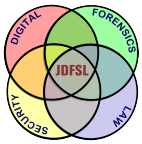

THE JOURNAL OF
DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 1 | Number 3 Article 1

2006

Designing a Data Warehouse for Cyber Crimes

Il-Yeol Song
College of Information Science and Technology, Drexel University

John D. Maguire
College of Information Science and Technology, Drexel University

Ki Jung Lee
College of Information Science and Technology, Drexel University

Namyoun Choi
College of Information Science and Technology, Drexel University

Xiaohua Hu
College of Information Science and Technology, Drexel University

See next page for additional authors

Designing a Data Warehouse for Cyber Crimes

Authors
Il-Yeol Song, John D. Maguire, Ki Jung Lee, Namyoun Choi, Xiaohua Hu, and Peter Chen

Designing a Data Warehouse for Cyber Crimes

Il-Yeol Song
College of Information Science and Technology
Drexel University
song@drexel.edu

John D. Maguire
College of Information Science and Technology
Drexel University

Ki Jung Lee
College of Information Science and Technology
Drexel University

Namyoun Choi
College of Information Science and Technology
Drexel University

Xiaohua Hu
College of Information Science and Technology
Drexel University

Peter Chen
Department of Computer Science
Louisiana State University

ABSTRACT

One of the greatest challenges facing modern society is the rising tide of cyber crimes. These crimes, since they rarely fit the model of conventional crimes, are difficult to investigate, hard to analyze, and difficult to prosecute. Collecting data in a unified framework is a mandatory step that will assist the investigator in sorting through the mountains of data. In this paper, we explore designing a dimensional model for a data warehouse that can be used in analyzing cyber crime data. We also present some interesting queries and the types of cyber crime analyses that can be performed based on the data warehouse. We discuss several ways of utilizing the data warehouse using OLAP and data mining technologies. We finally discuss legal issues and data population issues for the data warehouse.

1. INTRODUCTION

Development of information technology is a double-edged sword: On one hand, information technology provides us with infinite possibilities of designing various information systems for effective management of information. On the other hand, vulnerability of the information assets in digital forms has resulted in more chances for intrusion, damage, and destruction by various types of attacks. These attacks include financial fraud, sabotage of data/networks, theft of proprietary information, unauthorized

system accesses, denial of service attacks, cyber stalking, identity theft, virus attacks, hacking by ID hopping, industrial espionage, interruption of e-commerce business, and breaches in national security. The attackers, termed computer hackers, crackers, and cyber terrorists frequently have displayed remarkable levels of sophistication in their attacks. Their goals run the gamut from the relatively benign, such as responding to technical challenges or basic human curiosity, to the misguided attempts to expose and publicize system vulnerabilities, to the purely criminal, seeking system destruction for political or financial gain. In combating the activities of these cyber criminals, law enforcement personnel, security specialists, and systems administrators have had to be technically adept, as well as at least partly clairvoyant. They have made use of consultants, packing special software toolkits to gather evidence. Over the past few years, they have also been able to employ new bodies of law that have changed the rules governing the prosecution of cyber crimes. The collection and analysis of these computer attacks are termed cyber forensics [13].

Utilizing database technologies in cyber forensics domain seems promising. There has been an increasing demand of centralized systems to store criminal information so that users can retrieve the information as necessary[1, 4-6, 14, 21]. By making use of a database technology, analysts could store and retrieve the "5 W's of a crime" – Who, What, When, Where, and Why. Moreover, utilizing combinations of database technologies will offer efficient ways to analyze and report crucial information about cyber crimes. Traditional database structures, however, are not powerful and efficient enough in analyzing cyber crime patterns, finding relationships among various data, or generating complex reports. Data warehousing, Online Analytic Processing (OLAP), and data mining technologies can be used to resolve the limitations.

Data warehousing and OLAP technology have been successfully used in industry. A data warehouse is a data repository that contains historical data for effective data analysis and reporting processes [12]. Data warehouses are designed to support decision-making by studying and analyzing complex sets of data. A data model used for designing a data warehouse or a small-sized data mart is called a *dimensional model* [12]. A typical dimensional model is composed of a fact table and a set of dimension tables. A fact table stores the data to be analyzed, whereas dimensional tables contain descriptive data used for browsing, filtering, and grouping the fact data. An example of a fact table in a cyber forensics data warehouse is cyber crime data. Examples of dimensions are cyber attack, date and time of attack, target of attack, attacker, and law enforcement personnel. With these dimensions, we can easily analyze cyber crime patterns from various combinations of the dimensional data.

Utilizing data warehouse technologies could open a new perspective for the analysis of cyber crimes. Some studies have defined and described cyber

crimes and cyber forensics at the conceptual level [8, 9, 23]. However, to our knowledge, there were no studies in cyber forensics research providing a data warehouse design for analysis of cyber crimes information. In this paper, we present three different dimensional model schemas that can be used for developing a data warehouse to analyze cyber crime data. In the context of cyber forensics, designing an effective data warehouse model is significant in that it will offer crime analysts with diverse views and methods to investigate criminal records; hence it provides them with useful preventive information about cyber crimes. This information might give specialists, administrators and law enforcement agencies information and tips to prevent further similar attacks, as well as the more direct value of solving similar cyber crimes. Our dimensional model for cyber forensics also helps identify the taxonomy of cyber forensics in accordance with the information needs of cyber crimes analysts.

Data warehouses support the use of OLAP (Online Analytic Processing) and data mining technologies for analyzing cyber crimes. By applying OLAP technology, more diverse and complex reports at various levels of abstraction can be generated. By applying data mining technology, cyber crime patters and association among the cyber crime data elements can be identified. We believe that this study will contribute to cyber forensics research not only to provide a conceptual map (i.e., a taxonomy of cyber crimes analysis) for the design of cyber crime data warehouse model, but also to serve as a basis for further development of a robust cyber forensics analysis system.

The rest of the paper is structured as follows: Section 2 reviews research on cyber forensics and the use of database technology for cyber forensics. In Section 3, we present three different dimensional models that can be used for designing a data warehouse for cyber crimes. In Section 4, we discuss how we utilize the dimensional model in terms of query types, OLAP, and data mining. In Section 5, we briefly discuss legal issues and data population issues. Section 6 concludes our paper.

2. LITERATURE REVIEW

In this section we briefly review cyber forensics concepts and investigate the implications of database technology in the cyber forensics domain.

2.1 Cyber Forensics Concept Explication

The field of cyber forensics is concerned with a series of activities in relation to investigation and law enforcement of cyber crimes. The activities include gathering, processing, interpreting, and analyzing digital evidence in the process of reaching a conclusive description of cyber crimes. However, "cyber forensics" is often interchangeably used with other terms such as digital forensics, network forensics, computer forensics and software forensics [15, 19, 22]. What makes people use the terms interchangeably without careful

discrimination is the connection of concepts embedded within those terminologies which can be represented in broad characteristics as Hall and Davis [8] summarize:

- Interrogation and testimony skills
- Chain of custody formalisms
- Data recovery techniques
- Investigation techniques providing input to process improvement; and
- Investigation techniques providing input driving security research.

Incidence and *attack* are important concepts used in designing a dimensional model in our paper. We briefly review these two concepts. Incident is broadly defined to describe possible criminal events and is often related to reporting the events to authorities. Shultz and Shumway [18] briefly state that an incident is defined as an "adverse event" that results in a security threat to computer systems and networks. Events can include any types of abnormal activities in computers or networks including "system crashes, packet flooding within a network, unauthorized use of another user's account, unauthorized use of system privileges, defacement of one or more web pages, and execution of malicious code that destroys data" [18]. Prosise, Mandia, and Pepe [17] succinctly define a computer security incident as "any unlawful, unauthorized, or unacceptable action that involves a computer system or a computer network". They summarize that those actions include the following activities:

- Theft of trade secrets
- Email spam or harassment
- Unauthorized or unlawful intrusions into computing systems
- Embezzlement
- Possession or dissemination of child pornography
- Denial-of-service (DoS) attacks
- Tortuous interference of business relations
- Extortion; and
- Any unlawful action when the evidence of such action may be stored on computer media such as fraud, threats, and traditional crimes.

Many of those events are in violation of public law that may lead to legal actions. Therefore, in forensics perspective, when an incident first occurs, reporting and sharing information about the incident with law enforcement authorities or appropriate industry members is important since it will serve as a critical component in the cycle of incident-investigation-prevention.

An incident is considered a precursor to an attack. Not every incident may lead

investigators to think there is an attack. For example, a series of incidents may be that a server at a bank, one a school, and one at a retailer crashed. All are incidents, but they may or may not be related. If suspicious, the incidents need to be tracked, since it might not be until later that a pattern may emerge.

An attack implies criminal intention in some way and is defined in relation to attacker, incident, and victim. For example, a Denial-of-Service (DoS) attack is a method that attackers use to prevent legitimate users from accessing to a system. An attack pattern is defined as any interrelationships among incidents that led to an attack or other misuse that may be observed by victims.

2.2 Database Technologies in Cyber Forensics

Database technologies including data warehouse, data mining, and OLAP could be adopted as part of a toolkit for cyber forensics in analysis of data obtained from occurrences of cyber crimes. We review some interesting uses of those database technologies in cyber forensics.

Early attempts at cyber forensics include, for example, basic profiling of criminal records. The FBI's Computer Crime Adversarial Matrix makes broad generalizations about the attributes of computer attackers based on stereotyping [21]. The Matrix focuses on four broad general characteristics: organizational, operational, behavioral, and resource. The Matrix also consists of three primary kinds of attackers with two sub-categorizations: crackers are divided into groups and individuals, criminals are categorized as espionage and fraud/abuse, and vandals are categorized as strangers and users. However, their system has not been evaluated as successful mainly because of the broad categorizations and lack of empirical foundation.

In order to overcome empirical deficiencies in cyber crimes database systems, recent developments tend to be collaborative efforts between business and law enforcement. Law enforcement agencies and a group of businesses gathered their resources to constitute an information-sharing system that is specifically designed to combat phishing[1]. Titled Digital PhishNet, the database aims at serving as a common information repository for law enforcement and industry [5]. Crime investigators from participating entities will input phishing-related information into a database at the National Cyber-Forensics & Training Alliance, where crime analysts from the FBI analyze patterns and pass that information along to agents. Some of the major sponsoring companies are Microsoft, America Online, Lycos, EarthLink, Network Solutions, and VeriSign. The FBI, the Federal Trade Commission, the Secret Service, the U.S. Postal Inspection Service, and some undisclosed U.S. banks are also participating in the project.

[1] Use of e-mail and/or fake web sites to gather personal information for the purpose of identity theft. The stolen identities will then be used in further fraudulent activities.

In the academic field, there are also enthusiastic endeavors to develop efficient systems for the use of cyber forensics. There are a number of studies that contribute to the automated criminal network analysis and visualization of the network [4, 24, 25]. In these studies, especially, data mining technologies play critical roles in finding structural properties of the criminal network such as subgroups in the criminal hierarchy, interaction patterns between those subgroups, and who plays the central role in the network. The studies commonly argue that knowledge about the structure and organization of criminal networks is important for both crime investigators and system developers to formulate effective strategies to prevent crimes. Brown and colleagues [2] also present a mining system for cyber forensics. With its image mining ability, they argue that, it provides the services for training the system to detect the image evidence as well as for correcting and refining search results. The Bayesian networks algorithm is used to provide a compact and efficient means to represent joint distributions over a large number of random variables and allows effective inference from observations. Hence, their mining algorithms offers methods to understand probabilistic and causal relationships through updating criminal knowledge based on supplied evidence.

3. DIMENSIONAL DESIGN OF DATA WAREHOUSE FOR CYBER FORENSICS

In this section, we present three different dimensional models that can be used for a data warehouse for cyber crime data.

3.1 Dimensional Models of Cyber Crime Data Warehouse

In developing cyber forensics dimensional models, we follow the Kimball's design process, which has been widely accepted in industry [12]. The design process consists of the following four-steps:

Step 1: Identify the business process, representing an activity we want to model

Step 2: Determine *the grain* of a *fact table*, representing the level of the detail of the data warehouse data record to be analyzed

Step 3: Identify *the dimensions* used to analyze the fact table

Step 4. Identify the *measure data* of the fact table

The first step is selection of a business process to model. We adopted the cyber crime investigation activity as our business process. Thus, our fact table will contain the measure data about cyber crimes.

The second step is to select the grain of the fact table. As the grain of fact table, we can think about two choices - incidence and attack. As we defined in Section 2.1., an incidence is an abnormal activity that may or may not result in an attack. Figure 1 shows the dimensional model whose fact table models an attack as the grain, while Figure 2 shows the one with the incidence as the

grain. If we just want to analyze cyber attacks that actually resulted in crimes or damages, we can use the Attack fact table. On the other hand, if we analyze cyber crimes at each incidence level, we can use the Incidence fact table. Since many incidences, whether they may or may not result in any attack, are still important to track down, we think the Incident fact table is more powerful. The Incident fact table, however, may result in a larger number of rows than the Attack fact table. We call Figure 1 the *Attack schema* and Figure 2 the *Incidence schema*.

The third step is to identify dimensions that can be used to analyze the fact table. In the Attack schema shown in Figure 1, the selected dimensions are Date, Attacker, Attacker Demographics, Attack pattern, Attack status, Law enforcement, Target, Target Agency, and Incidence Summary. We note that we created a dimension called Incidence summary that summarizes multiple related incidences from Incidence instance table. In Figure 1, the table entitled Incidence instance is called a secondary dimension or an outtrigger table [12] as it is not directly related to the fact table. Attacker demographic and Target agency dimensions are called mini-dimensions. They could have been included in Attacker and Target dimensions, respectively, but they were separated out to remove redundant data storage. In addition, by adopting them as mini-dimensions, they could directly participate in the analysis of the fact table.

In the Incidence fact table shown in Figure 2, the selected dimensions are Date, Attacker, Attacker Demographics, Attack pattern, Attack status, Law enforcement, Target, Target Agency, and Attack. Here, because the fact table grain is each incidence, we modeled attacks as a dimension. With this design, all the related incidences for a single attack can be easily aggregated for the attack.

The fourth step is to identify the measure data. We selected the same measure data for both Attack and Incidence fact tables. We first included *Cyber Crime ID*, which is the primary key of the source database from which the cyber crime data came. This attribute will be useful in connecting the source database and the data warehouse. This attribute thus supports real-time analysis using the data warehouse. Other selected measures are *Loss in Dollars, Cost for fix, Actual Downtime, Cost for Downtime*, and *Cost for Exposed Confidential Data*. Other measure data could be added, depending on the specific purpose of the data warehouse and analysis types.

The Attack and Incidence schemas show the basic framework of the cyber crime data warehouse. Each dimension needs to include detailed textual data that can be used for browsing, grouping, or filtering the data. We note that dimensions in a dimensional model are usually denormalized. Thus, all the data related to each dimension by one-to-many relationships can be denormalized into the dimension. The problem becomes more complicated if relationships between two data elements in a dimension become many-to-

many. For example, the following are many-to-many relationships; the tools used by attackers, political affiliations joined by attackers, institutions the attacker attended, multiple Websites attacked by attackers, skills owned by attackers, etc. These data could be useful in analyzing cyber crimes. In Figure 3, we show how to model the information within our framework. Figure 3 is based on the Incidence schema, but the many-to-many data elements can also be easily added to the Attack schema.

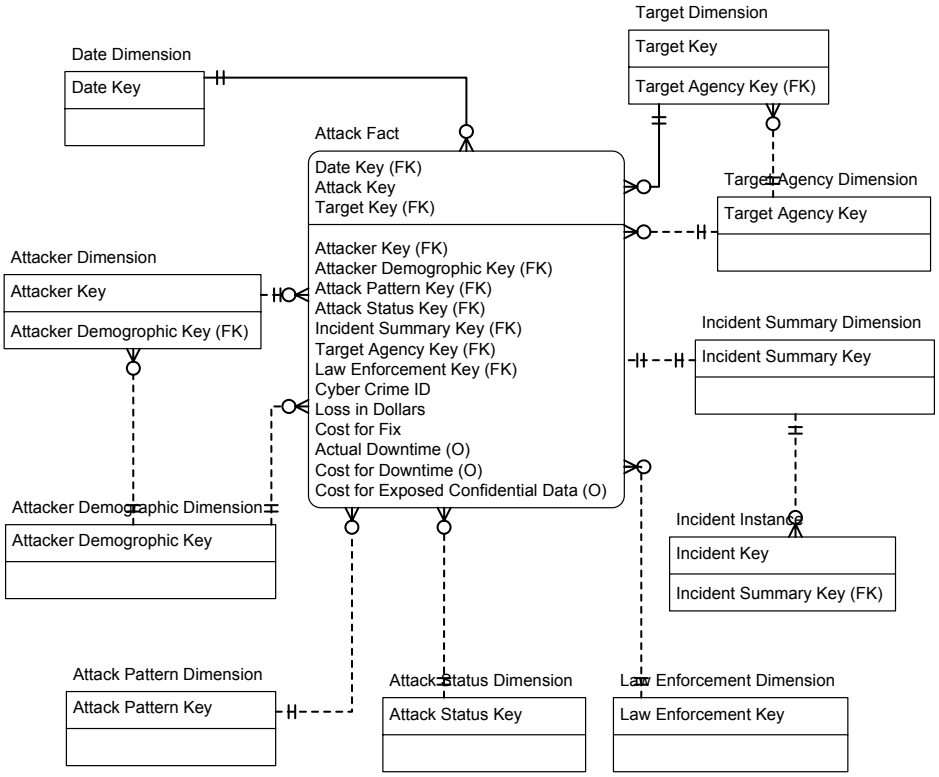

Figure 1. The Dimensional Model with *Attack* as the Grain (Attack Schema)

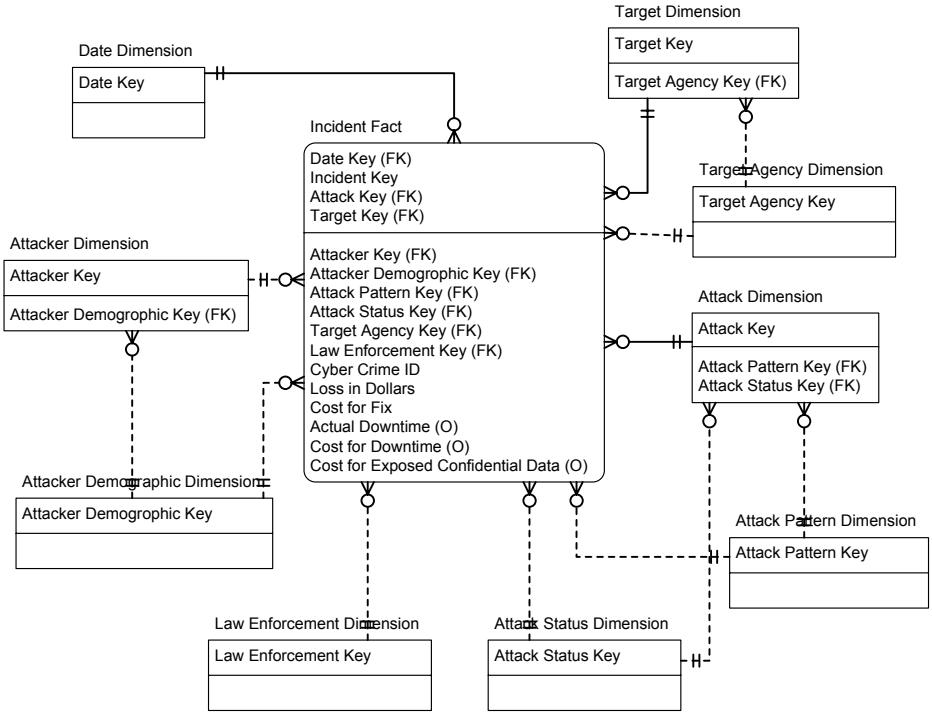

Figure 2. The Dimensional Model with *Incidence* as the Grain (Incidence Schema)

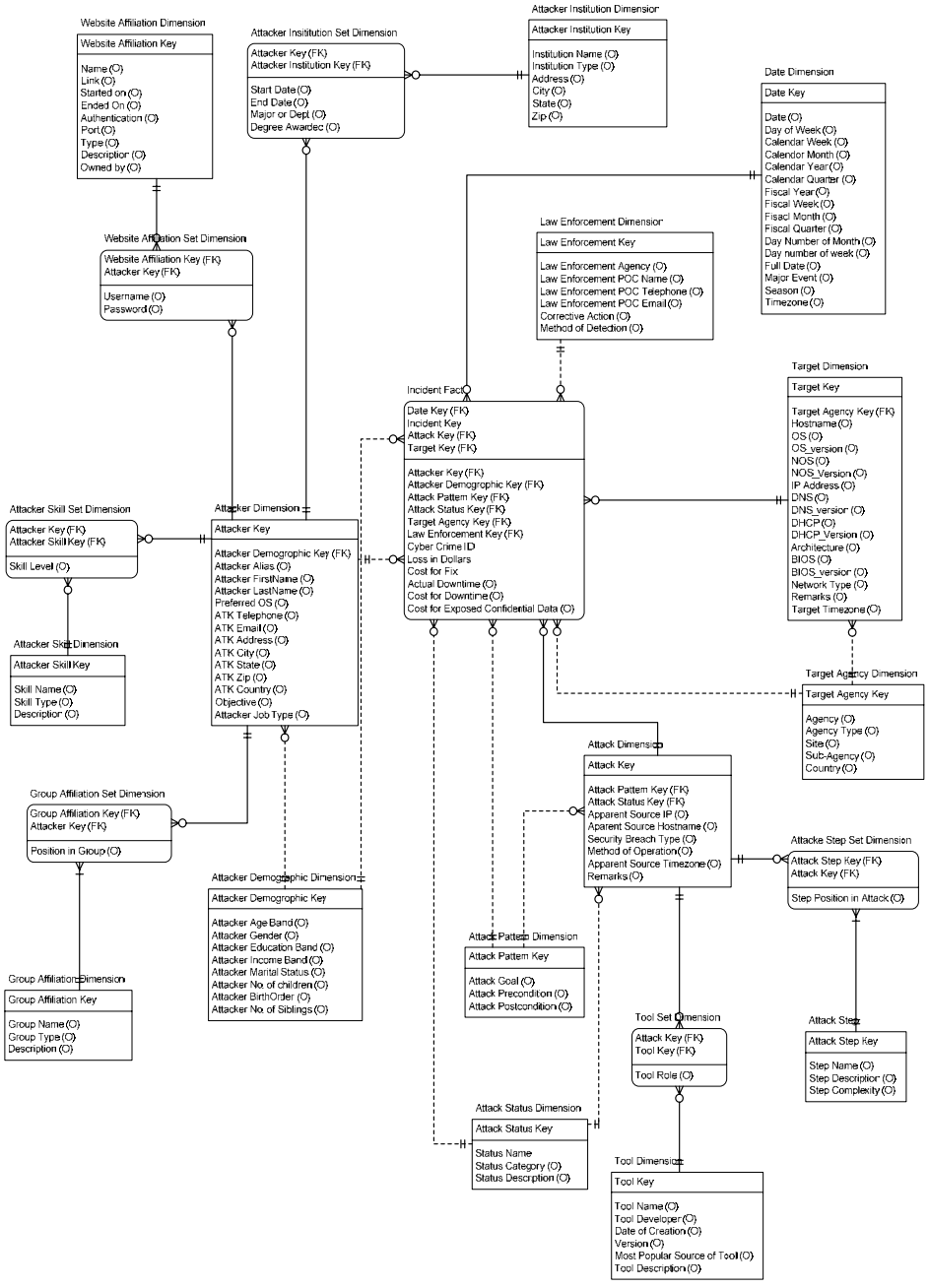

Figure 3. The Detailed Incidence Schema with Supporting Many-to-Many Information

In identifying the attributes of the dimension tables, we primarily used Kruse and Heiser [13] and Prosise, Mandia, and Pepe [17] as sources. They support the data points that investigators would have to collect. Other useful sources were described in various studies [10, 11, 16]. Howard and Longstaff [10] describe a taxonomy of terms with which to describe an incident. Further, their view of the decomposition of the incidents makes construction of analyses models much easier. Icove [11] presents the concept of classifying the criminal as well as the crimes into groups. This type of classification can be useful in predicting which computer criminal may lean towards a particular type of attack, or may tend to be part of a larger group. Last, Moore and his colleagues [16] present some very intriguing ways of describing the attacks themselves. Their use of attack profiles is very similar to the templating techniques used by military intelligence analysts.

4. USING OLAP AND DATA MINING WITH CYBER CRIME DATA WAREHOUSE

In this section, we present different ways of utilizing the data warehouse shown in Figure 3 using OLAP and data mining technologies.

4.1 Crime Analyses Using OLAP

OLAP enables a user to effectively extract and view information from different points-of-view. OLAP can locate the intersection of dimensions and report them.

From the dimensional model shown in Figure 3, we can perform a number of analyses. If our focus is the attacker, then we can run queries that would tell us who has performed what certain types of attacks in the past, who tends to work in groups, and who would be a leader in those groups. We can query for recidivism, levels of technical skills, and affiliations. This last would be of particular interest to those agencies involved in anti-terrorist and homeland defense effort.

Should the focus of our investigations be attacks, then the model supports queries that would show which agencies were targeted, what tools were used, what was expected to be gained, and what types of skills were required for a given type of attack. Target-related investigations would be able to query for agencies that were highly targeted, and if the attacks were successful or vulnerable. These queries could also help identify groups of hackers that might be involved in such targeting.

The model also supports analysis of vulnerabilities, specifically addressing what systems, architectures, and operating systems that were most vulnerable. While the press is generally full of articles saying which OS has a security problem, the query results would provide more reliable proof.

Other types of analyses that can be done are:

- How many invasion attacks have exploited a specific vulnerability each week?
- What time block tends to have the greatest activity by type of attack?
- Show attack counts per month by affiliated institutional backgrounds
- Show attack counts per period by tools used for each target system.
- Find attacks with the same attack category where at least 4 "Attack Steps" within the attack pattern matches the current case.
- Find attacks, across targets or agencies, which use the same apparent source IP or hostname
- Identify relationships between attack patterns and attack methodologies

Based on Figure 3, we further developed various types of crime analyses as in Cunningham, Song, and Chen [7], including Attack Analysis, Attack Pattern Analysis, Attack Step Analysis, Attacker Analysis, Attack Group Analysis, Incident Analysis, Target Agency Analysis, Tool Analysis, and Web Site Analysis by using some dimensions delineated in the dimensional model. Moreover, more meaningful queries can be designed in conjunction with other fields in the fact table. In Table 1, some examples of types of cyber crime analyses are presented.

Table 1. Types of Cyber Crime Analyses

Category	**Analysis**
Attack analysis	What kind of attack is the most frequent?
Attack analysis and tool analysis	What kind of attack is conducted with what kind of tools?
Attack pattern analysis	What type of conditions existed prior to attacks?
Attacker analysis	What are the demographics of well- known attacke
Attacker analysis and Attacker skill analysis	What type of skills do attackers use?
Attacker group analysis	Do the attackers belong to certain criminal groups? What are the group's characteristics?
Incident analysis	What are the incidents? How are they treated?
Target agency analysis	What types of agencies are attacked?
Tool analysis	What are the tools used for attacks?
Vulnerability analysis	How can we protect vulnerable attack points?
Web site analysis	Was there a unique ID entrance co-occurred with attacks?

4.2 Data Mining

Although OLAP is a key component of analytical process, it alone is not a sufficient tool for better understanding of cyber crime data and designing preventive methods against the cyber attacks. Some of the challenging issues cannot be answered by OLAP only. For example, to answer the following question "If a password theft attack happens, what is the type of attack most likely to happen next?" it is very difficult or even impossible to find a satisfactory answer based solely on the OLAP from the cyber forensic data warehouse. But the answer to the above question is very important to help the organizations/institutes reduce the damage caused by the attack. If password theft happens first, then we can take extra precautions concerning sensitive information.

Data mining techniques are used to identify patterns in a set of data. It looks for patterns where one event is connected to another event (i.e., association), patterns where one event leads to another later event (i.e., sequence or path analysis), and new patterns (i.e., classification). It can also offer visual combination of newly documented facts (i.e., clustering), and analysis of patterns in data that can lead to reasonable predictions about the future (i.e., forecasting) [20].

Data mining can be applied to various log analysis and intrusion detection systems [1, 10]. A lot of mining algorithms and methods such as association algorithm, decision tree, and others can be applied for mining the cyber forensic data warehouse to derive insightful knowledge rules to help understand the attacks and protect the network security. Below we briefly discuss some key algorithms and how these algorithms can help to solve some of the challenging problems for the cyber forensics. A deep discussion is beyond the scope of this paper.

(1) Association Rules: Association rule algorithms were originally designed to analyze market basket data to find correlations in items purchased together, as in "If a customer buys product A, what is the likelihood that he will buy product B?" In the cyber forensics, association rule algorithms can be used for analyzing the correlations in attack, target dimension, and attacker demographics, etc. For example, association rules can find out if there is a strong connection between authorization failure attacks with certain operating system platforms. This may suggest that the operating system of that platform may have some potential defects in the design, indicating the vendor may need to fix/redesign the authorization checking mechanism of the operating system.

(2) Classification Rules: Classification is a very popular data mining technique to build a model based on the training data and then apply the model to assign a new item to a certain class. There are many algorithms such as decision trees, neural networks, Bayesian networks, and probability theory for classification.

For example, to understand the denial of service attack, you can use decision tree algorithms to build a model, which may reveal such patterns as: If for the last 5 seconds, the count of one-way connections to the host IP address is 2000 from the same source IP, then most likely it is a denial of service attack.

5. DISCUSSION

In this section we discuss some issues that surfaced while investigating and designing a data warehouse for cyber forensics.

5.1 Legal Issues

In this section, we discuss legal issues related to data collection for cyber forensics. Although an in-depth review of the legal implications of cyber forensic is beyond the scope of this paper, legal issues are mandatory considerations in performing the forensics activities. Even with the recent changes made to laws governing system security, privacy, data collection, and monitoring, there are still a significant number of legal hurdles that must be crossed in the proper conduct of an investigation and prosecution of computer attacks [23]. While it is generally understood that the computers used at places of work (whether government at any level, corporate, or small business) are owned by the business, the users still have some expectation of privacy. Thus the cyber investigator or systems administrator must follow very distinct procedures to gather evidence that would be useful in the legal sense. They must ensure that it is collected properly (such as using bit stream copies) and preserved correctly (pulling the hard drives permanently).

Computer crime has an exceeding broad definition, covering areas of national security, financial fraud, theft, interruption of interstate/international commerce, industrial espionage, and racketeering. Title 18 of the US Code lists dozens of definitions of those particular areas that make up computer crime. Most parts of the areas directly related were significantly strengthened in the USA Patriot Act of 2001. This law amended many portions, easing the rules of prosecution, lowering criminal thresholds, and more clearly defining the rules of evidence, as well as clarifying the definitions of a number of specific crimes themselves. Indeed, some civil liberties experts find some of the changes to be nothing short of chilling.

Collecting data for cyber crimes databases is difficult for a variety of reasons [3]:

- Many security compromises go unnoticed for long periods of time
- Many companies do not report these crimes for fear of public embarrassment
- Many crimes, such as theft of proprietary information, are hard to quantify monetarily in terms of negative publicity, loss of competitive

advantage, or lost productivity when breaches occur or networks are down.

Thus, it is necessary to create policy and support to obtain crime data from various existing heterogeneous sources.

5.2 Data Population Issues

The data warehouse built on the dimensional model can also be populated via a number of steps.

First, data should be populated to a purpose-built relational database, populated interactively, perhaps via a web page, by law enforcement and computer incident investigative agencies. Data from this database could then be moved into the data warehouse using commercially available ETL (Extraction, Transformation, and Loading) tools.

We found that there are many agencies charged with the investigation of computer crime. We can import the data from these existing databases to our data warehouse. The existing crime databases range from the FBI's National Computer Crime Squad (NCCS) and the National Infrastructure Protection Center (NIPC) at the federal level, to state agencies operating as part of the state attorneys general or state police forces. There are also the incident reporting organizations, which include the DoD centers (ACERT, AFCERT, NAVCIRT, etc.), as well as industry-specific organizations such as the banking industry's Financial Services Information Sharing and Analysis Center (FSISAC).

Attempting to move all these agencies and organizations to a single collection point would be a tremendous effort in terms of both time and cost. Difficulties would also be faced in addressing the host of privacy and legal issues from the number of jurisdictions, as well as security classification problems.

6. CONCLUSION

In this paper, we have presented three dimensional models for a data warehouse for cyber forensics. We have also discussed ways of utilizing the data warehouse by considering the types of analysis as well as using OLAP and data mining technologies. We contend that our data warehouse model could be used as a central repository for analyzing various crime data and will enhance various OLAP and data mining activities against cyber crimes.

Further investigation on the cyber forensics dimensional model is necessary. The dimensional models we presented are draft models that were developed based on our conceptual analysis of literature. Our model should be further enhanced when the actual crime data are available. Further work could seek direct involvement of security specialists and law enforcement agencies, for in depth technical details as well as to ensure that the queries used do yield results that will be truly useful for both law enforcement agencies and prosecutors.

Among the areas that could be further researched are possible integration of the data warehouse with other forensic databases as forensic image databases, Virus and Worm Signature Database, Attack Tool Signature Database, Law Enforcement Cyber-Attack Contact Database, and Integrated Biometric Database, etc. The integrated comprehensive data warehouse will better serve its purposes.

7. ACKNOWLEDGEMENTS

The authors would like to thank you our students - Jojo John, Shelly Gupta, and Keith Gerritsen who contributed to a survey of literature and model developments for this project.

8. REFERENCES

1. Bhaskar, R. State and local law enforcement is not ready for a cyber Katrina. *Communications of the ACM*, *49* (2). 81-83.
2. Brown, R., Pham, B. and de Vel, O. Design of a digital forensics image mining system. in Khosla, R., Howlett, R.J. and Jain, L.C. eds. *Lecture Notes in Computer Science*, 2005, 395-404.
3. Cap, C.H., Maibaum, N. and Heyden, L., Extending the data storage capabilities of a Java-Based smartcard. in *Sixth IEEE Symposium on Computers and Communications*, (Hammamet, Tunisia, 2001), 680-685.
4. Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W. and Schroeder, J. COPLINK: Managing law enforcement data and knowledge *Communications of the ACM*, *46* (1). 28-34.
5. Claburn, T. Banks, law agencies team up to fight Phishing, 2004.
6. Common Digital Evidence Storage Format Working Group Standardizing digital evidence storage. *Communications of the ACM*, *49* (2). 67-68.
7. Cunningham, C., Song, I.-Y. and Chen, P.P., Data warehouse design to support customer relationship management analyses. in *7th ACM international workshop on Data warehousing and OLAP*, (Washington DC, 2004), ACM Press, 14-22.
8. Hall, G.A. and Davis, W.P. Toward defining the intersection of forensics and information technology. *International Journal of Digital Evidence*, *4* (1). 1-20.
9. Hannan, M.B., Turner, P. and Broucek, V., Refining the Taxonomy of forensic computing in the era of E-crime: Insights from a survey of Australian Forensic Computing Investigation (FCI) teams. in *4th Australian Information Warfare and IT Security Conference*, (Edith Cowan University, Perth, Western Australia 2003), 151-158.

10. Howard, J.D. and Longstaff, T.A. A common language for computer security incidents *Sandia Report*, Sandia National Laboratories, 1998.
11. Icove, D.J. Collaring the cybercrook: An investigator's view *IEEE Spectrum*, 1997, 31-36.
12. Kimball, R. and Ross, M. *The data warehouse toolkit*. Wiley, New York, 2002.
13. Kruse, W.G. and Heiser, J.G. *Computer forensics: Incident response essentials*. Addison-Wesley, 2002.
14. Kurlander, N. Fighting crime and terrorism through data integration, 2005.
15. Marcella, A.J. and Greenfield, R. (eds.). *Cyber forensics: a field manual for collecting, examining, and preserving evidence of computer crimes*. Auerbach Publications/CRC Press, Boca Raton, FL, 2002.
16. Moore, A.P., Ellison, R.J. and Linger, R.C. Attack modeling for information security and survivability *CMU SEI Technical Note*, CMU Software Engineering Institute, 2001.
17. Prosise, C., Mandia, K. and Pepe, M. *Incident response: computer forensics*. McGraw-Hill, New York, 2003.
18. Schultz, E.E. and Shumway, R. *Incident response: A strategic guide to handling system and network security breaches* New Riders, Indianapolis, 2002.
19. Solomon, M., Barrett, D. and Broom, N. *Computer forensics jumpstart*. SYBEX, San Francisco, 2005.
20. Thomsen, E. *OLAP solutions: Building multidimensional information systems*. Wiley, New York, 2002.
21. Turvey, B.E. *Criminal profiling : an introduction to behavioral evidence analysis* Academic Press, San Diego, CA, 2002.
22. Vacca, J.R. *Computer forensics: computer crime scene investigation*. Charles River Media, Hingham, MA, 2002.
23. Wegman, J., Legal issues in computer forensics. in *Allied Academies International Conference*, (New Orleans, LA, 2004), 45-49.
24. Xu, J.J. and Chen, H. CrimeNet explorer: A framework for criminal network knowledge discovery. *ACM Transactions on Information Systems*, *23* (2). 201-226.
25. Xu, J.J. and Chen, H. Criminal network analysis and visualization. *Communications of the ACM*, *48* (6). 100-107.