Identifying Vulnerabilities and Critical Requirements Using Criminal Court Proceeding

Travis D. Breaux[1], Jonathan D. Lewis[1], Paul N. Otto[1,2], Annie I. Antón[1]

Department of Computer Science, North Carolina State University[1]
School of Law, Duke University[2]
{tdbreaux,jdlewis,pnotto,anton}@ncsu.edu

ABSTRACT

Information systems governed by laws and regulations are subject to civil and criminal violations. In the United States, these violations are documented in court records, such as complaints, indictments, plea agreements, and verdicts, which thus constitute a source of real-world software vulnerabilities. This paper reports on an exploratory case study to identify legal vulnerabilities and provides guidance to practitioners in the analysis of court documents. As legal violations occur after system deployment, court records reveal vulnerabilities that were likely overlooked during software development. We evaluate established requirements engineering techniques, including sequence and misuse case diagrams and goal models, as applied to criminal court records to identify mitigating requirements that improve privacy protections. These techniques, when properly applied, can help organizations focus their risk-management efforts on emerging legal vulnerabilities. We illustrate our analysis using criminal indictments involving the U.S. Health Insurance Portability and Accountability Act (HIPAA).

Categories and Subject Descriptions

D.2.1 Requirements/Specifications, Methodologies

General Terms

Design, Security, Legal Aspects

1. INTRODUCTION

The requirements of information systems are increasingly affected by U.S. government laws and regulations such as the Health Insurance Portability and Accountability Act (HIPAA)[1], which governs the privacy of electronic patient health information. For example, an emerging problem affecting information systems involves insider threats in which an employee or contractor engages in activities that create civil or criminal liability. In two successive years, the annual E-crime Watch Survey revealed that over one-third of surveyed security executives and law enforcement officials identified insider threats as the greatest cause of damage to information systems [18]. Herein, we illustrate how to identify critical requirements using examples from a criminal case, United States v. Ferrer, that details insider threats to patient medical records involving a HIPAA violation [6].

[1] Pub. L. No. 104-191, 110 Stat. 1936 (1996)

This paper offers three contributions: (1) background on the U.S. federal legal process and terminology that requirements engineers must know to replicate this analysis; (2) our experiences and time-saving guidance for future researchers on acquiring relevant court documents needed to perform this analysis; and (3) a comparative evaluation of three notations that we use to analyze legal vulnerabilities with examples that link vulnerabilities to mitigating requirements.

The remainder of the paper is organized as follows: Section 2 reviews related work; Section 3 provides background including relevant legal terminology, procedures and documents from the U.S. federal court system; Section 4 presents the research method used to conduct this study; Section 5 shows our analysis results comparing the different notations; and Section 6 concludes with future work.

2. RELATED WORK

Misuse and abuse cases are used in security to elicit and reason about system vulnerabilities. McDermott and Fox first defined abuse cases as user-system interactions that result in harm to the system [12]. Engineers use the abuse case notation to elicit these interactions from customers and document them. A related concept to the abuse case is the misuse case introduced by Sindre and Opdahl [15]. An important difference from abuse cases is that misuse cases provide links to security use cases that are intended to mitigate the case for misuse. Misuse cases have been popular in engineering practice as important artifacts in early requirements [15] and for performing trade-off analysis [1] and risk analysis [16]. This is the first time they have been used to identify legal vulnerabilities.

In requirements engineering, goals describe intended states to be maintained or achieved by the system [5]. Anti-goals are resolved by creating new goals to mitigate or prevent the obstructing goals [11]. Regnell et al. propose using hierarchical use case models to iteratively decompose goals into sequential user actions using use case, flow and sequence diagrams [14]. In this paper, we contrast the use of misuse case and sequence diagrams with KAOS diagrams for representing legal vulnerabilities using anti-goals.

In recent years, researchers have been drawn to the challenges that legal requirements pose to information systems [13]. Challenges include accurately acquiring legal requirements and maintaining traceability [3, 4]. To the authors' knowledge, this paper reports the first time that legal

vulnerabilities or software requirements have been identified in criminal case law.

3. LEGAL BACKGROUND

To help the reader understand the context surrounding criminal proceedings, we provide a cursory overview of the process governing U.S. criminal law. A few simplifying assumptions have been made for clarity and brevity. The discussion focuses on the federal court system, which is governed by the Federal Rules of Criminal Procedure (FCRP)[2] [7]; various states may have different rules at each stage of the process. We describe how the federal system handles non-capital felonies, glossing over differences in misdemeanor proceedings. Many white-collar crimes are initially investigated by administrative agencies, rather than police; the discussion briefly notes how the process differs in such cases. Our discussion emphasizes phases involving the specific court documents analyzed for this paper.

Once a violation is suspected, the investigative process begins: enforcement officials (e.g., police, prosecutors) determine whether a crime was committed, identify the perpetrators, gather evidence linking the perpetrators to the crime, and locate the perpetrators [10]. If substantial evidence exists linking the suspects with the crime, official charges in the form of a *criminal complaint* — a formal document accusing a suspect of committing some criminal act — is typically filed [8]. The complaint, governed by FCRP 4 [7], tends to be a brief assessment of the specific acts performed by the accused that constituted a criminal statute's violation [10]. At this stage, the accused becomes a defendant and the court creates an official *docket*, a record of all proceedings and filings involved in the case [8]. With the passage of the E-Government Act[3], all federal courts must provide online access to court information, including full case dockets.

In many cases, a grand jury determines whether there is sufficient evidence against the defendant to justify advancing to trial [10]. If the grand jury finds the evidence sufficient, it can issue an *indictment*, or "formal written accusation" of criminal conduct [8]. The indictment, specified by FCRP 7(c), "shall be a plain, concise and definite statement of the essential facts constituting the offense charged" [7]. The indictment provides the first full account of the suspected violation(s) and details of substantiating facts. The indictment supersedes the complaint as an account of the suspected criminal act [10].

After an indictment is issued, the defendant may enter a plea at arraignment. After arraignment, *plea bargaining* may begin in which a substantial majority of defendants will exchange a guilty plea for reduced charges or lesser sentencing [10]; if accepted, the offer will be detailed in a *plea agreement*.

Defendants have a right to a *jury trial* in all felony prosecutions; a trial by judge is called a *bench trial* [10]. If a guilty verdict is entered, whether by judge or jury, a judge generally determines the defendant's sentence. Probation officers will provide a *presentence investigation report* for sentencing purposes, which details the "convicted defendant's educational, criminal, family, and social background" [8]. There are three broad categories of sanctions: restitution, probation, or incarceration [10].

4. CASE STUDY DESIGN

This research employed an exploratory, multi-case study design [19] to answer a two-part research question: can we identify software vulnerabilities from civil and criminal cases and, if so, which documents are most relevant and which notation best represents the information contained in relevant case materials? In this section, we describe the case study materials, the units of analysis and analysis procedure.

4.1. Case Study Materials

Relevant civil and criminal cases and corresponding court documents can be identified and acquired in different ways. The U.S. federal government centrally manages federal court records through the Public Access to Court Electronic Records (PACER) database system. In addition, privately managed databases, such as LexisNexis and WestLaw, contain numerous court records. These databases charge a subscription-based or per-page fee to retrieve court documents. The per-page fee includes the number of pages for each requested document, including pages from search results. The cost of keyword searches, as opposed to looking up specific case numbers, may be prohibitive for businesses or engineers with a small discretionary budget.

A lower-cost, indirect method to identify cases is through news reports and press releases. Due to the recency of HIPAA, we used the relevant sections of the United States Code (U.S.C.) and Code of Federal Regulations (C.F.R.) for HIPAA to identify cases in news reports, and then cross-referenced the case number in PACER. In addition, we scanned press releases from the regional offices of the U.S. Department of Justice for indictments and convictions.

For this study, we chose to examine recent criminal cases that include at least one violation of HIPAA regulatory law, given our experience in analyzing HIPAA regulations. Using PACER, we acquired the full dockets for the following seven cases, which were identified as the only seven HIPAA-related court cases to date by an Assistant U.S. District Attorney for the Western District of Washington [17]:

1. United States v. Gibson – a hospital insider acquires patient medical records to commit wire fraud.
2. United States v. Ferrer – an insider acquires patient medical records to commit Medicare fraud.
3. United States v. Hungerford – a health insurance insider acquires patient medical records to commit wire fraud.
4. United States v. Occident– a hospital insider acquires patient medical records to commit wire fraud.
5. United States v. Ramirez – a primary care provider insider attempts to sell a patient medical record to a drug trafficker.
6. United States v. Williams – a healthcare clearinghouse insider acquires and sells patient medical records.
7. United States v. Williams and Adjei – a healthcare clearinghouse insider acquires patient medical records to file fraudulent tax returns.

[2] Abbreviated as "Fed. R. Crim. P." in legal works.

[3] Pub. L. No. 107-347, § 205, 116 Stat. 2899, 2913-15 (2002)

Each case includes a distinct docket for each defendant, resulting in 22 dockets in all. Reviewing each docket, there are a total of 1141 entries; Table 1 presents a subset of the 238 different types of entries we identified from these dockets. Among these eight cases, only four originated with official complaints.

Table 1: Types of available documents for the seven cases examined in this study

Type of Docket Entry	No. of Entries
Complaint	8
Indictment	38
Plea Agreement	16
Transcript	25
Minute Entry	154
Judgment	26

As our analysis focuses on identifying legal vulnerabilities affecting software, we selected documents where software systems might directly or indirectly be used to commit the crimes as charged. We focused our analysis primarily on the indictments from each case, or secondarily on the plea agreements if an indictment was not available (as was the case in United States v. Gibson). Several different versions of an indictment may exist; this explains the 38 indictment entries in the dockets for only 22 defendants. The various types of indictments are described in Section 3. We found that plea agreements contain no more detail than the corresponding indictments and, in fact, contain additional information irrelevant to this analysis (e.g., waivers of rights, penalties imposed). Lastly, we found that sentencing transcripts can be used to generally rate the case's severity.

4.2. Units of Analysis and Procedure

The units of analysis for this study consist of descriptions of actors, actions and events involved in criminal charges. These units were prescribed by our choice of notations: sequence, misuse case and KAOS diagrams. This limited focus may have caused us to overlook other important features that are relevant to develop legally compliant software.

The analysis procedure was performed in two passes over selected court documents by two researchers working in tandem. The first pass identifies actors and events using heuristics from the Goal-Based Requirements Analysis Method [2]. The second pass is limited to parts of the document in which events are identified and is repeated for each notation. This repetition, as opposed to deriving one diagram from another, avoids bias introduced by the limitations of any one notation. Domain-specific linguistic devices in the text and limitations in the notation are identified and recorded for discussion. Finally, to check consistency and completeness, the actors from the first pass are cross-checked with the actors from the second pass to identify missing events.

5. FINDINGS AND DISCUSSION

We illustrate the results of our analysis using criminal indictments to identify real-world software vulnerabilities.[4] Our objective was to develop software requirements that will thwart future insider efforts to exploit these vulnerabilities. This analysis entailed deriving sequence, misuse case and KAOS diagrams from the indictments and charges. We introduce the notations and illustrate the analysis using United States v. Ferrer (Case 2 from Section 4.1) that describes an insider threat [6]. Consider the following excerpt from the corresponding indictment, paragraph (6) in which actors are *italicized* and events are underlined:

"6. From on or about May 23, 2005, and continuing through on or about June 26, 2006, at Broward County, in the Southern District of Florida, and elsewhere, the defendants,

FERNANDO FERRER, JR.,
and
ISIS MACHADO,

did knowingly and willfully combine, conspire, confederate, and agree with *each other* and with *others* known and unknown to the Grand Jury, to defraud the *United States* and to commit certain other offenses against the *United States*, namely:

a. to knowingly and with intent to defraud, exceed authorized access to a protected computer, and by such conduct further the intended fraud to obtain things of value exceeding $5,000, in violation of Title 18, United States Code, Sections 1030(a)(4) and (c)(3)(A);
b. during and in relation to a felony violation of Title 18, United States Code, Chapter 47, to wit, Title 18, United States Code, Sections 1030(a)(4) and (c)(3)(A), to knowingly possess and use, without lawful authority, a means of identification of *another person*, in violation of Title 18, United States Code, Section 1028A(a)(1); and
c. to knowingly and for a reason other than permitted by Title 42, United States Code, Chapter 7, Subchapter XI, Part C, obtain individually identifiable health information relating to an *individual*, with the intent to sell, transfer, and use, and cause to be used, individually identifiable health information for personal gain, in violation of Title 42, United States Code, Sections 1320d-6(a)(2) and (b)(3)."

This excerpt highlights several key findings observed throughout all cases identified in Section 4.1. First, paragraph (6) begins with the dates of the violations, reported as a period of time, and the summary violations in paragraphs (6)(a)-(c) do not include specific dates. At this point in the legal process, the exact dates may not be known. Second, the number of parties involved in the violation may not be known, as illustrated in the above excerpt "others . . . unknown to the Grand Jury." This missing information affects the quality of scenario and goal analysis in different ways, which we discuss here. Finally, the indictments trace from each violation in paragraphs (6)(a)-(c) to specific paragraphs in corresponding laws that were violated. These references indicate potential "hotspots" in regulations that

[4] DISCLAIMER: Statements made in this paper are intended to reflect the actual charges stated in the indictments and are not intended to suggest guilt or innocence of the defendants.

can be used to prioritize requirements by surveying multiple indictments.

An important observation not shown in this excerpt is that subsequent, numbered paragraphs include backward references to this paragraph. These cross-references are used to refer back to details that are shared across these different contexts, including actors and events. Similar to the regulatory analysis method employed by Breaux et al. [3], analysts must incorporate these details in each new context to accurately represent the individual charges.

5.1. Sequence Diagrams

Sequence diagrams are an Object Management Group (OMG) standard included in the popular Unified Modeling Language (UML). Using sequence diagrams, engineers can describe the functions of individual objects in a linear-time, total-order notation; see related work on state charts for a partial-order notation that supports concurrency [8]. Because engineers are familiar with sequence diagrams, others have used this notation to describe scenarios and the actions of actors in an analogous manner [8].

Figure 1 shows a sequence diagram acquired from paragraph (6), above. The disadvantage of sequence diagrams, observed in modeling this excerpt and other indictments we considered, is the missing temporal information required to create event sequences. For example, in paragraphs (6)(a)-(c), the engineer must infer the order of the events "exceed authorized access," "possess and use" and "sell, transfer and use" presented in Figure 1. These inferences include deciding that the phrase "things of value" in paragraph (6)(a) includes both "a means of identification of another person" and "individually identifiable health information" in paragraphs (6)(b) and (6)(c), respectively, which may or may not be accurate.

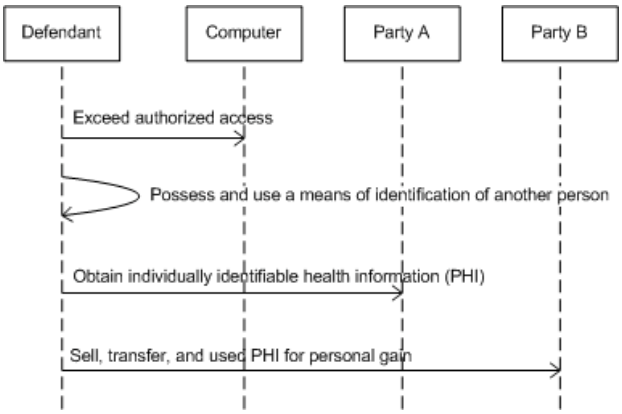

Figure 1: Potentially Inaccurate sequence diagram from paragraph (6), United States v. Ferrer

5.2. Misuse Case and KAOS Diagrams

Sindre and Opdahl introduced the misuse case diagram [15]. In misuse case diagrams, actors are linked to misuse cases that represent misuses of the system. Figure 2 presents the same subset of events from paragraph (6) that appear in the sequence diagram in Figure 1. In Figure 2, the phrase "commit certain offenses against the United States" from paragraph (6)(a) is mapped to a misuse case and refined by the sub-cases "exceed authorized access to a protected computer" from paragraph (6)(a) and "possess and use a means of identification" from paragraph (6)(b). We identified these sub-cases using the phrase heuristics for identifying purposes and instruments in an activity description [3].

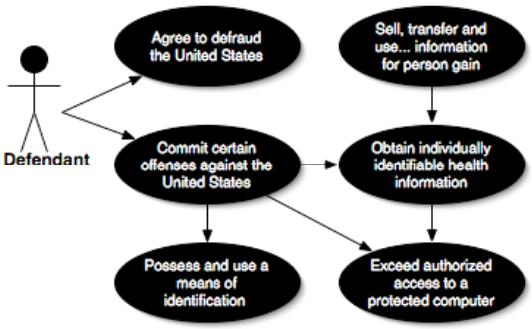

Figure 2: A misuse case diagram from paragraph (6), United States v. Ferrer

Similar to misuse case diagrams, KAOS diagrams can represent anti-goal and associated threat agents [11]. Unlike misuse case diagrams, anti-goals are refined using logical "AND" and "OR" nodes to represent possible alternatives. During refinement of an anti-goal model, the threatening agents are refined to be responsible for a leaf level anti-goal. Figure 3 presents the same subset of events from paragraph (6) that appeared in Figures 1 and 2, this time using the KAOS method to represent anti-goals.

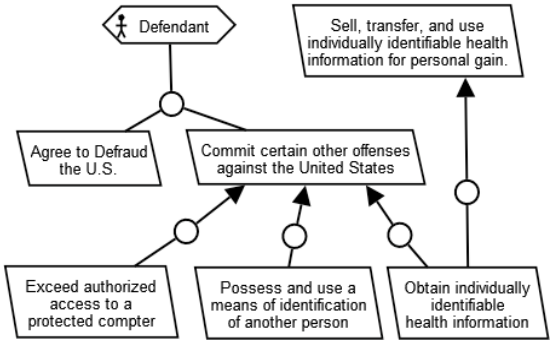

Figure 3: A KAOS diagram from paragraph (6), United States v. Ferrer

To illustrate the benefit of refinement using "AND" and "OR" links, we separate the events "sell, transfer and use" into three anti-goals. The English conjunction "and" is logically ambiguous and can be interpreted as a logical disjunction [3]. Thus, the analyst can create mitigating requirements for each event as if they occur independently. For example, one can prevent "use" by encrypting the information, assuming the threat agent does not have the means to decrypt the information. However, sales and transfers are not prevented or mitigated by encryption; thus another goal is required to prevent or mitigate these threats. Alternatively, misuse case diagrams provide the "excludes" link, which we did not employ in this study, but which may be relevant in the analysis of other indictments.

6. CONCLUSION AND FUTURE WORK

This study demonstrates that criminal case documents, notably complaints and superseding indictments, are rich sources for identifying legal vulnerabilities and that all of sequence, misuse case and KAOS diagrams present different strengths in this new domain. To a great extent, the applicability of each diagram notation depends heavily on the presence or absence of relevant information.

In Section 5.1, we observe an instance in which sequence diagrams cannot be accurately created from criminal indictments. While not always the case, we generally found that misuse case and KAOS diagrams could represent actors and events using refinement hierarchies. As discussed in Section 5.2, the notable difference was that KAOS diagrams provide an additional distinction through "AND" and "OR" refinement links that are necessary to capture the exclusivity of separate charges described in the indictment, despite the appearance of events occurring in conjunction to achieve some overall goal. This approach is still amenable to identifying mitigation strategies that address a single event, even if the single event occurs repeatedly. For example, the act of exceeding authorized access to health information, as a single event, may be difficult to mitigate. This is especially true if observable behavior includes authorized access that is normally granted to the malicious user. However, as a sequence of multiple, similar-type events, it may be possible to discern that access exceeds normal behavior using operational profiles, in which normal frequency of use helps highlight behaviors that are out of the norm.

However, goal-oriented models that do not express temporal relations, such as the misuse case and KAOS diagrams that we examined, will fail to capture a class of vulnerabilities that is only exploitable through transactions or a sequence of dissimilar events. For example, the acts of exceeding authorized access (for unauthorized purposes and in numbers beyond the average operational profile) to health information and subsequently using the information to file fraudulent insurance claims is a complex vulnerability: the health information is usable to file claims independent of the health care provider maintaining the information. Requiring that claims be filed using a secret known only to the provider and the agency, called a *shared secret*, mitigates this vulnerability as it renders the act of acquiring the information useless in the process of filing claims. The insurance agency would thus reject claims filed without the shared secret. Observing the applicability of this mitigation strategy, however, benefits from the explicit representation of temporal relations between events. While trial transcripts may contain this information, the value of expending this additional effort to analyze these transcripts must be determined by future work.

The limits of this study reveal fertile ground for future work. For example, this case study did not examine cases that went through the U.S. federal appeals process. Cases that are appealed are used to decide legal precedent and constitute an extension or retraction to statutory and/or case law. The decisions in these cases can be used to reinforce prior decisions regarding known vulnerabilities or to yield insight into new vulnerabilities through new interpretations of existing laws.

7. ACKNOWLEDGEMENTS

We thank Professor Sara Sun Beale of Duke University School of Law for her feedback. This work was funded by the IBM PhD Fellowship (RTP CAS) and NSF #032-5269 and NSF #043-0166.

8. REFERENCES

[1] I. Alexander, "Initial industrial experience of misuse cases in trade-off analysis," *IEEE Joint Int'l Conf. Req'ts Engr.*, pp. 61-68, 2002.

[2] A.I. Antón. *Goal-based Requirements Analysis Method*, PhD Thesis, Georgia Tech, 1996.

[3] T.D. Breaux, M.W. Vail, A.I. Antón. "Towards compliance: extracting rights and obligations to align requirements with regulations," *IEEE Int'l Conf. Req'ts Engr.*, pp. 49-58, 2006.

[4] T.D. Breaux, A.I. Antón. "Analyzing regulatory rules for privacy and security requirements," *IEEE Trans. Soft. Engr., Special Issue on Soft. Engr. for Secure Sys.*, 34(1): 5-20, 2008.

[5] A. Dardenne, A. van Lamsweerde, S. Fickas. "Goal-directed requirements acquisition", Science of Computer Programming. 20:3-50, 1993.

[6] United States v. Ferrer, et al. Case No. 0:06-CR-60261-JIC, S.D. Fl., Dec. 7, 2006.

[7] *Federal Rules of Criminal Procedure*, as amended December 2007.

[8] B.A. Garner, Ed., *Black's Law Dictionary*, 8th ed., Thompson West, 2004.

[9] M. Glinz. "Improving the quality of requirements with scenarios", *2nd World Congress for Software Quality*, 55-60, 2000.

[10] Y. Kamisar et al. *Modern Criminal Procedure: Cases, Comments, and Questions*, 11th ed., St. Paul, Minn.: Thomson/West, 2005, pp. 2-20.

[11] A. van Lamsweerde, "Elaborating security requirements by construction of intentional anti-models," *IEEE 26th Int'l Conf. Soft. Engr.*, pp. 148-157, 2004.

[12] J. McDermott, C. Fox, "Using abuse case models for security requirements analysis", 15th Computer Security Applications Conf.,, pp. 55-64, 1999.

[13] P.N. Otto, A.I. Antón, "Addressing legal requirements in requirements engineering," *15th IEEE Int'l Req'ts Engr. Conf.*, pp. 5-14, 2007.

[14] B. Regnell, M. Andersson, J. Bersrand. "A hierarchical use case model with graphical representation", *IEEE Int'l Symp. and Workshop on Engr. of Computer-based Sys.*, pp. 270-277, 1996.

[15] G. Sindre, A.L. Opdahl. "Eliciting security requirements with misuse cases", Req'ts Engr. 10:34-44, 2005.

[16] D. Verdon, G. McGraw, "Risk analysis in software design," IEEE Security & Privacy, 2(4): 79-84, 2004.

[17] P. Winn, "Confronting the threats of medical identity theft," *Health Information Privacy/ Security Alert*, July 24, 2007.

[18] S. Yanovitch, K. Kimberland, "2007 E-crime watch survey shows security incidents, electronic crimes and their impact steady versus last year," *CSO Magazine*, Sep. 2007.

[19] R.K. Yin. *Case Study Research*, 3rd ed. Applied Social Research Methods Series, v.5, Sage Pubs., 2003.