

Dartmouth College Computer Science Technical Report TR2001-403



A site to classify and organize the risks of performing business on the
Internet

by
Aidan Stanley Marcuss

Honors Thesis

Advisors: Fillia Makedon, Carey E. Heckman

Department of Computer Science
Dartmouth College
Hanover, New Hampshire

June 1, 2001

Abstract

As the use of the Internet and other computer networks to transact business grows, there is an ever increasing need for those taking part in those transactions to understand the risks of doing so. While there are many web sites that have created valuable databases of specific vulnerabilities for certain types of hardware and software, there is a lack of focus on attempting to analyze the interaction of businesses, their systems, computer networks, and their customers and the risks that are created by either intended or unattended interactions. EcomRISK.org is a web site that presents a clear taxonomy to classify these risks and provides other features to aid in the general discussion of e-commerce risk. The site, and the taxonomy at the center of it, creates a database of these incidents so they can be clearly searched. This paper discusses the creation of EcomRISK.org, from vision to birth.

Contents

1	Introduction	1
1.1	Project History	3
1.2	Acknowledgments	6
1.3	Technical Note	7
2	Business Analysis	8
2.1	Market Analysis	8
2.1.1	PEST Analysis	8
2.1.2	SWOT Analysis	10
2.2	Mission Statement and Purpose	12
2.2.1	Mission Statement	12
2.2.2	Purpose	12
2.3	Technological and System Goals	13
3	Basic Site Architecture	15
4	The EcomRISK.org Taxonomy	18
4.1	Risk Incident Submission Form	18
4.1.1	Submitter Information	19
4.1.2	Incident Information	19
4.1.3	Solution Information	24
4.1.4	Quantifiable Ramifications	26
4.1.5	Company Information	27
5	Other EcomRISK.org Tools	28
5.1	Forum	28
5.2	In the News...	29
5.3	Working Papers	29

5.4	Resources	30
5.5	Other GREeCOM.org Sites	31
5.5.1	eJETA.org	31
5.5.2	DePolicy.org	32
6	Conclusion	33
6.1	Project Status and Future Work	33
6.2	Personal Conclusions	33
	Bibliography	34

List of Figures

1.1	A screen-shot of EcomRISK.org's main page.	4
1.2	GREeCOM.org's Logo	5
1.3	eJETA.org's Logo	5
1.4	DePolicy.org's Logo	6
3.1	A diagram of how the server handles the typical page request .	17

Chapter 1

Introduction

Electronic commerce (e-commerce) has become a critical part of everyone's life, whether they directly realize it or not. More and more businesses and consumers are using the Internet, along with other electronic networks, to transact business. The use of electronic networks to transact business spans from the average consumer buying a book from Amazon.com to General Motors using a web-based parts and service system for their dealers and suppliers [1]. E-commerce has the same potentials for abuse as regular commerce: people can steal, cheat, over-charge, illegally collect personal information, and, in general, defraud one another. However, e-commerce, as a class of business, can fundamentally change the ways these crimes are perpetrated and the frequency of certain types of crimes.

EcomRISK.org is a new web site that creates a on-line community based on the discussion and examination of the risks of performing e-commerce. The site features many useful tools to foster an on-line community. The purpose of the site, as stated in the Professor Makedon's original proposal, is to "document and collect e-commerce risk cases and misuses in a quantifiable manner." [2] The very first question one might ask is, what is meant by e-commerce? Professor Makedon states that "E-Commerce is a new socio-economic force which provides business solutions based on the use of the Internet." [2] A slightly more precise definition that I found on the Web is:

E-commerce may be defined as 'the technology, processing, and operations which occur when business transactions are done automatically over networks, using IT.' In other words, the scheme of the total system, even including operation. Included in the term

e-commerce are transactions between businesses and consumers (B to C), businesses and businesses (B to B), as well as operations internal to companies, the conversion of government procurement to electronic systems, and transactions between consumers. [3]

This definition encompasses all of the prevalent forms of e-commerce that we want to address.

Now we come to the question, what is an e-commerce risk? Professor Makedon states “To understand the risks of electronic commerce specifically, is to understand the risks of abusing or misusing different types of information and at different levels of transactions.” [2]. As a result of the last nine months of work, this definition has become somewhat expanded. An e-commerce risk is the improper use of information at any level of storage or transaction or the violation of the behavior intended by the designers of a network for the clients of the network. The first part of the definition is fairly self explanatory. Business transactions involve the transfer of information. It is the misuse of this information that presents the main risk. The second part of the definition, relating to the violation of intended behavior, is meant to add the class of denial of service (DoS) problems to the realm of those that fall under e-commerce risks. Some examples of typical high-level risks are:

- Security breaches of hardware and software systems supporting e-commerce transactions.
- Un-ethical mining of databases, by both the owners of a database and the users, allowing the person to glean new, private, and not specifically released data about a person.
- Insider problems of company or government information abuse.
- DoS attacks to commercial, government, or private web-sites or other types of information servers.

This list is intended to give just a high overview of what potential risks are. One need not go very far to read about examples of these risks in the real world. A quick browse through the news section¹ of EcomRISK.org can reveal many current examples:

¹This can be found at <http://devlabserver.cs.dartmouth.edu/sites/ecomrisk/news/>. Our main news feed comes from the Institute of Security Technology Studies at Dartmouth College.

- The Computer Emergency Response Team (CERT) at Carnegie Mellon University was hit with a DoS attack lasting for 30 hours during 23rd and 24th of May 2001. [4]
- The official White House web-site (<http://www.whitehouse.gov/>) was also hit with a DoS attack the same week as CERT. [5]
- Data mining is already in heavy use by many corporations, such as “Bank of America, People’s Bank, Sundance, Equifax, Reader’s Digest, Group 1, Marriott, and The Washington Post.” [6]

The above list just scratches the surface of the risks faced by those involved in e-commerce.

The first task in the EcomRISK.org’s creation was to create a taxonomy that would allow us to accept and classify these risks. This taxonomy lies at the heart of the site and took much of our development time. However, the development of the site did not stop with the creation of a classification system. We added a large set of features to the site that we felt would aid in the discussion of e-commerce risks. In Figure 1.1 a sample of the main page of EcomRISK.org is shown.

Over the course of the rest of the paper, I will discuss how we went about creating our site. From a discussion of the business analysis Carey Heckman brought to the group, which helped narrow in on a target audience, to the technological underpinnings of the site, design of the taxonomy, other features of the site, and, finally, the status and future plan for the site.

1.1 Project History

During the fall term of my senior year (2000-01) as an undergraduate majoring in Computer Science at Dartmouth College, I became involved with the department’s DEVLAB and Professor Makedon’s projects on e-commerce. The projects centered around the creation of a number of on-line resources relating to e-commerce and the wide array of issues brought up by this new phenomenon. The parent site, named GREeCOM.org² (GREeCOM.org’s logo is shown in Figure 1.2) , was originally the main focus of development. Our intention was to develop a site the provided a forum to discuss and classify the risks of performing e-commerce, a place to learn about the latest

²GREeCOM stands for Global Research & Education in E-commerce.

The screenshot shows the EcomRISK.org website. At the top, there is a navigation bar with links for Home, Search, Site Map, About Us, and Log Out. The main content area is divided into several sections:

- Introduction:** Not sure what "EcomRISK" is? Click here for a short introduction.
- Forum:** Our special topic for the month of April 2001 is **Medical Privacy**. Add your voice to our forum by clicking [here](#) now.
- Risk Incidents:**
- Working Papers:** Working Paper of the Month: **The Ethics of Online Medical Records** by Charles D. Aboujaoude. Abstract: This paper describes the ethical implications of developing a national online medical database. Such a database would contain the lifetime health records of every U.S. resident by combining information... **Keywords:** Medical records, national online databases, medical privacy, medical ethics, medical technology. **Download the Paper!**
- In the News...:** From **ECOMMERCE TIMES**: **Critical Errors in Online Banking** by Lou Bova. Many financial institutions risk losing market share because they are concentrating too much on developing new online and wireless products, rather than improving the experience customers have on their existing Web sites, according to a report released Tuesday by Jupiter Media Metro. **Full story here...**
- Resources:** Selected Resource for April 2001: **The Berkman Center for Internet & Society at Harvard Law School**. The Berkman Center for Internet & Society is a research program founded to explore alternatives, share its study, and help promote its development.
- Survey Results:** Questions: What is your greatest fear(s) when using the internet for E-Commerce? Results: 54% of our readers said Spamming. **Let me take it!**

The footer contains copyright information (Copyright ©2001 Trustees of **Dartmouth College**. All Rights Reserved), logos for Mac OS & Server, and site maintenance information (Site maintained by webmaster@develab.cs.dartmouth.edu. Last modified: 14 May 2001 20:49pm).

Figure 1.1: A screen-shot of EcomRISK.org's main page.

technological developments relating to e-commerce, and a place to provide a certification of users showing that they had reached some above average level of knowledge about the world of e-commerce. Originally it was our intention to place all of this functionality into one site. However, the group that had assembled to work on the site felt that in order to maximize the value of each piece, we should separate the main functions we were attempting to provide into several sites.

Professor Makedon had laid out her intentions in this area in a proposal to create the EcomRISK³ Data Center. When I joined the project I took over the creation of this data center. We turned her idea of a database

³EcomRISK is a mixture of the term e-commerce and the word risk: the RISK in the name is not an acronym.



Figure 1.2: GREeCOM.org's Logo

into a web site named EcomRISK.org. The other two functions originally planned for GREeCOM.org were separated into their own sites as well, named eJETA.org⁴ (eJETA.org's logo is shown in Figure 1.3) and DePolicy.org⁵



Figure 1.3: eJETA.org's Logo

(DePolicy.org's logo is shown in Figure 1.4). GREeCOM.org became the umbrella site handling user registration and other services common across the three sites.

⁴eJETA stands for the "Electronic Journal on E-commerce Tools & Applications".

⁵The history of the name "DePolicy" is quite lengthy - however it is instructive enough to know that DePolicy.org is "The Center of Learning and Self-Testing on E-Commerce Policies & Technologies".

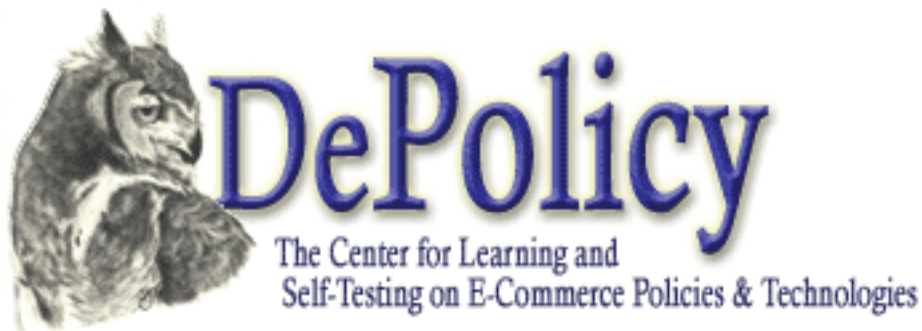


Figure 1.4: DePolicy.org's Logo

I became the main developer EcomRISK.org, however, I would be remiss if I did not give a large amount of the credit for the development of the look and feel to our web master Sue Anne Johnson and our server's administrator Tilmann Steinberg. My intention was to build the site with Professor Makedon's proposal as a guide.

While my main focus was the development of EcomRISK.org, over the course of the last three terms I have been involved in the development of the GREeCOM.org center as a whole. Many of the technologies I developed for the EcomRISK.org site have been adopted for use in the other sites (eJETA.org and DePolicy.org) such as the resources database, the working paper database and viewer, the news system, the forum system, the survey system, and a few others.

1.2 Acknowledgments

I would like to thank several people who have helped with this project:

Fillia Makedon - One of my two thesis advisors and the head of the GREeCOM.org project.

Carey E. Heckman - My other thesis advisor who helped steer much of the business analysis of the project, allowing us to define our goals clearly.

Tilmann Steinberg - A PhD student who served as a system administrator for the whole project and an advisor on countless design

details.

Sue Anne Johnson - The web master for the whole DEVLAB team who helped create the current design and layout of the site.

Richard Adams - A fellow senior computer science major who joined the team during the Spring term of 2001 and provided critical help in the analysis that went into the creation of our mission statement and purpose.

Mason Kortz - A fellow senior computer science major who joined the team during the Spring term of 2001 and provided critical help in site development.

Lou Scerra and Nick Morinigo - Fellow senior thesis writers in English and Government, respectively, with whom I spent many a late night.

1.3 Technical Note

As of the release date of this document, we are still in the process of tying the domain name EcomRISK.org to our server. The web site can be accessed at <http://devlabserver.cs.dartmouth.edu/sites/ecomrisk/> .

Chapter 2

Business Analysis

2.1 Market Analysis

2.1.1 PEST Analysis

In order to determine what the potential user base for our site was and, therefore, what design principles we should follow we used the well known PEST analysis. The letters of PEST stand for political, economic, social, and technological. This analysis attempts to identify the trends in these four areas in order to determine if the trends exist to provide a wide enough audience for our site.

1. Political
 - (a) On-line criminal activity is more prevalent.
 - (b) Law enforcement is having to deal with the regulation of e-commerce.
 - (c) Does tax policy need to be re-written to take e-commerce into account?
 - (d) Privacy (or Big Brother) is becoming more of an issue.
 - (e) E-Commerce is beginning to transcend national boundaries.
 - (f) There is increasing concern and interest among the population at large.
 - (g) The U.S. Government itself is doing more and more of its own business on-line..

- (h) There is a general shift in the policy focus from libertarian ideals to more conservative ideals.

2. Economic

- (a) E-Commerce is a new way of conducting business (B2B and B2C).
- (b) .COMs are declining.
- (c) Larger companies are just starting to come on-line and will bring a continued growth in technology and systems using electronic networks to perform business.
- (d) Government is beginning to take a look at the effect on-line businesses have on the economy and how they should be regulated.
 - i. There is a lack of any sort of tax for on-line transactions which could mean a serious loss of tax revenue for the government.
 - ii. On-line business is largely unregulated by law. Risks could transform rapidly into serious legal issues if they are exploited.

(e) Insurance

- i. The historical way to deal with risks is to insure against them, however, there has not been much of an entrance into the market by large insurance firms.
- ii. On the consumer level, credit cards are starting to become e-commerce aware and deal with the risks of doing business on-line (American's idea of third party liability, "on-line fraud protection").

- (f) The invasion of technology into all parts of a business' operation.

3. Social

- (a) More and more people are making the Internet a part of their everyday lives.
- (b) People are starting to form purely on-line communities.
- (c) On-line companies are advertising in the "real" world.
- (d) People are relying on technology and the Internet to be able to function in everyday life.

4. Technological

- (a) Pre-packaged "e-commerce" systems are starting to be heavily used rather than custom built systems.
- (b) More everyday devices are connected to the Internet and other networks.
- (c) Higher bandwidth is available across the access spectrum.
- (d) Interaction with computers is moving beyond the screen and keyboard.
- (e) There is a centralization of data and computing.
- (f) There is an increased accuracy in personal information.
 - i. Smart Cards
 - ii. Biometric information
- (g) There are more wireless connections.
- (h) Autonomous agents are starting to become widely used.
- (i) Innovation is surpassing the threshold past which people can understand all of the interactions taking place.
- (j) A wider spectrum of people are using technology.

This list of trends is by no means meant to be exhaustive, but does suggest that there are a number of trends suggesting that there will be a large amount of people and businesses concerned with the risks of e-commerce. The potential user base for a site like EcomRISK.org will likely grow dramatically in the next few years.

2.1.2 SWOT Analysis

The SWOT analysis is an attempt to examine the overall picture of what resources we have as a group, where our opportunities are, where we are vulnerable, and what poses a direct threat to our project. The letters in SWOT stand for strengths, weaknesses, opportunities, and threats. This analysis is an attempt to determine how feasible it is for us to take on a project of this nature. The paints a picture of the playing field we would bring our site on to.

1. Strengths

- (a) Human Resources (i.e. Smart students, good faculty, etc.)

- (b) Not constrained by business factors (i.e. profits, shareholders, etc.)
- (c) Dartmouth name
- (d) Dartmouth resources
- (e) Faculty resources
- (f) Constant influx of new ideas
- (g) Interest from financial sponsors
- (h) Non-Profit status
- (i) Institute of Security Technology Studies (ISTS) connection
- (j) Can use other organizations on campus

2. Weaknesses

- (a) D-Plan's lack of continuity
- (b) Lack of business discipline
- (c) Lack of experience
- (d) Reliance on grants as sources of funding

3. Opportunities

- (a) Fill a niche in security web sites
- (b) Create a knowledge base that can serve as a teaching tool
- (c) Bridge the gap between the academic study of computer science and the business implementations of these technologies
- (d) Enrich the content provided by the GREeCOM.org sites
- (e) Provide a quasi-academic setting in which companies can freely share knowledge allowing them to realize a mutual, as well as individual, gain
- (f) Bring students a valuable extra-curricular activity
- (g) Provide the ISTS at Dartmouth a connection to the undergraduate and graduate students

4. Threats

- (a) Hard to become known among business communities without large advertising expenditures
- (b) Difficult to keep the continuity of the sites maintenance and development with the varying staff size and short tenures of staff
- (c) E-commerce security is a hot topic which will require us to move fast if we want to fill our niche
- (d) Operating budget is much smaller than most companies
- (e) Staff size is much smaller than most companies
- (f) Staff are often working on this site while at the same time maintaining full academic workloads

This analysis shows that we have several strong opportunities to which our strengths can speak. Also, by examining our weaknesses we have become aware of what hurdles we must overcome and be constantly vigilant for. It is also important to have a good idea of what are some direct threats to our project. We must always be aware of what might potentially make us obsolete, diminish our value, or prevent us from being successful.

2.2 Mission Statement and Purpose

To serve as a guide in our development I came up with a mission statement and purpose for EcomRISK.org. As first it may seem odd to break these two up, however, Collins and Lazier suggest doing so in *Beyond Entrepreneurship: Turning your Business into an Enduring Great Company* [7] and I found it useful.

2.2.1 Mission Statement

Our mission is to become the number-one web site for electronic commerce risk news and assessment by bring high quality risk classification to business.

2.2.2 Purpose

Our purpose is to create an on-line community for those involved in electronic commerce which fosters learning about the risks of electronic commerce through the sharing of information.

2.3 Technological and System Goals

Nobody likes boring web pages, least of all people who use computers all day. As all of the EcomRISK.org developers were aware of this, we strove from the beginning to create as dynamic a site as possible. A site can be dynamic because it is constantly updated by human operators or because the code that runs it constantly generates new content. A human operator is more likely to make the best choices about what content is relevant, what is valuable, and where it should be placed. However, one of the weaknesses touched upon in our SWOT analysis (see Chapter 2.1.2) is that we are chronically short on people and people may only work on the site for short amounts of time. Therefore, we decided that one of our main goals would be to create a dynamic, largely automated site through the use of code.

A second goal was to build a site that used the latest in technology. This serves two purposes. First, it allows our own site to be a case study in information storage and management using the latest technologies. The management of our own site requires us, the developers, to stay on top of the latest trends in technology in order for us to stay current and relevant to our audience. No one would buy a computer from a store which used abacuses to calculate a customer's bill. Similarly, a web site that claimed to be "The Source for E-Commerce Risk News & Assessment" would not provide much interest to the users of the Internet if it did not employ the latest technologies.

The second purpose of building a web site with the latest technologies is to serve as an instruction tool for those students who become involved in its development. By providing a place, albeit small, in the computer science department that is involved with the latest web technologies, the site offers the students a chance to get involved in computer science outside of the classroom. This use of "hot" technology is a strong attraction for students who, we must interest, as they are are only source of developers.

A third main goal was to build a site that could be used to test different sorts of data mining techniques. As much of the incidents we are concerned with collecting center around the use of data, our site's value would be greatly increased if we could ourselves showcase some of the latest algorithms and thinking in the field.

Our fourth and final main goal centered around devising new ways to manage the knowledge we acquire through user input. We collect vast amount of information about users whether it be from their own submissions or our own

log of what actions they have performed. In some ways this is an extension of our third goal, but involves separate disciplines. The site attempts to bring together many separate, but inter-related forms of data. By attempting to create intuitive ways to view and access the data, we could greatly increase the visibility, and thus the value, of the data in our site. An example of this would be to create a meta-search that could be used to query all of the differing types of data we have in one coherent interface. By allowing a user one main entry point into the data, we can expose them to data that they might not have even realized. This also will allow us to examine relations between data sets that we may not have foreseen.

Chapter 3

Basic Site Architecture

We set out to build a site that created a dynamic, vibrant on-line community centered around learning and discussion. As discussed in Chapter 2.3, we set out four main goals that would act as our guiding principles in the development of EcomRISK.org. We wanted to build a dynamic, content rich site that, while showcasing the latest technology, provided a test bed upon which we could try new data mining and knowledge management techniques. We had to decide what to use for a base platform, web server software, server-side scripting language, and database server.

To achieve these goals we set about choosing the best mixture of tried and new technologies we could. For a base platform for our server, we wanted to try a new technology and therefore choose Macintosh OS X Server (Mac OS X Server)¹. Mac OS X Server is, at its core, a variant of the Berkeley Software Distribution of UNIX. Due to our limited budget, it was important to choose a platform that would allow us to use the wide range of free, open source tools available on the Internet. Choosing Mac OS X Server allowed us to download and compile a wide range of freely available software for web site creation.

Next, we choose the software to actually serve our site. This choice was fairly simple. The Apache Server² is the most widely used web server software on the Internet. The Apache Server is a robust, efficient web server. By choosing Apache, we left our options for what server-side scripting language

¹For more information about Macintosh OS X Server, see its web page at <http://www.apple.com/macosx/server/>.

²For more information on the Apache Server, see its web page at <http://www.apache.org/>.

to use wide open.

There are many server-side scripting languages in use across the Internet and this choice was less obvious. For our own curiosity, we wanted to try a fairly new technology and thus choose PHP³. As a scripting language, PHP is very powerful language with C-like syntax. PHP can be intertwined with HTML code to allow for fast and efficient development. Again, the choice of PHP left open our choice of database systems.

To round out our server, we needed to find a reliable and fast database system. We choose MySQL⁴. MySQL is an open source relational database management system. Using the prevalent SQL language as its interface, MySQL fit well into our architecture. PHP has built in routines for accessing MySQL which allowed for us to tightly integrate the web pages and database.

By choosing Mac OS X Server, Apache, PHP, and MySQL for our web serving platform, we created a mixture of tired technologies with new ones. Figure 3.1 shows how a typical page request requiring a database access from a user's web browser would be handled by our setup. The steps are outlined below.

1. A web browser makes a request for a page.
2. The Apache Server receives the request and, after doing some basic logging, passes it to the PHP module.
3. PHP compiles and runs the script that constitutes the page requested and makes any and all database queries required by the page. If a page requires multiple queries, steps 3 and 4 will be repeated as necessary.
4. The MySQL server makes the queries and returns the result set to the PHP script.
5. The PHP script runs to completion and sends its output back to Apache.
6. Apache then sends the web page back to the client browser.

None of this technology breaks new ground. However, it was not our intention to do so. We wanted to build a robust system which we could use to reach our goals.

³For more information on PHP, see its web page at <http://www.php.net/>.

⁴For more information on MySQL, see its web page at <http://www.mysql.com/>.

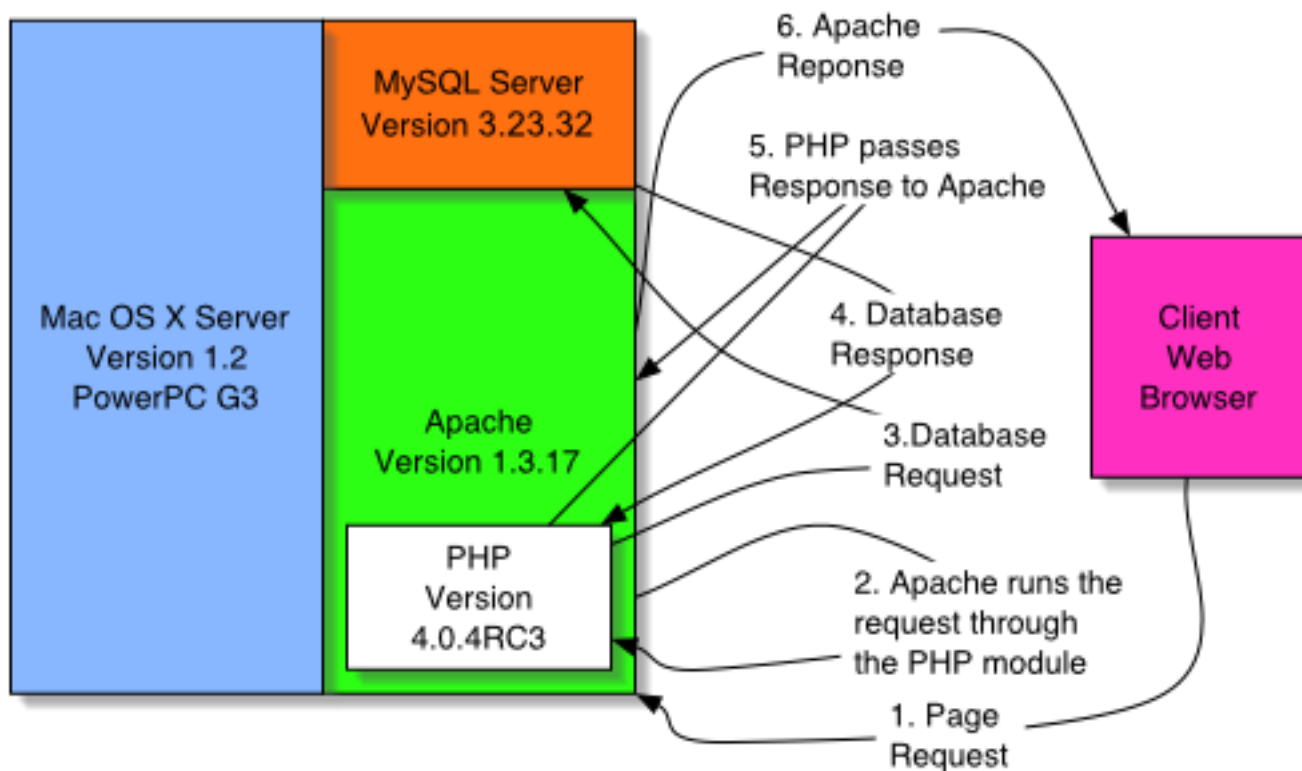


Figure 3.1: A diagram of how the server handles the typical page request

One important consideration in the design of our system was its scalability. Currently our system has sufficient response time, however, as the number of users increases, we will need to modify our design to increase performance. One of the first changes we could implement to improve performance would be to move the MySQL database to another machine. This would reduce load on the main web serving machine. Also, we could consider upgrading our base platform to one with multiple processors or changing the operating system to one that supported the clustering of servers. Mac OS X Server supports machines with up to two processors, but not clusters of machines. Apache, MySQL, and PHP can also be compiled and run on other UNIX and Windows operating systems. Therefore, if the need arose, we could transport the web site with little work to another operating system that supported machine clustering to increase usage capacity.

Chapter 4

The EcomRISK.org Taxonomy

It is the taxonomy presented in the incident submission system of EcomRISK.org that makes it unique among the web sites on the Internet that attempt to discuss and address e-commerce. The vision was to create a system that returns relevant information concerning the types of e-commerce problems faced by companies and how they were solved. What is particularly unique about this taxonomy is that it combines technical analysis with a quantifiable results analysis. A main goal in the creation of the taxonomy was to create a classification system so that we could return meaningful search results for those who wanted to browse our database of incidents. The questions in the taxonomy may seem odd upon their first reading, but they attempt to create a coherent framework into which each incident can be placed.

In this chapter I will first present related work in the field of security taxonomies and then go into detail about how our taxonomy is structured, using our incident submission form as a guide. It would be useful to examine the incident submission form when reading this chapter and it can be found at <http://devlabserver.cs.dartmouth.edu/sites/ecomrisk/cases/> .

4.1 Risk Incident Submission Form

The risk incident form is broken up in to 5 sections: “Submitter Information”, “Incident Information”, “Solution Information”, “Quantifiable Ramifications”, and “Company Information”. In the following sections I will go into each section of the form in-depth, explain each question, and how the

answers help classify the incident. Some questions have predefined answers and others a free-form. The questions with predefined answers allow us a way to standardize the classification. The free-form questions provide the submitter the chance to relate the details of a particular incident.

4.1.1 Submitter Information

The first section asks for information about the submitter. To reduce the number of random incident submissions, we require the user to be logged in to our site before they can make a submission. In order to login, a user must have created an account on the GREeCOM.org site. The GREeCOM.org site handles all of the user registration details so that each of its three member sites, of which EcomRISk.org is one, can share one user system. During the registration process a user is required to tell us their e-mail address and their full name. Optionally, they may also list a mailing address and some other personal information.

The end result of using the GREeCOM.org's user registration system is that the incident form can be light on personal information. The user does have the option to make the submission anonymous, in which case the submission is stored with no associated user name, the IP address of the submitter is not recorded, and the type of browser that is being used is not recorded. If the submission is not made anonymously, these three pieces of information are recored simply for logging purposes. The only other question asked in this section is the user's relation to the incident they are submitting. Their are two answers to choose from: "Personally Involved" and "Read about it in the News". I ask this question to get some sense of how deeply involved a person is in the incident they are reporting. Submissions are not necessarily of less value if the person was not personally involved, they may simply lack the depth of information that someone who was directly involved may have. There is a high value in their being a high numbers of incidents in our database because, the more incidents we are able to collect, the more valuable searches of our database will be. It is for this reason that we intentionally did not make personal involvement a requirement.

4.1.2 Incident Information

The second section of the form asks for information about the incident itself. At the core of each incident lies a computer system and, therefore, when

attempting to create a classification system I researched the available literature on computer security taxonomies. It is important to understand that at the center of every incident lies a computer system flaw. I will discuss the literature as it becomes relevant to the questions below.

The first question in this section asks, “Where did the cause of this incident originate?” The submitter must choose either “Internal” or “External”. This question is to classify the incident by the location of the cause.

This rest of this section’s questions center around classifying the incident’s computer system flaw. This is by far the most technical part of the submission process. One of the most useful papers I found on the subject of computer security taxonomies was *A Taxonomy of Computer Program Security Flaws* [8]. It is important to the understanding of the “Incident Information” section to give an overview of the paper here. The Landwehr et al. paper presents a clear taxonomy for organizing computer security flaws. All of the incidents we hope to collect have at their root some security flaw. The authors use three general parameters to classify security flaws: what was its genesis, when was it introduced, and where was it introduced. The authors then break down each parameter such that an error fits into one sub-category in each question. Landwehr’s taxonomy focuses entirely on the technical side of the question and is extremely useful when attempting to understand what, technically, occurred. I have chosen to use the Landwehr taxonomy for this section of the form. I have made slight modifications to their taxonomy, largely to simplify some of the categories because, sometimes, their granularity is too fine for our uses.

The second question in this section asks about what the result of the incident was. Landwehr states that “computer security flaws are any conditions or circumstances that can result in denial of service, unauthorized disclosure, unauthorized destruction of data, or unauthorized modification of data” (Landwehr et al. 211). I agree with Landwehr’s analysis and will use his classification system for this taxonomy. It may seem odd at first to ask what the result was, but it is useful here because it suggests the approach we are using. All of our incidents are submitted after the fact, and therefore a certain amount of working from effect to cause is required. There are four predefined answers for this question.

Denial of Service - A submitter would choose this answer when the incident resulted in a denial of service to one or more systems. It is important to note that in the case of a malicious DoS attack,

the denial of service is both a symptom as well as a result.

Unauthorized Disclosure of Data - This answer would be chosen if the incident resulted in unauthorized data access. An example of this might be a hacker cracking a web site's credit card database such that they could see user's credit card information.

Unauthorized Destruction of Data - This answer would be chosen if the incident resulted in unauthorized destruction of data. An example of this occurring would be if the user of a system was allowed to destroy other user's data without proper authorization.

Unauthorized Modification of Data - This would be chosen if the incident had resulted in the unauthorized modification of data. An example incident involving this would be a bank employ being able to inappropriately change a customer's balance in the computer system.

These four categories are prevalent throughout the literature on computer security flaw taxonomies and provide very good general categories for the result of an incident.

The first part of Landwehr et al.'s taxonomy focuses on the genesis of the flaw. As I state the question on the form, "How did it enter the system?". This question has eight predefined answers which come directly from the Landwehr et al. taxonomy. I have removed the last level of their categorization because it is too detailed to be of use in our taxonomy. I will also not explain exactly what each category means for two reasons. First, they are largely self explanatory, and second the authors of the taxonomy did a much better job than I could hope to and leave it to the reader to examine their paper.

- Intentional - Maliciously inserted (Backdoor, Trojan Horse, etc...)
- Intentional - not Maliciously inserted (Covert Channel, etc...)
- Inadvertent - Incomplete or Inconsistent parameter Validation
- Inadvertent - Implicit Sharing of Privileged/Confidential Data
- Inadvertent - Inadequate serialization of process
- Inadvertent - Identification/Authentication Inadequate

- Inadvertent - Violable Constraint Error
- Inadvertent - Exploitable Logic Error

Just to provide a few examples of how different incidents might fall into these categories, a DoS attack would fall under the “Intentional - not Maliciously inserted (Covert Channel, etc...)” category and a web site the allowed access to its credit card database would fall into the one of the “Inadvertent” categories, depending on the exact circumstances. The credit card database access could be due to poor design and fall into the “Inadvertent - Exploitable Logic Error” category or may not properly check a user’s privilege , in which case it would fall into the “Inadvertent - Identification/Authentication Inadequate” category. Following this question there is a free-form question that allows the submitter to explain in detail how the flaw entered the system.

The third question asks “When did it enter the system?” Again, this question and its answers come from the Landwehr taxonomy. For this categorization, I left their categorization choices as they state them in their paper. There are five predefined answers to this question.

- During Development - Requirement/Specification/Design
- During Development - Source Code
- During Development - Object Code
- During Maintenance
- During Operation

To continue with the examples used to describe the “how” question, a DoS attack would fall into the “During Operation” category and an unauthorized access to the credit card database might fall into any of them, depending upon the exact circumstances. How could the unauthorized credit card database access fall under any of the categories? Say the database is protected by a password, perhaps the requirements for the password were no robust enough and, therefore, easily cracked. This would fall into the “During Development - Requirement/Specification/Design” category. If, instead, the code did not properly check passwords, it would fall into the “During Development - Source Code” category. Again, this question is followed by a free-form question that allows the submitter to describe exactly when the flaw entered.

The fifth question asks “Where did it enter the system?” Again, this question and its answers come from the Landwehr taxonomy. Landwehr’s classification is of an appropriate detail so I do not amend the categories. There are eleven predefined answers for this question.

- Software - Operating System - System Initialization
- Software - Operating System - Memory Management
- Software - Operating System - Process Management/Scheduling
- Software - Operating System - Device Management (including I/O, networking)
- Software - Operating System - File Management
- Software - Operating System - Identification/Authentication
- Software - Operating System - Other/Unknown
- Software - Support - Privileged Utilities
- Software - Support - Unprivileged Utilities
- Software - Application
- Hardware

To keep up the examples used in the previous questions explanations, a DoS attack would fall under “Hardware”, for it is a router’s or server’s inability to handle the load, and unauthorized access to a credit card database could fall under any of the categories. The unauthorized credit card database access could be the fault of the operating system not properly authenticating users, in which case it would fall into the “Software - Operating System - Identification/Authentication” category, or there may be a utility that routinely performs a function on the database that is not properly secured, in which case it would fall under the “Software - Support - Privileged Utilities” category. This question is followed by a free-form question which allows the submitter to describe the exact circumstances.

The second to last question of this section seeks to determine how long the incident effected the system. This is important to determining the severity

of the incident. Here the submitter is limited to choosing from predefined questions so that we may categorize the incident.

The final question allows the submitter, in their own words, to describe the incident and give an overall summary. This free-form questions allows us to receive detailed information about the unique characteristics of the incident.

4.1.3 Solution Information

The third section of the form asks for information about what the nature of the solution to the incident was. Again, the questions in this section are a mixture questions with free-form and predefined answers. We are attempting to classify the solution not only by what it took technically to fix, but also what, if any, policy changes needed to be made to reach a final solution. The questions concerning policy fixes are meant to extract the underlying policies that companies run by from the technologies they use. Before I began work on this section I read two papers which proved to be very useful. The first was *The Remedy Dimension of Vulnerability Analysis* by Lindqvist et al. [9] and the second was *A Critical Analysis of Vulnerability Taxonomies* by Bishop and Bailey [10]. I would suggest both to the reader interested in work on taxonomies in this area.

The first question, “Technically, what did the fix involve?”, has five predefined answers.

Nothing at all - In this scenario the fix to the problem was not technical. For example, a DoS attack would not require any technical fix. There are many problems that may just involve changing the way a company operates and not the technology they use.

Reconfiguration of system - This might be chosen because the incident did not require any re-working of code or design, however requires a change to the way the system is setup. For example, a company might have accidentally left a web server setup so that it would provide directory listings, instead of always going to a web page. This might expose files and data that the company does not want to be public. In this case, most web servers have a simple setting to turn this off and fixing it is as simple as changing this setting.

Applying a patch - A constant problem in maintaining servers is keeping the programs on them running up-to-date with the latest patches. Patches are fixes to known bugs that are distributed by the creator of the software. It is a never-ending job to keep all programs up-to-date. A user might select this on the form when the incident was caused by a fault that had already been fixed in a patch.

A re-design and re-implementation of the system - If a user were to select this choice it would mean that their original design for the system was inherently flawed. In order to fix the flaw the system had to be re-designed and re-implemented using the new design. An example incident in which the user would choose this answer would be when a system's authentication process was designed with a flaw that allowed improper authentication of users and the only way to fix it was a re-design of the system.

Just a (corrected) re-implementation of current design - This is similar to the previous answer, however, it certainly happens in the real world that a design is not flawed, however the implementation is flawed. This solution requires a re-working of code, but not a total re-design.

Next we allow a free-form text answer to describe what the solution was, based on the classification chosen in the previous question. This allows the user to explain exactly what was required in their technical fix. To round out the question about the technical fix, we ask for the URLs of any resources that might have been used in the solution. This provides vital information about what Internet resources are truly useful for security problems.

Now that the user has told us how, if at all, their technology needed to be fixed, we ask the same question, except this time we are looking at how their policies might have changed. Policy may seem somewhat nebulous, but here refers to the practices used in the running of a system. We have three predefined answers for "In terms of policy, what did the fix involve?"

Nothing at all - In this scenario the fix to the problem was did not require a change in the companies policy. This suggests that the nature of the change lies in the technology used, not the policy surrounding its operation.

Re-design of policy - One might choose this when making their submission because the solution required a change in their policy. An example of this type of incident would be if a system administrator noticed a user misusing their account in such a way that was not wrong, but inappropriate, and had no place higher up in an organization to bring this information to. The system administrator may not be tasked with making a decision about the user so they ignore it. Upon the realization of the problem, the organization would have to re-design their policy so that there was a place to bring such information so that it could be dealt with.

Re-education on policy - What a organization decides upon as their policy towards their systems may not always be followed. Often there are rules or procedures which are there for show and not used in real practice. An example of this would be a developer adding new code to a product without going through the proper channels. This change may cause a flaw that could have been prevented had the code gone through review. In this case, an organization would have to re-educate their programs as to the proper procedures regarding code changes.

Following this question is a free-form question which allows the submitter to describe exactly what the policy changes where, if there were any at all.

4.1.4 Quantifiable Ramifications

The purpose of this part of the form is to quantify what the ramifications of the incident were. This section has two questions: "How many hours did it take to analyze and solve the problem?" and "How much revenue was lost as a result of the incident?". The first of these questions classifies the risk by its severity in terms of time. The longer an incident takes to analyze and solve, the more of a drain on a organization's resources it is. In this questionnaire we have created seven predefined answers. They range from "5 hours or less" to "100 hours or more". There may also be incidents that take no time to solve, but still take time to analyze. A DoS attack is a good example of a situation in which the realization that you were under attack might take some time, but there is currently no solution you could implement to stop it.

The second question asks about how much revenue was lost as a result of the incident. This, like the first question in this section, helps to determine

the ramifications of the incident. Some incidents, while they may take require drastic measures to fix them, do not, in the long run, cause much harm in terms of revenue. Loss of revenue is an important factor in determining the overall affect a incident has.

4.1.5 Company Information

This final section of the incident submission form asks for information about the company that was involved in the incident. Knowing what business an incident occurred in can help to create a more detailed picture of the types of risks faced by different businesses.

Submitters are allowed to make a specific company anonymous, however they are required to choose what type of business (or businesses) the company involved in the incident is in. For this question we use the North American Industry Classification System (NAICS) created by the United States Census Bureau¹. The NAICS is assigned to a business by the Internal Revenue Service (IRS) when it receives their federal tax forms. We are currently using the 1997 NAICS as this is the standard. This classification system is not only used in the United States, but also in Mexico and Canada. The European Union is considering adopting this and will likely do so in 2002. By using a widely accepted classification system for business types, we can classify incidents much more accurately, and not rely on someone's own choice as to what type of business he or she thinks a business is in.

If a submitter does not choose to make a company anonymous, they may enter the name of a company, its stock symbol if it is publicly traded, and an official contact person. The name and stock symbol can help us to find even more information about a company and aid in further classification of the incident. The official contact information can help us establish a link to the company to attempt to open up a dialog about this incident.

As the taxonomy is new, we do not take full advantage of this information. A future upgrade of the system could take this information and automatically determine a large amount about the company. Also, once the number of incidents grows, this type of information will greatly aid in generating statistics about what specific risks are faced by a sector of business.

¹For more information about the NAICS you go visit the Census Bureau web site about it at <http://www.census.gov/epcd/www/naics.html>.

Chapter 5

Other EcomRISK.org Tools

The development of the site did not stop with the creation of the incident system. In order to become a site which becomes a stop on user's daily trips around the news web-sites, we thought it would be critical to sections that would bring daily content and human interactions on a basis less formal than the incident system.

Our goal is to create a on-line community in which people can interact on many different levels. To this end, we added several features to the site that bring the user the chance to catch up on the latest news, discuss the news, incident submissions, or anything else relating to e-commerce, read up about the latest trends in working papers, and in general add to the overall depth of knowledge available on the site. Moreover, tying the site into the GREeCOM.org family of sites, it brings a critical piece to the overall center as a whole.

5.1 Forum

Our choice to use PHP as the server-side scripting language and MySQL as the database allowed us to choose from many "off-the-shelf" forum systems. We decided to use the "electrifiedForum Version 0.93" released on March 27, 2001¹ as our starting point. The system consisted of several PHP scripts and required several tables in the MySQL database.

The system implemented all of the features we desired. There could be

¹More information can be found about this system at <http://www.electrifiedpenguin.com/apps/forum.php>.

many top level topics, within which messages could be organized. Replies to a particular message were "threaded" so that the user had some sense of how a series of messages related to one another. The package also included an searching function to search the messages.

After installing the package I made fairly large modifications to the internal workings of the system to integrate our user system, in place of the one provided, and to change the look and feel of the system.

5.2 In the News...

Our aim with the "In the News..." section is to provide links to relevant news stories concerning the Internet. Towards fulfilling its purpose this section consists of two main news sources. First, we pull down an XML version of the Slashdot.org home-page every half-hour and parse out the submission headlines and the links to the submissions. Slashdot.org is a site that contains news items submitted by users and the runs a threaded discussion off of each news item. The headlines pulled from Slashdot.org are those typically found at news web-sites, however, they provide a good snapshot of the hot topics around the Web from users of the Web. By updated the news stories every half-hour, the site has at least one section that is updated all the time.

Our second source of news is a daily news briefing prepared by the ISTS at Dartmouth College. The headlines and links to the articles from this daily e-mail are manually extracted and inserted into a database table. Currently, the "In the News..." main page pulls the twenty-five most recent postings from the database and displays them. The links provided are actually links to another script, which takes the record ID of the particular news items database entry as a parameter, counts this click as a "view" and redirects the browser to the site with the actual news piece. This simple click-counting method allows us to see which news items people are most interested from their headlines.

5.3 Working Papers

Towards our goals of becoming an information rich site, we thought it would be valuable to include working papers of a non-technical nature as a knowledge base on current trends and issues in e-commerce. The working papers

not only provide a important amount of knowledge for the users of the site but, also, provide a place for students to become involved in the EcomRISK.org community. At the time of this writing we had over 70 working papers available on the site.

I constructed a system to handle the submission of working papers (in PDF format) so it would be easy to search and sort them by author, keyword, title, abstract, references, or full text. While, currently, only administrators can upload new working papers, the system is easy enough to use that eventually it could made available to the public for use at large. A PHP generated web page provides the interface to upload the papers, providing space to specify the author, title, abstract, keywords, references, full text of the paper, a PDF file containing the paper, and a PDF file containing presentation slides if one accompanies the paper. The information about the paper is then stored in a database table and the file is stored in a directory inaccessible to the web-server. Each paper is given a unique identifier² so that the author can reference it.

Any user can search the database of papers seeing the paper's title, author, abstract, and keywords. Only a registered user can download the paper. A user can download the paper after accepting an agreement concerning the acceptable uses of the paper. Once again, the download link in fact links to PHP script so that we can store in our database which users download which papers to what IP address at what time.

5.4 Resources

Perhaps the section with the least realized potential, the resources section is meant to provide users with a list of sites that we have found to be useful. The list of sites in this section comes from ones found by us, the developers. For each link we have specified several pieces of information.

Level - In this category a site can either be 'top' or 'specific'. This classification is meant to deal with whether this resource is a general one, such as a whole web-site, or just one page from a web-site.

²This identifier looks like "WP-EC-010329-18". The letters "WP" stand for working paper. The letters "EC" stands for EcomRISK.org. The 6 digit number is the data upon which the paper was submitted in the form YYMMDD. The last number states that it was the eighteenth paper submitted that day.

Type - A site can be one of seven types: ‘news’, ‘background’, ‘research’, ‘solutions’, ‘certification’, ‘conference’, or ‘legal’. This is meant to classify the sites by what type of content it holds. A site may be one or many of these types.

Description - This is a small description about a site either taken directly from a site’s own pages or written by the contributor. The intention of this section is to provide a short description of what the resource contains.

Keywords - These are always determined by the contributor. The intention is for there to be a fairly small number of keywords that characterize what the resource contains.

These classification fields allow a user to create more meaningful searches when they are looking for specific information. While the site does not currently take full advantage of this extensive classification, the framework exists.

Again, as with so many other parts of the site, when a user clicks on a link to a resource, they are in fact click on a link to another PHP script which counts this click as a “view” and redirects them to the page. This provides us with valuable information as to which resources users find useful based on the information we have provided.

5.5 Other GREeCOM.org Sites

As discussed in the section entitled Project History (see Chapter 1.1), Ecom-RISK.org is just a part of a larger family of sites. The parent site, GREeCOM.org is responsible for controlling user registration and provides a central point to navigate to the parts of any of the other sites. some of the sites contain deep links into the features on one of the other two sites. To give the reader a general sense of the feature set in GREeCOM.org, I will give a cursory examination of the main features of the other two sites.

5.5.1 eJETA.org

eJETA.org’s main feature is a database of refereed papers written on the latest research methods, tools and applications in electronic commerce by experts in the field. These papers will come form a wide range of disciplines.

As the site states, one of its main goals is to "To develop an interdisciplinary international forum which has a scientific basis and which provides technology transfer between university research and commercial/industrial communities."

This database of papers will provide an important knowledge base for the users of EcomRISK.org. The papers can be cross-referenced with the incident database in EcomRISK.org to provide a user the latest news and information on a particular risk.

5.5.2 DePolicy.org

DePolicy.org's has two main features. First, the site provides comprehensive evaluations of web sites that offer on-line courses, training, and certification on e-commerce issues. EcomRISK.org users might be very interested in this, as they are business professionals. The second feature of DePolicy.org is its own self testing system. Using a quiz system developed by Mason Kortz, a user may take quizzes on a wide variety of topics dealing with e-commerce using their web browser. The quizzes are short and all of the questions are multiple choice. Upon completion of the quiz, the user is not only told how many questions he or she answered correctly and incorrectly, but also a list of resources that provide more information about questions that were answered incorrectly. These resources can often be found on other GREeCOM.org sites and provide a way to encourage participation in the center as a whole.

Chapter 6

Conclusion

6.1 Project Status and Future Work

As of June 1, 2001 the site will be live and open for anyone on the Internet to use. As of this date, the system as a whole should be considered to be in a version 1.0 release. However, in order to realize the full potential of the site, there is still much more work that can be done. Our main goal was to have the site in a stable state with a fairly comprehensive feature set. The underlying framework of the site has been designed such as to limit as little as possible the directions future developers want to take the site.

Under Professor Makedon's guide the team involved with the GREeCOM.org project has continued to grow and there is no doubt that the site will continue to be expanded and move in new directions.

6.2 Personal Conclusions

Over the past nine months I have learned a great deal working on this project. From computer risk taxonomies to business analysis, this project has dramatically expanded my experience at Dartmouth College. I have greatly enjoyed being able to work with fellow computer science students, both undergraduate and graduate. I look forward to watching EcomRISK.org growing over the next few years.

Bibliography

- [1] E. Cone, *New E-Commerce Unit For General Motors* (2001). Available at <http://www.zdnet.com/intweek/stories/news/0,4164,2311854,00.html> (15 May 2001).
- [2] F. Makedon: *E-Commerce Security Resource: "ECOMRISK" Data Center"*
- [3] Y. Hiramatsu, *Electronic Commerce: Trend and Future* (2000). Available at <http://www.okibusiness.com/oki/otr/html/nf/otr-183-02-3.html> (27 May 2001).
- [4] S. Costello, *UPDATE - CERT hit by DDoS attack for a third day* (2001). Available at <http://www.idg.net/go.cgi?id=481360> (27 May 2001).
- [5] D. McGuire, *White House Confirms Denial Of Service Attack* (2001). Available at <http://www.washtech.com/news/regulation/9997-1.html> (27 May 2001).
- [6] R. Groth, *Deriving meaning through mining* (2000). Available at <http://www.indiaserver.com/businessline/2000/12/11/stories/211177wp.htm> (27 May 2001).
- [7] J. C. Collins and W.C. Lazier, *Beyond Entrepreneurship: Turning your Business into an Enduring Great Company* (Prentice Hall, Englewood Cliffs, 1992), p. 95-134.
- [8] Landwehr, C. E., Bull, A. R., McDermott, J. P., Choi, W. S., "A Taxonomy of Computer Program Security Flaws." *ACM Computing Surveys* 26:3 (1994): 211-254.

- [9] Lindqvist, U., Kaijser, P., Jonsson, E., “The Remedy Dimension of Vulnerability Analysis”. *Proceedings of the 21st National Information Systems Security Conference* (1998): 91-98.
- [10] Bishop, M., Bailey, B., “A Critical Analysis of Vulnerability Taxonomies”. *TR CSE-96-11 Dept. of Comp. Sci., University of California at Davis* (1996).