data transactions, security events, and log entry content. *IDs correspond to IDs in Table 1.*

ID	Data Transactions	Security Events	Log Entry Content	Total
ASTM	5	4	8	**17**
CCHIT	9	10	5	**24**
CHI	4	5	8	**17**
DoD	5	12	4	**21**
HL7	35	36	16	**87**
HALR	4	20	9	**33**
HITSP	0	0	0	**0**
IEEE	0	19	0	**19**
IHE	39	19	1	**59**
MU1	4	0	3	**7**
MU2	0	4	0	**4**
NEMA	5	29	7	**41**
OB	1	5	6	**12**
PCI	3	5	6	**14**
SANS	4	11	8	**23**
NIST	0	0	9	**9**
Total	**118**	**179**	**90**	**387**

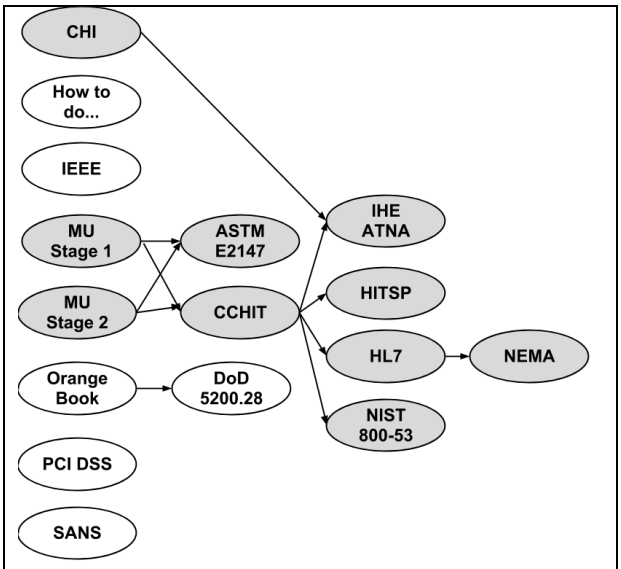

Figure 1. Summary of relationships among the included documents. *IDs correspond to those presented in Table 1. Arrows represent associations between documents.*

4.2 Data transactions collected

RQ1: What data transactions should be logged in EHR systems? Before combining similar items, we collected a total of 118 individual data transactions from the source documents. Eighty-nine out of 118 data transactions were categorized as atomic data transactions (fitting into exactly one of create, read, update, or delete). Additionally, we identified 29 total non-atomic data transactions. After categorization and grouping similar items, we identified four unique atomic transactions and seven unique non-atomic transactions. Table 3 summarizes our data transaction categorization.

Table 3. Summary of data transactions after categorization

Category	Number of Data Transactions
Atomic Transactions	
Create	37
Read	19
Update	20
Delete	13
Total	*89*
Non-atomic Transactions	
Cancel	9
Resolve/Complete	8
Disperse/Deliver	3
Assign	5
Deassign	2
Merge record	1
Unmerge record	1
Total	*29*
Total	**118**

4.3 Security events collected

RQ2: What security events should be logged in EHR systems? Before combining duplicates, we collected a total of 179 security events from the source documentation. After categorization, we identified 77 distinct security events. Table 4 summarizes the security event categories collected from the source documents[8], including the number of events grouped within each category and the total number of extracted individual security events.

Table 4. Summary of security events after categorization, sorted by number of events within each category

Category	Number of Events
User Events	11
Access Control	8
Authentication	4
Detection of Malicious Activity	1
Administration Events	
Application Administration	21
System Administration	18
Log Administration	7
General Administration	7

4.4 Log entry content collected

RQ3: What log entry content should be captured for each log entry in EHR systems? Before combining duplicates, we collected a total of 90 individual log entry content from the source documents. After categorization, we identified 22 distinct groupings of log entry content. Table 5 summarizes the data elements for log entry content collected from the 16 source documents.

Table 5. Summary of log entry content data elements after categorization, sorted by number of data elements within each category

Category	Number of Data Elements for Log Entry Content
Object Affected	7
Source Identification	4
User Identification	4
Breach Access	1
Destination Identification	1
Event Description	1
Priority	1
Reason	1
Success/Failure	1
Timestamp	1

[8] A full, detailed list can be found at http://go.ncsu.edu/loggingcatalog

5 Discussion

In Section 4, we presented our raw data findings that answer RQ1, RQ2, and RQ3. In this section, we discuss our findings for RQ4, RQ5, and RQ6, which require more in-depth analysis and discussion.

RQ4: What data transactions, security events, and log entry content are not included in healthcare documents, but are included in non-healthcare documents? For **data transactions** collected, all 11 data transactions were included in the healthcare documents. For **security events**, we identified 14 individual events that were included only in non-healthcare documents:

- Exhausted resources
- Access to cryptographic keys
- Application/component installation
- Change audit configuration
- Change of access protocol/port
- Disable access protocols/ports
- Enable access protocols/ports
- Change to access control
- Functions initiated by automated data processing operators
- Hardware additions
- Hardware deletions
- Invalid input
- Override of human-readable output markings
- System hardware maintenance actions

Four source documents (IEEE, "How to Do Application Logging Right", Department of Defense Directive 5200.28, and the "Orange Book") collectively provide these 14 security events. The majority of these events are not application-specific security events performed by EHR users. Instead, the 14 security events relate to network, hardware, and software configuration components of application and system administration. Since the 14 security events are not EHR or healthcare-specific, they may not have been considered for inclusion in the healthcare documents. However, for a comprehensive logging mechanism, each of the 14 security events should still be considered when implementing EHR systems. For example, a user who systematically provides invalid input into text fields in the EHR system could potentially be a malicious user attempting a SQL injection attack. If the user's repeated attempts to enter invalid input are not logged, and the user succeeds in launching an injection attack, the logging mechanism does not provide an auditable trail of the user's repeated attempts to attack the system through each text field. The malicious user could potentially successfully deny launching the attack without the EHR system having any proof otherwise.

Similarly, no healthcare document includes the logging of changes to audit configurations, which may include specifying which events to log and where to store log files. Suppose a malicious administrative user disables logging for prescription creations. An administrative user might then have the ability to conspire with a healthcare provider to commit prescription drug fraud by submitting fake prescription orders. Without logging who changed the audit configuration, the malicious administrative user could deny doing so. Therefore, EHR systems should log all changes to the audit configuration to help promote administrative accountability.

For **log entry content**, we identified two data elements not included in healthcare documents:

- Before/after values
- Priority

Both before/after values and priority could provide useful information when reconstructing traces of user behavior. For example, before and after values for an update to prescription information would include the original dosage value and the updated dosage value. This information could help investigators discover issues associated with medical malpractice. Here, nonrepudiability is important since the healthcare employee could not deny changing the dosage value to an improper or erroneous value in the event of patient sickness or death. Likewise, with an event priority captured for logged entries, auditors could focus on higher-priority function executions (such as failed backups) or PHI accesses that have been flagged by the system with a high priority for being audited. For example, suppose an emergency situation arises where a patient's declared healthcare provider is not immediately available. The logging mechanism could indicate that the PHI access has a high audit priority because of no existing relationship between the provider and the patient. Both before/after values and priority could be useful pieces of information for auditors.

RQ5: Which document offers the most detailed specifications for data transactions, security events, and log entry content that should be captured by EHR logging mechanisms? The three documents that suggest the most data transactions, security events, and log entry content, overall, are HL7 (87 items extracted), IHE ATNA (59 items extracted), and NEMA (41 items extracted). The HL7 and IHE ATNA specifications were very similar in phrasing and content, suggesting a strong relationship between the two documents. Likewise, the NEMA specification was referenced multiple times within the HL7 document. Figure 1, presented previously in this paper, graphically displays these relationships.

For data transactions, in particular, the top three documents are IHE ATNA (39 data transactions extracted), HL7 (35 data transactions extracted), and CCHIT (9 data transactions extracted). For security events, the top three documents are HL7 (36 individual

security events extracted), NEMA (29 individual security events extracted), and Chuvakin & Peterson's "How to Do Application Logging Right" (20 individual security events extracted). For log entry content, the top documents are HL7 (16 data elements extracted), "How to Do Application Logging Right" (9 data elements extracted), and NIST (9 data elements extracted).

While ASTM E2147 is strongly referenced by meaningful use standards as a primary resource for logging mechanism implementation requirements, the ASTM E2147 document includes only 17 data transactions, security events, and log entry content. Most of the ASTM E2147 data transactions have very general descriptions compared to HL7 and IHE ATNA. For example, ASTM E2147 specifies that EHR systems "Record or report type of access (authentication, signoff, queries, views, additions, deletions, changes)". However, HL7 presents much more specific data transaction descriptions, including creating, updating, merging, and unmerging subject of care record accesses; creating, admitting, and updating encounters or visits; and initiating, updating, completing, or canceling orders and order sets. While ASTM E2147 presents very broad, general CRUD actions that should be logged, HL7 and IHE ATNA include specific data objects and specify which CRUD actions apply for each data object, giving a more detailed, comprehensive understanding of what data transactions should be logged.

ASTM E2147 also presents a limited list of security events that should be logged: user logins, user logouts, printing/faxing, and copying. The ASTM E2147 document does not include software, network, or administrative events that should be logged. Therefore, an EHR system developer solely relying on ASTM E2147 may overlook logging security events that are vital to reconstructing user behavior. However, ASTM E2147 was also the only document out of the 16 total documents to include the copying of data as a security event. Therefore, we conclude that no single document should be considered when implementing EHR system logging mechanisms. Instead, developers should consider a comprehensive catalog of multiple documents and resources that specify data transactions, security events, and log entry content that should be captured by logging mechanisms.

RQ6: What minimal set of documents covers 100% of the cataloged data transactions, security events, and log entry content? With 16 source documents, 11 data transactions, 77 security events, and 22 data elements for log entry content, we wanted to identify a minimal set of documents necessary to cover 100% of the extracted items.

For **data transactions**, HL7 covers 10 out of the 11 transactions identified. For complete coverage of the data transactions, developers would have to also consider any one of the following sources: ASTM, CCHIT, or IHE ATNA. Therefore, the minimum set of documents needed to discover 100% of the cataloged data transactions is:

$\mathbf{S_{data\_transactions}}$**:** {HL7 && (ASTM || CCHIT || IHE ATNA)}

For **security events**, the minimum set of documents required for 100% coverage is large. However, Meaningful Use Stage 1, NIST, and HITSP offered no security events in this study. HL7 covers only 34 out of the 77 total security events. NEMA covers an additional 18 security events. IEEE covers an additional nine events. DoD 5200.28 covers an additional four events. CCHIT covers an additional three events. CHI covers an additional three events. HALR covers an additional two events. MU2 covers an additional two events. OB covers an additional one event. Finally, ASTM E247 covers an additional one event. Therefore, the minimal set of documents required to discover 100% of the 77 security events (listed in descending order by number of security events contributed) is:

$\mathbf{S_{security\_events}}$**:** {HL7, NEMA, IEEE, DoD 5200.28, CCHIT, CHI, HALR, MU2, OB, ASTM}

For **log entry content**, HL7 covers 12 out of 22 data elements. CHI covers an additional four data elements. SANS covers an additional two data elements. ASTM, HALR, IHE ATNA, and NIST each cover an additional one data element apiece. Therefore, the minimal set of documents required to discover 100% of the 22 data elements for log entry content (listed in descending order by number of log entry content data elements contributed) is:

$\mathbf{S_{log\_entry\_content}}$**:** {HL7, CHI, SANS, ASTM, HALR, IHE ATNA, NIST}

Overall, to obtain a complete set of *all* data transactions, security events, *and* log entry content, the union of $S_{data\_transactions}$, $S_{security\_events}$, and $S_{log\_entry\_content}$ must be calculated. Therefore, the following set of 13 out of the 16 total documents (in alphabetical order) must be considered when designing logging mechanisms:

$\mathbf{S_{All}}$**:** {ASTM, CCHIT, CHI, DoD 5200.28, HALR, HL7, IEEE, IHE ATNA, MU2, NEMA, NIST, OB, SANS}

Figure 2 presents the breakdown of coverage by the documents in $\mathbf{S_{All}}$. For 50% coverage of collected data transactions, security events, and log entry content, only {HL7} must be considered. For 80% coverage, {HL7, NEMA, IEEE, and CHI} must be considered. For 90% coverage, {HL7, NEMA, IEEE, CHI, DoD, HALC} must be considered.

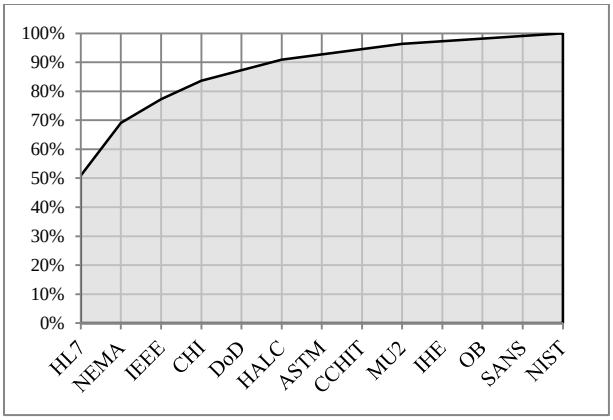

Figure 2. Graph showing coverage of the cataloged data transactions, security events, and log entry content for $\mathbf{S_{All}}$, sorted in increasing order by percentage contributed.

6 Limitations

Our cataloged set of 16 documents may not fully represent all the necessary standards, requirements, regulations, and logging mechanism specifications available in the field. However, we intend the set of 16 documents to be an initial, comprehensive, and cross-domain representation of logging mechanism specifications that can be expanded over time with additional documentation.

Additionally, we may have incorrectly interpreted certain descriptions of the data transactions, security events, and log entry content during categorization. However, software developers often encounter such ambiguous, high-level descriptions of some standards and regulations when developing software. In this study, we erred on the side of caution and did not group any items that were ambiguous.

Likewise, HL7 and IHE ATNA provide lists of specific PHI data objects along with the CRUD transactions that apply to each data object. However, our current methodology diminishes the benefit of having the specific list of PHI data objects and the CRUD actions that apply to each. In previous work [24], we concluded that specific auditable events (like those presented in HL7 and IHE ATNA) are more beneficial when implementing and evaluating EHR logging mechanisms.

Finally, nonrepudiability involves more than just data transactions, security events, and log entry content. If log entries can be fabricated, altered, deleted, or tampered with in any way, the log is untrustworthy and no longer offers solid proof to counter repudiation threats. Immutability concerns should be addressed by both the software (such as through the use of digital signatures or provenance tracking), as well as organizational policies. Organizational policies for immutability may include restricting physical access to servers that store PHI, or policies that require the use of write-once, read-many hardware components for log storage.

7 Future Work

In healthcare [25], viewing of PHI in EHR systems is a key concern with regard to privacy of an individual's sensitive medical data. However, a 2004 report by the President's Information Technology Advisory Committee suggests that many healthcare software systems are not configured to record accesses or viewing of data [26]. Since our current catalog diminishes the benefit of PHI data objects and the CRUD actions that apply to each, one avenue of future work involves developing a systematic process for identifying both the data objects and CRUD actions that apply to each. This process would be tailored to the specific EHR system to obtain a relevant, accurate, and system-specific representation of the EHR system's possible data transactions.

In addition, we plan to use our catalog as a basis for developing and validating a set of software security metrics to help measure the degree to which a software system's logging mechanism promotes nonrepudiability. However, logging mechanism nonrepudiability includes additional considerations, such as immutability of the log files (making sure log entries cannot be altered, fabricated, or deleted), reliability of timestamps, and log backup procedures. We plan to incorporate these considerations when developing our security metrics for logging mechanisms.

8 Conclusion

Standardization organizations and/or governing bodies need to create more comprehensive, centralized specifications for logging mechanisms, especially for outlining security events that developers should consider. In the meantime, we create an initial catalog of 11 data transactions, 8 categories of security events (77 individual security events), and 10 categories of log entry content (22 individual data elements) that should all be considered when implementing logging mechanisms in EHR systems.

Overall, EHR developers should not rely on a single document to provide an adequate, comprehensive list of data transactions, security events, and log entry content. While all 11 data transactions cataloged may be extracted from just two of the 16 total documents, both security events and log entry content required at least eight out of the 16 documents to extract 100% of the security events and log entry content cataloged in this study. Even though a prominent standard such as ASTM E2147 identifies all of the atomic CRUD data transactions, it fails to identify 94.8% of the total security events cataloged in this study. Furthermore,

while the healthcare documents, as a whole, adequately identify data transactions, they fail to identify 16 out of 77 security events in our catalog.

Each EHR system contains different features and security considerations. We do not claim that *all* EHR system developers should implement *all* 11 data transactions, 77 security events, and 22 log entry content data elements cataloged. However, EHR system developers should consider the widest variety of suggested data transactions, security events, and log entry content possible, then decide whether each applies and is appropriate to log in the individual EHR system.

Acknowledgements

This work is supported by the USA National Security Agency (NSA) Science of Security Lablet. Any opinions expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSA. We also thank the RealSearch group for helpful feedback on this research.

References

[1] S. Hernan, S. Lambert, T. Ostwald, and A. Shostack. (2006, October 30). *Uncover Security Design Flaws Using the STRIDE Approach*. Available: http://msdn.microsoft.com/en-us/magazine/cc163519.aspx

[2] "Medicare and Medicaid Programs; Electronic Health Record Incentive Program Stage 1," in *45 CFR Parts 412, 413, 422, and 495* vol. 75, ed: United States Department of Health and Human Services, 2010.

[3] "Medicare and Medicaid Programs; Electronic Health Record Incentive Program Stage 2," in *42 CFR Parts 412, 413, and 495* vol. 77, ed: United States Department of Health and Human Services, 2012.

[4] "Electronic Health Record Privacy and Security Requirements," in *Version 1.1*, ed: Canada Health Infoway, 2004.

[5] (2011, April 9, 2012). *Meaningful Use Definition & Objectives*. Available: http://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives

[6] "Common Audit Message Version 2.x," ed: Health Level Seven International, 2001.

[7] J. King, B. Smith, and L. Williams, "Audit Mechanisms in Electornic Health Record Systems: Protected Health Information May Remain Vulnerable to Undetecte Misuse," *International Journal of Computational Models and Algorithms in Medicine*, vol. 3, p. 19, April-June 2012.

[8] A. Chuvakin and G. Peterson, "How to do application logging right," *Security & Privacy, IEEE*, vol. 8, pp. 82-85, 2010.

[9] (2011). *CCHIT Certified 2011 Ambulatory EHR*. Available: http://www.cchit.org/certify/2011/cchit-certified-2011-ambulatory-ehr

[10] (2007). *Information system audit logging requirements*. Available: http://www.sans.org/security-resources/policies/info_sys_audit.pdf

[11] "IEEE standard for information technology: Hardcopy device and system security," ed: IEEE Standard, 2008, pp. 1-177.

[12] "Common Weakness Enumeration: A Community-Developed Dictionary of Software Weakness Types," ed: MITRE Corporation, 2013.

[13] "Logging Cheat Sheet," ed: The Open Web Application Security Project (OWASP), 2013.

[14] "National Vulnerability Database Version 2.2," ed: United States Department of Homeland Security National Cyber Security Division/Us-CERT.

[15] A. Chuvakin and G. Peterson, "Logging in the age of web services," *IEEE Security and Privacy*, vol. 7, pp. 82-85, 2009.

[16] "T15: Collect and Communicate Security Audit Trail Transaction," in *Version 1.1*, ed: Healthcare Information Technology Standards Panel, 2007.

[17] "NIST Special Publication 800-53: Security and Privacy Controls for Federal Information Systems and Organizations," ed: National Institute for Standards and Technology, 2013.

[18] "E2147-01 Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems," ed: American Society for Testing and Materials, 2013.

[19] "Payment Card Industry Data Security Standard," in *Requiremens and Security Assessment Procedure, v2.0*, ed: Payment Card Industry Security Standards Council, 2010.

[20] "Department of Defense Standard: Trusted Computer System Evaluation Criteria," ed: United States Department of Defense, 1985.

[21] "Department of Defense Directive 5200.28: Security Requirements for Automated Information Systems," ed: United States Department of Defense, 1988.

[22] "Audit Trail and Node Authentication," ed: Integrating the Healthcare Enterprise, 2012.

[23] "Security and Privacy Auditing in Health Care Information Technology," ed: Joint National Electrical Manufacturers Association, European Coordination Committee of the Radiological and Electromedical Industry, and Japan Industries Association of Radiological Systems Security and Privacy Committee, 2001.

[24] J. King, B. Smith, and L. Williams, "Modifying Without a Trace: General Audit Guidelines are Inadequate for Electronic Health Record Audit Mechanisms," presented at the ACM SIGHIT International Health Informatics Symposium, Miami, Florida, USA, 2012.

[25] "Health Insurance Portability and Accountability Act," ed: United States Department of Health & Human Services, 2007.

[26] "Revolutionizing health care through information technology," National Coordination Office for Information Technology Research and Development, Arlington, Virginia, USA2004.