2007

Architecture for an automated IRC investigation tool

Dugald A. Brown
West Virginia University

Architecture for an Automated IRC Investigation Tool

Dugald A. Brown

A Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Computer Science

Roy S. Nutter, Ph.D., Chair
John M. Atkins, Ph.D.
Bojan Cukic, Ph.D.

Lane Department of Computer Science
and Electrical Engineering

Morgantown, West Virginia
2007

Keywords: architecture, automation, chat room, crime, identity theft, investigation, IRC

Abstract

Architecture for an Automated IRC Investigation Tool

Dugald A. Brown

Chat rooms provide a safe haven to discuss and at some times perform some types of crimes involving the internet. The anonymity provided by chat rooms combined with a general lack of records for most chat servers creates an environment where those who would perpetrate a crime may flourish. Current forms of investigation such as manually monitoring rooms or analyzing logs post-mortem require many man-hours with little to show or provide results too late to gather additional information.

In this thesis some methodologies are discussed in the design of a tool to automate the task of monitoring, logging, analyzing, locating users in a chat room discussing or taking part in criminal activities. Through the proposal of an architectural design for an internet chat room investigation tool which will automate the collection and analysis of chat room conversations, the burden on investigators to monitor the plethora of available chat room channels decreases. Using this design, a working implementation of the tool can be put into operation which will benefit investigating agencies currently using a manual or post-mortem approach to their investigations. This tool will reduce the number of man-hours invested in monitoring activity within IRC chat rooms or waiting for text searches of large chat logs to complete.

Table of Contents

Abstract ii
Table of Contents iii
List of Tables and Figures v
1 Introduction 1
1.1 Chat Rooms and Crime 1
1.2 Current Situation 2
2 Problem Statement 3
3 About Chat Rooms 4
3.1 What is a Chat Room 4
3.2 Personal Messages 5
3.3 Why Use a Chat Room 5
4 Chat Room Crimes 7
4.1 Some Typical Crimes 7
4.1.1 Identity Theft 7
4.1.2 Pirated Media 8
4.1.3 Fraud 8
4.1.4 Disaster Predators 8
4.1.5 Sexual Predators 9
4.2 Why the Appeal 10
4.2.1 Anonymity 10
4.2.2 Lack of Records 11
4.2.3 Legal Issues 11
4.3 Difficulties 11
4.3.1 Lack of Logs 12
4.3.2 Identifying Users 13
4.3.3 Legal Obstacles 14
5 Previous Investigations 16
5.1 Law Enforcement 16
5.2 Non-Official Groups 17
5.3 Post-Mortem 17
6 The Dugald Automated Investigation Tool 18
6.1 The Collection Module 19
6.2 Storage Module 21
6.3 Analysis Module 27
6.4 Alert Module 30
6.5 Locator Module 30
6.4.1 Paid-Account Rooms 32
6.4.2 Static-Account Rooms 32
6.4.3 Dynamic-Account Rooms 33
7 Initial Implementation 35
7.1 Additional Research 35
7.2 Database Schema Design 36
7.3 Collection, Parsing, and Storage 38

7.4 Keyword Analysis 38
7.5 Alerting Investigators 39
7.6 Locating Users 39
7.7 Return to Research 40
8 Summary 41
8.1 Benefits of Automation 41
8.2 Conclusion 42
8.3 Future Work 42
References 45
Appendices 47
Appendix A – Laws 47
Fourth Amendment 47
California Rule of Professional Conduct 1-400 47
US Code – Title 18, Part 1, Chapter 119, §2511 47
Appendix B – Detailed Module Diagram 48
Appendix C – Capture Sample 49
Appendix D –Preliminary IRC Message Classification 52
Appendix E – Collection Module Pseudo Code Algorithm 53
Appendix F – Database Schema 55
Appendix G – Database Schema Description 56
Appendix H – Extreme Programming 58
Appendix I Example Data through the Tool 59
Message with Credit Card Number 59
Message with User Joining the Channel 60
Message with General Text 61

List of Tables and Figures

Table 1 - Data Template 20
Table 2 - Keyword Examples 28
Table 3 - Phrase Examples.. 28
Table 4 - Regular Expression Examples.. 28

Figure 1 - Module Arrangement 19
Figure 2 - Collection Module.. 20
Figure 3 - Example Database Schema for Message Information 22
Figure 4 - Storage Module 26
Figure 5 - Analysis Module 27
Figure 6 - Alert Module 30
Figure 7 - Locator Module.. 31
Figure 8 - Detailed Module Diagram.. 48
Figure 9 - Detailed Example Database Schema.. 55
Figure 10 - Extreme Programming Model.. 58
Figure 11 - Use Case with Credit Card Number.. 59
Figure 12 - Use Case with Login Message 60
Figure 13 - Use Case with General Text Message.. 61

1 Introduction

1.1 Chat Rooms and Crime

Chat rooms provide a safe haven to discuss and at some times perform some types of crimes involving the internet. The anonymity provided by chat rooms combined with a general lack of records for most chat servers creates an environment where those who would perpetrate a crime may flourish.

In the case of some criminal and civil crimes, the internet has improved the means to commit said crimes. According to Charles Cote, the Country Manager for Fortinet Australia, a security solution provider, “In 2006, it was more financially rewarding in the United States to trade in illegal financial information than to sell drugs [8].” Crimes such as identity theft and digital media piracy have increased as the internet has gained popularity. If not the traditional news sources CNN or MSNBC, then within the technical news outlets such as the ACM Tech News, it is difficult for one to go a day browsing without finding some reference to an impending law or a recent crime having to do with these well known forms of internet crime.

This illegal activity within the internet presents investigators with the need to develop tools and methods to investigate crimes committed using the internet and bring the criminals to justice. Current forms of investigation such as manually monitoring rooms or analyzing logs post-mortem require many man-hours with little to show or provide results too late to gather additional information. This document will present one proposal for the implementation of a tool to automate the investigation of chat room based discussion and perpetration of criminal activity.

1.2 Current Situation

Currently, in order to detect criminal acts in the process of happening in a chat room, a person must be sitting at a computer logged into the same chat room viewing the act taking place. Also, after they take place, there is still a need for direct human interaction in collecting the evidence and analyzing chat logs. With either scenario, this creates a burden on man-power. To the best of our knowledge, the law enforcement entity, be it the local police or the Federal Bureau of Investigation (FBI), will need to reduce the number of men in the field so they can perform log analysis. Alternatively, they could under-man the analysis team in order to keep men in the field, or would be forced to hire more investigators devoted to analyzing the chat room logs. However, more people may raise budgetary concerns. In staffing a position such as this, the hiring agency should be aware that chat rooms operate twenty four hours a day for every day of the year with activity generally happening around the clock. For one agent at a time to monitor chat rooms as many at a time as he or she can handle around the clock will still costs 8766 man-hours per year (24 hours/day * 365.25 days/year) in lost time where a person is watching a screen and waiting for activity rather then pursuing a current investigation. This situation may lead some to ask why not automate this task to record and search logs at a later time with a search tool; but, as will be explained later, the time delay will allow the perpetrator time to cover his or her tracks. This predicament in cost and time raises the need for an automated tool that will reduce the workload for investigators and provide results in a timely manner.

2 Problem Statement

This thesis proposes the architectural design of an internet chat room investigation tool which will automate the collection and analysis of chat room conversations in order to reduce the burden on investigators. This proposal will handle only incidents of identity theft; more specifically, trading and selling credit card information. This scope was chosen in order to concentrate on the design of the tool rather than the complexity of seeking out the multitude of common internet chat crimes referenced throughout Section 4.1. The analysis tool designed within this proposal will be accomplished through the completion of five modules:

- Collection of Conversation Data
- Storage of Evidence and Analysis Results
- Analysis of Data being Captured and Stored
- Locating the Offending User
- Alerting Necessary Human Data Analysts

The tool will also focus on Internet Relay Chat (IRC) chat rooms to further reduce the scope for the initial design. IRC is an internet chat protocol defined under RFC 1459 [7] where users on many computers establish a connection to a remote service hosting the chat room. Additionally, IRC has been chosen for its notoriety in the world of online crime due to the anonymity it provides its users [8]. Keeping the need for a new method in mind, some background and previous approaches must be reviewed in order to better understand what is currently being implemented.

3 About Chat Rooms

3.1 What is a Chat Room

Chat rooms are web pages or web-enabled applications used on many internet sites to enable users of that site to join in conversation on a certain topic or topics. Internet chat room topics vary in scope and offer a wide variety of choices. Some chats have very narrow topics with potential examples such as "Help Using Your New Dell® Laptop" or "Understanding Moray Eels as Pets." Other chat rooms are very broad in their assigned topics. These subjects could be similar to such topics as "Vacation Stories" or "General Interests for Ages 40-50." Some chat rooms do not even have a topic of discussion, these rooms invite users to enter and discuss whatever comes to mind. Many servers offer said broad based chat room topics; an example of one such server would be would be the xchat channel (or room) on the ChatJunkies IRC server.

When users log into a particular chat room, they will either be assigned or asked to choose a username. A username, sometimes referred to as a handle or nickname, is a unique identifier for the person joining a chat room. Many users will keep the same username, if it is available, across all the various chat sites which they use. There are two distinct types of usernames used in a chat room: static and dynamic. A static username is generally found in a site where a user must register before using said site. This username is reserved for that person until the account is removed. A dynamic username however would be associated with those sites which provide or require a user to provide a username every time they log on to the room. This username does not necessarily need to be new and unique to those handles previously used on the site, only not in use at the time the person is logging into the site. This not only gives the user the option of changing their name at will, it also presents the possibility that more than one user on a given site may use that handle, albeit during different login sessions.

3.2 Personal Messages

When speaking of logging conversations of and identifying individual users, who could be identified as the criminal in question by their handle, it is worth noting that a chat room, for the purpose of this thesis, will not include personal messaging. A personal message, also known as an ‘instant message’ or IM, is one which is sent from one user to another, outside the public forum of the chat room. This process is quite similar to the way a chat room works. However, only two users will be involved in the conversation.

Personal messaging could be likened to whispering in the ear of a friend, passing a note or being on a telephone call. These are messages not meant to be read by the other users, especially those of the chat room. This paper will avoid personal messages due to legal issues inherent in intercepting a message where the parties involved have a reasonable expectation of privacy. This means that federal wiretapping laws and the Fourth Amendment of the United States Constitution [5] apply to conversations held using personal messages whereas a conversation in a chat room is similar to holding a conversation in a crowded room where one’s expectation of privacy is virtually nonexistent.

3.3 Why Use a Chat Room

Definitions and differences aside, why do people use chat rooms? Just like chat room topics, reasons users partake in the services offered by chat rooms vary greatly. Three broad-based uses for a chat room are the gathering of information, sharing experiences, and for socializing. Gathering information in a chat room relates to rooms with topics such as the previously mentioned rooms directed at users of a certain product. These users log into a specific room in order to learn more about that assigned topic of discussion in the chat room. The users in a room dedicated to sharing experiences could be new or expecting mothers and/or fathers or possibly be victims of a crime spree or disaster for two examples. Users of a social room typically enter in order to talk to their friends using the same chat room or make new friends of those users chatting on that particular channel. Rooms such as these also include rooms of common interest such as dating rooms or forums dedicated to a particular religion. These chat rooms types are general enough that in many cases one room seems that it could fall under multiple classifications. In observing the multitude of chat room topics and offerings, the appeal to various criminal types begins to grow clearer. In the next section, select varieties of chat room crimes will be discussed in order to better understand the situation before delving into the investigation techniques.

4 Chat Room Crimes

4.1 Some Typical Crimes

Chat rooms provide an ideal site for criminals of all varieties, albeit for those whose crimes are of a more physical nature, this medium only provides a means to locate the site of their next crime or discuss a previous crime rather than the opportunity to carry one out. So, to introduce the crimes, the discussion must be limited to a subset of crimes for which a chat room is ideal. These crimes include:

1 Identity Theft and Trading Stolen Information

2 Acquisition and Distribution of Pirated Music, Movies, and Software

3 Perpetrators of Fraud

4 Predators of Disaster Victims

5 Sexual Predators/Pornographers.

4.1.1 Identity Theft

Based on observation of the sample data in Appendix C, identity theft in a chat room appears to primarily consist of cashiers and perpetrators of phishing schemes. Those individuals/groups not typically found in rooms known for identity theft are those responsible for breaking into computer systems owned and managed by credit card companies or online businesses. These criminals, referred to as "The Kids" according to Cote, sell the stolen information in blocks of 5000 accounts at a time [8]. Chat rooms known for identity theft are populated by users attempting to sell stolen credit card information or users trading tips and experiences concerning the actual theft of the information. Again, based on observation, it appears much of this activity occurs out in the open rather than in private discussions which would be expected.

4.1.2 Pirated Media

The most common and well known forms of piracy are software, music, and motion pictures. Piracy has gained notoriety in recent years with crackdowns by the Recording Industry Association of America (RIAA) and Motion Picture Association of America (MPAA) on those downloading illegal copies of music and movies respectively. These rooms are inhabited by users attempting to distribute, either for free or for a fee, copies of their respective digital media. No matter the cost, this typically violates the copyright held by the original owners of these works. In some cases, this media will have been acquired legally, but in other instances the work in question originated from piracy as well.

4.1.3 Fraud

This particular brand of crime covers a very broad spectrum. This can include such crimes as the actual theft of identity information or creating fake credit cards. Another prevalent crime online is phishing, or setting up a fake login site in order to capture the login information of an unsuspecting user. The last type of fraud to be mentioned in this thesis will be overseas money-laundering schemes. These usually come in the form of a user being contacted by some form of foreign "royalty" asking to transfer some money to the person. Then this person is supposed to write a check and keep some portion of the money. This scheme is also sometimes used to get access to someone's bank account information.

4.1.4 Disaster Predators

There exist some chat rooms dedicated to survivors and families of disasters and crime. The unprofessional activity of attorneys seeking out clients in rooms such as these is not only a prohibited act in some jurisdictions, it is downright immoral. While not exactly a crime, there are regulations in certain states prohibiting this. One such example, California Rule of Professional Conduct 1-400, prohibits attorneys from soliciting clients in chat rooms for victims and families dealing with tragedy or disasters [14]. This prevents attorneys from causing undue grief or preying on those with diminished decision-making capacity due to emotional duress through unsolicited offers and urgings for legal action. This may well be another violation which finds itself within the confines of personal messaging, but offers a very specific search criterion within a chat room unlike the sexual predators which will be discussed next.

4.1.5 Sexual Predators

The last set of well known criminals that lurk within the confines of social applications such as Facebook, MySpace, or the chat rooms of this paper's concern are vendors and distributors of child pornography and sexual predators. These particular varieties of online crime have gained notoriety as the popularity of the internet has risen for children. Digital, print, and television advertisements abound to warn parents of the danger of allowing their children unsupervised access to the internet. These are well known crimes and easy to stop by attentive parents; however the criminals themselves are difficult to track due to the personal nature of the crimes being committed. Criminal acts of this nature typically do not happen in the chat room itself; these predators will only hunt for their victims in the chat room. Crimes committed by sexual predators are an example of those which begin within the chat room, but typically result in the perpetrator exiting to personal messages before any actual crime is committed [13] rendering automated criminal investigations within chat rooms virtually useless due to the aforementioned wire tapping laws. The live-investigator methods in use for detecting pedophiles and child pornographers appear to be a more appropriate approach to this task. As reported in the Hartford Courant, police departments are investing in units dedicated to investigating and arresting perpetrators of these crimes [16]. These officers must log in live and be able to speak to the other users of a chat room in order for their work to be successful. The interactive requirement of this type of investigation does not lend itself well to an automated investigation method.

4.2 Why the Appeal

The previous section discussed the many types of internet crimes which occur in chat rooms on the web. Now, the question at hand is: why are these individuals and groups drawn to the medium that is the internet chat room? These are reasons such as:

1 User Anonymity

2 Lack of Conversation Records

3 Laws Restricting Logging

4.2.1 Anonymity

Anonymity is a luxury that can not be afforded in person beyond an individual not revealing his or her name. The handle or username that a person uses while online acts as a veil to mask the user's true identity. This quality goes hand-in-hand with a chat room's feature for giving users the ability to pretend they are someone they are not. While in a chat room users may assume the identity of someone by another name, a person of another gender or age group, or even create a whole new image or personality for themselves. This offers an initial layer of protection for the user which begins the multiple layers of discovery law enforcement must peel away during their investigation in order to determine the true identity of criminals in an internet chat room.

4.2.2 Lack of Records

The lack of conversation logs recorded by many chat sites is another difficulty possibly facing law enforcement officials in an investigation; and, by the same right, this is an aid to criminals in the online chat world due to the lack of readily available evidence. This lack of logging by server owners leaves the task to groups conducting investigations or their associated investigation tools to capture the logs that would be used in a criminal prosecution.

4.2.3 Legal Issues

There exists a potential need for investigators to collect conversation logs in chat rooms for evidence. However, law enforcement officers need to consider the legal aspects associated with this action. As stated previously, wiretapping laws do not apply in chat rooms, for the reason that speaking in a chat room is akin to speaking in a crowded room. Officers only need be aware of restrictions within the specific jurisdiction in which the case applies. An example of one such restriction is a New Hampshire law requiring both parties involved to consent to being recorded in order for the log to be admissible as evidence in court [2].

4.3 Difficulties

When tracking the perpetrator of a crime in an internet chat room from the handle used in the crime back to the user sitting behind a keyboard, there is an array of difficulties facing law enforcement officers that also coincide with some of the factors that lend appeal to the criminal element. These obstacles must be considered very carefully when determining how to proceed with an investigation.

There are three prevalent categories in which these issues may be grouped; these categories are:

1 Lack of Logs Maintained by Chat Owners

2 Identifying Users Based on Logon Information

3 Legal Obstacles Preserving the Rights of Chat Users

4.3.1 Lack of Logs

The lack of logs kept by owners of chat room hosting servers is perhaps the easiest to overcome of the challenges encountered by investigating officers in a criminal case. This obstacle comes in two forms: no records or records with a short life-span. Many sites only keep records of users rather than the discussions in chat rooms. On top of this, if a site does keep records of the conversation taking place in their chat rooms, most likely they will not exist for long. It is common practice in the business world to only keep certain types of data for a set period of time; that is to say that discussion records may be destroyed in the course of the business's day-to-day operations. This inconvenience may be overcome by the investigating officer in the chat room through implementation of a purchased or developed logging application (such as the algorithm proposed in Appendix E) for use while monitoring in the room. Leaving the investigator to capture logs also provides the benefit of having a copy of the evidence already stored for later prosecution. These logging applications could capture the text of the chat room conversation (example shown in Appendix C), or it could capture screen-shots to create a digital video similar to a flip book. This scenario leaves an investigating officer with the task of building a case and obtaining a warrant to gather the information on a company's server in a timely manner so as to acquire user information before it is removed from the server. This data removal leaves law enforcement less time to build a case, thus creating an additional challenge.

4.3.2 Identifying Users

The next major obstacle investigators encounter in their endeavors to trace a username back to the identity of a real person is being able to prove who was sitting at the keyboard at the time of the criminal act. This obstacle may result through several means:

- **Dynamic Usernames** – Allowing users to create a new handle every time they enter the room preventing investigators from associating a username with a particular individual that frequently uses that name.
- **Password and Identity Theft** – Stolen passwords or identities can allow a malicious user to create an account for a chat room using another person's identifying information, thereby impeding an investigation while the investigators attempt to find the true user.
- **Poor Computer or Network Security** – This could indicate something as simple as walking away from an unlocked account for a few moments or saving their password in a public location such as an internet café. Additionally, many home users now implement wireless networks within their home; however, some fail to secure their network. These unsecured networks offer the opportunity for a malicious user to park near the house and make use of this network, thus the evidence points back to the owners of the network rather than the malicious user.

- **Redirection by a Malicious User** –.This would involve the malicious user spoofing, or changing, their IP or MAC addresses. In the case of IP addresses, this could lead investigators to the wrong location; and in the case of the MAC address, the wrong computer at the correct location.

But even after these issues, the case comes down to either proving that a particular person was at the keyboard at the time of the crime such as those in an internet café, or determining for a fact that a crime was committed on a home PC and identifying the responsible party.

4.3.3 Legal Obstacles

Among the considerations, officers of the law must determine if the evidence gathered may be used in a court of law. All Americans, whether their acts have been deemed criminal or not, are constitutionally guaranteed against unlawful search and seizure [5]; meaning that the law enforcement official involved must obtain a warrant from a judge specifying precisely when, where, and what is to be searched. This constitutional protection also extends to the digital world where wiretapping (the act of intercepting wired transmission) is prohibited without a signed warrant. However, in the case of chat room investigation, the Supreme Court has ruled that conversations within a chat room offer no expectation of privacy [1]. The specific text of the ruling in *US v. Charbonneau* is as follows:

> *There is a limited expectation of privacy for emails sent/received on AOL. Email is like regular mail. When it is sent, the sender's expectation of privacy diminishes. Once an email (like a letter) is received the recipient controls it and the sender's expectation of privacy is gone. There is even less expectation of*

> *privacy in a chat room. When someone posts in chat room, he/she runs the risk that an undercover agent is in the chat room. Therefore, anything said in chat room is admissible in court* [1].

This ruling allows for the use of internet chat logs to be admissible as evidence in crimes where a chat room discussion was involved. In the case of a personal message, this would not be admissible without the perpetrator first inviting the investigating agent into the private chat. However, exceptions do exist; in states such as New Hampshire, there exist laws maintaining that even a chat room log is inadmissible as evidence in court unless all parties involved have consented to the recording of the conversation [2]. Therefore, investigating officers must be aware of laws in the jurisdiction of the offending user before pursuing criminal charges; thereby adding a degree of complexity in getting the user to consent to logging without being accused of entrapment.

5 Previous Investigations

In considering how to create faster and more accurate methods and tools to aid in the discovery, location, and conviction of internet criminals, some investigation is necessary into the past approaches of law enforcement officers. Among the three criminal acts and the rule from the Code of Conduct previously mentioned, investigations into child pornography and pedophilia receive the most mention. However, the current approaches undertaken require large commitments on the part of the agencies.

5.1 Law Enforcement

Currently, techniques to locate a pedophile or distributor/collector of child pornography by law enforcement agencies appear to be manual [9]. This work is difficult, tedious, costly, and time consuming [10]. This act is similar to going undercover for narcotics or weapons sales in order to catch the suspect implicating him or her self. This requires training as is outlined in section 3.1 of *Investigating 'Internet Crimes Against Children' (ICAC) Cases in the State of Florida* [4]. This paper also mentions specialized computer equipment and networks not associated with any law enforcement agency that the investigating officers are required to use. This equipment, while not prohibitively unique poses another budgetary problem for the investigating agency. The computers used for this investigation must have no hardware or software ties to a governmental agency due to spyware frequently implemented by the offender in order to judge his or her safety. This lack of revealing software extends as far as the network through which the computer accesses the internet.

5.2 Non-Official Groups

Another approach being seen more and more often is rather than law enforcement, a member or multiple members of parents' or children's rights groups posing as children online then collecting evidence to turn over to law enforcement officers upon approach by a suspected sexual predator. One such example of this is the watchdog group Perverted Justice with its force of volunteers [17]. With the proliferation of chat sites available to the public, requiring an investigator, sworn or private citizen, to watch each one is not sensible. As stated in the previous section, one set of chat rooms would require 8766 man-hours for a single person at a time to monitor the rooms for a year.

5.3 Post-Mortem

In the case of searching for instances of piracy and trading information obtained through identity theft, these processes already exist in a minimally automated way. The current problem exists in the form of timeliness. Institutions exist which capture chat logs as mentioned in the previous section. However, these logs are being examined after the fact; that is, the messages are not being examined as they arrive. These institutions are using tools such as GREP to search the text files being captured. GREP searches one or more input files for lines containing a match to a specified pattern. This leads to the issues of not searching a library of terms and potentially not being able to acquire user logon records captured by the chat provider that could possibly lead to the identification of the suspect in question due to timeliness involved in analyzing captured logs then contacting the chat provider before records are deleted in the regular course of business. This leads us to the need for not only an automated approach, but one which examines the logs in real-time as well.

6 The Dugald Automated Investigation Tool

In designing a chat room investigation and analysis tool, considerations must be made to define the scope of the problem. For the initial design, the chat room tool will be specified for any Internet Relay Chat (IRC) chat room. Along with the specifications for this type of chat room, exploration will be made into possible routes and modifications needed for individual tool modules to accommodate other types of chat rooms in the future. IRC has been chosen as the example because its chat rooms which play host to a plethora of criminal activities. The anonymity provided by IRC also lends itself well to undercover investigation techniques.

The complete tool will house five subsystems (displayed in Figure 1 on the next page with the detailed diagram of Figure 9 in Appendix B) which are:

1 The Collection Module
2 The Storage Module
3 The Analysis Module
4 The Alert Module
5 The Locator Module

These five subsystems working in tandem will comprise the investigation tool which will provide information with the potential to increase the number of arrests and reduce the investigator workload related to chat room crime investigations.

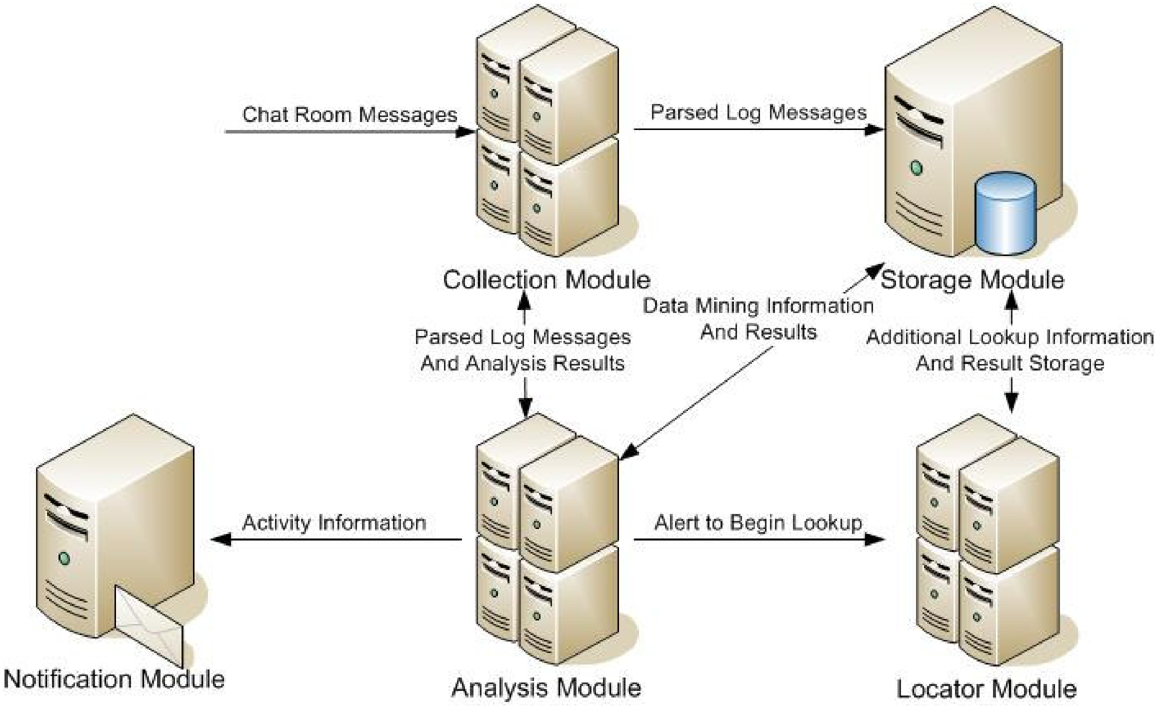


Figure 1 - Module Arrangement

6.1 The Collection Module

The collection module of the Dugald Automated Investigation Tool (DAIT) will be the subsystem illustrated in Figure 2 which captures data and parses it in preparation for analysis and storage. With respect to the IRC chat rooms, the collection module will make use of the protocol specified in RFC 1459 to connect to a specified channel on an IRC server and begin capturing the conversation as messages arrive. This data will arrive in a format as demonstrated in Table 1 and the sample of logs in Appendix C.

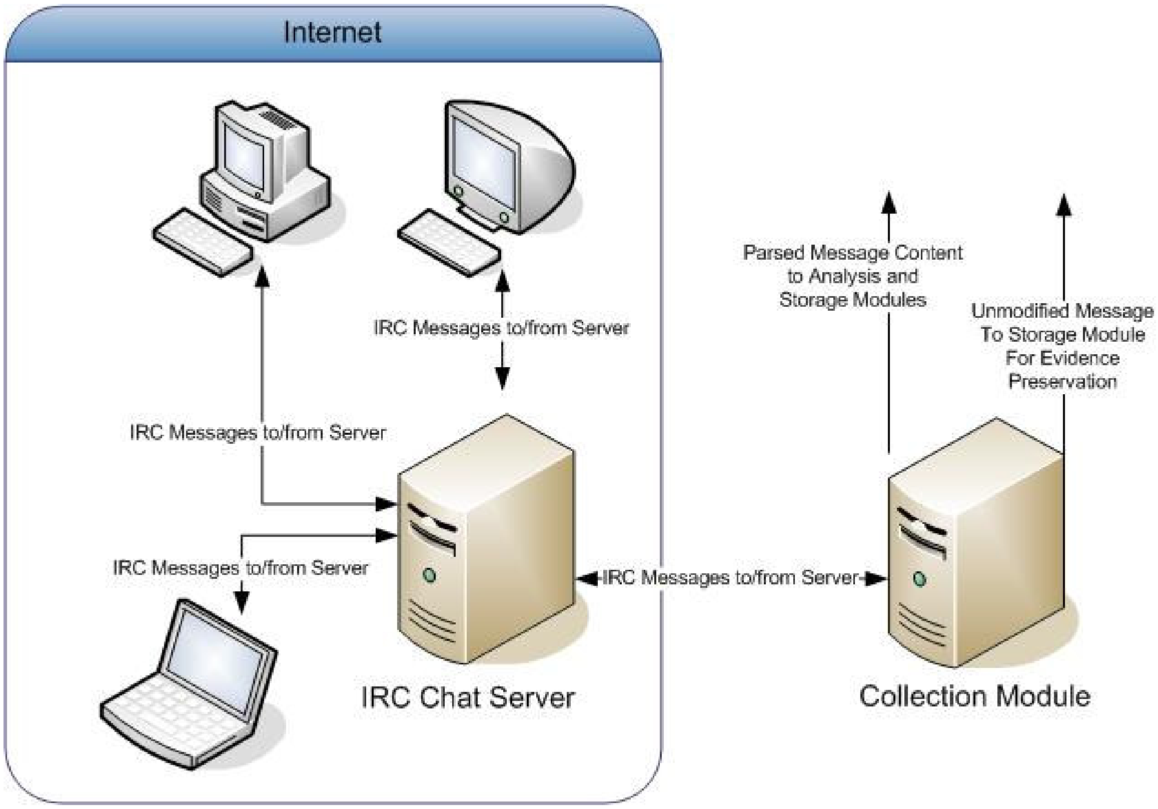

Internet
IRC Messages to/from Server
IRC Messages to/from Server
IRC Messages to/from Server
IRC Chat Server
IRC Messages to/from Server
Parsed Message Content to Analysis and Storage Modules
Unmodified Message To Storage Module For Evidence Preservation
Collection Module

Figure 2 - Collection Module

Table 1 – Data Template

Nov 20 22:10:22 <userA>	message.
Nov 20 22:10:23 <userB>	message.
Nov 20 22:10:35 <system>	announcement.
Nov 20 22:10:45 <userC>	message
Nov 20 22:11:04 <userB>	message
Nov 20 22:11:12 <userC>	message

The Collection Module should be written using an object-oriented programming language such as Java or C++. These languages provide built-in libraries for making connections over the internet and therefore lend themselves well to the solution. After collecting the message data from the IRC server, the Collection Module will store an original unedited copy of the message for evidence and parse the messages to find dates, usernames, hyperlinks, etc. Once the data has been parsed, it is simultaneously sent to the storage module for logging purposes and passed to the analysis module for immediate keyword detection. For investigating other chat room types which do not provide a protocol to design by, the design of this module must be modified. Some chat room types provide a convenient logging tool such as that functionality provided by the XChat client for IRC. By using this module within the DAIT, the data sent from the server will be broken into usable components to allow for eased analysis and extended storage for later examination or use as evidence.

6.2 Storage Module

The data storage module of the DAIT will be used to retain logs in their parsed form for long-term storage and enhanced search ability. A database, which is a collection of data arranged for ease and speed of search and retrieval, lends itself well to the requirement. In order to manage this database, a database management system (DBMS) would be the tool implemented to fulfill the role of this module of the investigation tool. A DBMS is the software implemented for the purpose of managing the databases created.

Before the data could be entered into the database it would need to be broken into usable fragments. Analysis of a chat log from the irc.hackcrew.cc IRC server offers a glimpse into a possible breakdown of the message types which are listed in Appendix D. This server (irc.hackcrew.cc) was chosen for message-type analysis due to a readily apparent proliferation of illegal activity which occurs within the channels of this particular server. It is worth mentioning, that when monitoring a new chat room type or server, the data may need to be reexamined in order to find new parsing rules based on the format of the data as presented by this new chat room server.

As new chat server logs are analyzed, message types will evolve. This will cause a need to change the parsing algorithm of the capture module as well. However, due to the very basic design of the database schema, new message types should not impact the design of the schema unless new information is extracted (similar to the hyperlinks and actions currently used). The schema of our database for the messages, illustrated in Figure 3, is comprised of nine tables and the associated relationships between the tables. This version does not include storage for analysis results as the data stored would be specified by the agency using the tool, however, a more complete version of the schema (still barring analysis results) is offered in Appendix F.

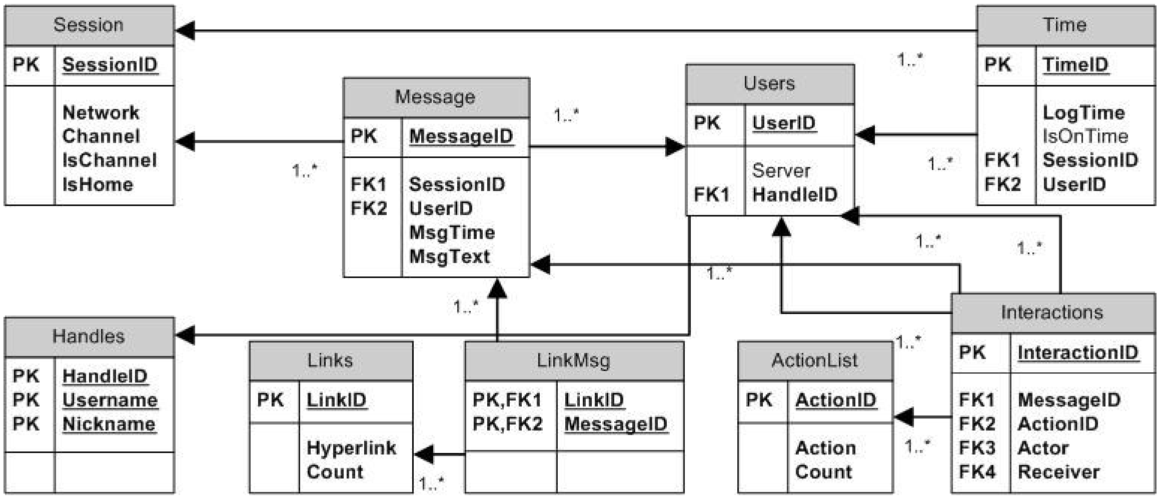


Figure 3 - Example Database Schema for Message Information

The first table in the database, ***Session***, is comprised of five attributes; the *SessionID*, *Network*, *Channel*, *IsChannel*, and *IsHome*. This table will be used to store information about the chat rooms being monitored. Session will posssess the following attributes:

- *SessionID* – An integer representing a unique number identifying the record in the table.

- *Network* – A text field in which to record the name of the IRC network (server).
- *Channel* – A text field identifying the name of the channel or chat room in which the log took place.
- *IsHome* – A Boolean field identifying if the room is the landing page of the server.
- *IsChannel* – A Boolean field identifying if the room is a private message.

Next, the ***Handles*** table is a table comprised of three attributes used for recording username and nickname combinations and consisting of the followint attributes:

- *HandleID* – An integer representing a unique identifier for the combinations.
- *Username* – A text field indicating the name the user registered with upon entering the chat server.
- *Nickname* – A text field containing the display name the user will be using within the chat room.

Then there is the table ***Users*** where user logon locations will be found. This table consists of the following three attributes;

- *UserID* – A unique integer identifier for records in the table.
- *HandleID*I – An integer field relating to corresponding *Username* and *Nickname* combination on the ***Handles*** table.
- *Server* – The information on a user's location which is provided upon login. The message during login typically has this format:
- *May 10 11:42:16 * Nava (~njvc@1CCED6DD.204D9C25.F48CD9E7.IP) has joined #ccpower*

- Nava is the user's display nickname, njvc is the username, and everything after the '@' can be used to identify the user's location. This information may be used later for a WHOIS lookup or, in the case of an IP address, may be used to find the user's location which will be discussed further in Section 6.4.

The next table, ***Message***, is again made of five attributes and is used for recording the message text from a room. These attributes follow:

- *MessageID – An identifier* stored in the same integer format as the other identifying attributes.
- *SessionID – Is* referencing the identifying fields of the ***Session*** table.
- *UserID* – Is referencing the identifying fields of the ***Users*** table.
- *MsgTime* – An attribute to store the time the message was sent
- *MsgText* – An attribute to store the text of the message.

Following this is ***Time***, a table used to record when a user logs on and off within a logging session with the following attributes:

- *TimeID* – A field uniquely identifying the records of the table.
- *SessionID* – A field referencing the identifying field of the ***Session*** table.
- *UserID* – A field referencing the identifying field of the***Users*** table.
- *LogTime* – A field saving the timestamp for the logon or logoff time.
- *IsOnTime* - A Boolean field storing a true value for a logon timestamp and a false value for a logoff timestamp.

Now for the ***Links*** table where information will be stored about hyperlinks occurring within the message text in a chat room in order to allow data analysts the opportunity to review the most prevalent links found during investigations.. This table will have three fields:

- *LinkID – A* field containing an integer identifying value.
- *Hyperlink* – A field to store the actual hyperlink text.
- *Count* – A field to tally the number of times this link has appeared in logging sessions.

This leads to the ***LinkMsg*** table which will relate hyperlinks to the messages in which they appeared thus allowing investigators to view the context involved with the link appearing in the chat room as well as connect it to a particular user sending the message. This table will only contain two attributes:

- *LinkID* – A reference to the specific hyperlink from the ***Links*** table.
- *MessageID* – A reference to the message in which the hyperlink appeared.

Next is the ***ActionList*** table which contains three attributes to describe interactions between users in a room providing more social insight into the behavior of users in these rooms rather than to be used as evidence to be used for criminal investigation. The attributes for this table are:

- *ActionID – A* field containing an integer identifying value.
- *Action* – A fieldholding the text of the interaction between two users.
- *Count* – A field to tally the number of times this action has appeared in logging sessions.

Finally, the ***Interactions*** table which is comprised of five attributes to describe the users, the message, and the action involved in a user interaction. This table is comprised completely of references to other tables within the database with the exception of the identifying key. This allows for a query that can show the complete set of interactions a particular user or users went through in a chat room or rooms. The attributes of this table are:

- *InteractionID* – A field which will identify the individual records.
- *MessageID* – A field to denote where the action appeared in the chat.
- *ActionID* – A field to show which particular action took place during the interaction.
- *Actor* – This field denotes either the user sending the message or the user executing said action.
- *Receiver* – Denotes the user to which the action was directed.

These tables and the associated relationships define the database which will contain the information retrieved during logging that will ultimately lead to detection of identity theft in the chat room. The data flow of the module is demonstrated in Figure 4 below.

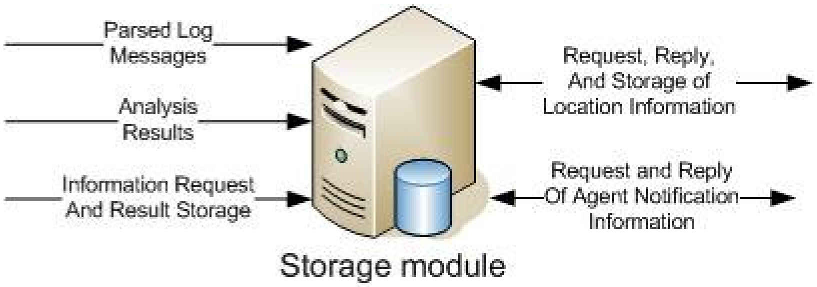

Parsed Log Messages
Analysis Results
Information Request And Result Storage
Request, Reply, And Storage of Location Information
Request and Reply Of Agent Notification Information
Storage module

Figure 4 - Storage Module

The additional information being saved such as links and interactions could possibly lead to additional evidence. If a user is detected who is distributing credit card numbers, then the interactions may determine other users to watch more closely. Or, if the user distributes hyperlinks, then these links may lead to sites supporting additional criminal activity.

6.3 Analysis Module

The Analysis module will be the portion of the investigation tool responsible for determining if a crime is being referenced to within a chat room conversation message. This module will be comprised of two components; the keyword analysis and the database analysis portions. As illustrated in Figure 5 below, the Analysis Module will function as two separate pieces working toward a common goal. The keyword analysis portion will search all incoming data for criminal activity identifiers, while the database analysis will use the large quantities of messages in the database to extrapolate new information. If either section finds a match to that subsystem's detection rules, then the information will be sent simultaneously to the locator module and to the alert module which will indicate to the investigating agent that a potential criminal act has been found.

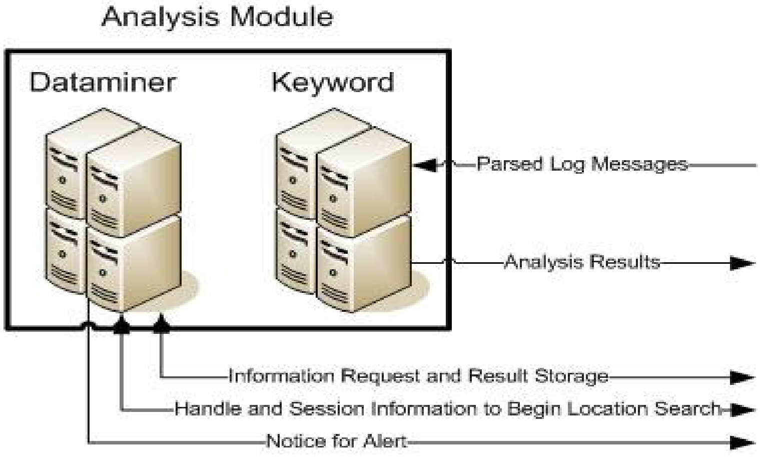

Analysis Module
Dataminer
Keyword
Parsed Log Messages
Analysis Results
Information Request and Result Storage
Handle and Session Information to Begin Location Search
Notice for Alert

Figure 5 - Analysis Module

The keyword analysis portion of the Analysis Module will be responsible for sorting through all the data being parsed by the Collection Module. As well as sending data to the Storage Module, the Collection Module will send data to the keyword analyzer to be checked for any evident markers to indicate criminal activity. The messages will be checked against dictionaries which will contain pre-established

- **Key Words** – Words typically found in messages referencing activity concerned with stolen identity information or trading of financial information such as bank accounts or credit cards. Examples found in Table 2.

Table 2 - Keyword Examples

Visa	Mastercard	Discover
Password	PIN	SSN
CVV	Cashier	CVN

- **Key Phrases** – Sets of words that, when found together, typically indicate message activity concerned with stolen identity information or trading of financial information such as bank accounts or credit cards. Examples found in Table 3.

Table 3 - Phrase Examples

American Express	Expiration Date	Mother's Maiden Name
Security Question	Security Word	Social Security Number
Routing Number	Account Number	Expiration Date

- **Regular Expressions** – Regular expressions are sets of rules used to identify certain types of text. Good examples are the varieties of credit card number all matching the same format. Regular expressions will be used to seek out these strings within the messages received by the DAIT. Examples found in Table 4.

Table 4 - Regular Expression Examples

[0-9] [0-9] [0-9]-[0-9] [0-9]-[0-9] [0-9] [0-9] [0-9]	{[0-9] [0-9] [0-9] [0-9]-?} {[0-9] [0-9] [0-9] [0-9]-?} {[0-9] [0-9] [0-9] [0-9]-?} {[0-9] [0-9] [0-9] [0-9]-?}
{CVN|CVV}{[0-9][0-9][0-9][0-9]?}	{[0-9][0-9]}{-|/}{[0-9][0-9]}?{[0-9][0-9]}

Upon finding these indicators, this subsystem will simultaneously send any needed information to the Locator Module in order for the user's location to be determined as well as to the Alert Module so that the investigating officer may be alerted to possible criminal activity.

The database analysis portion of the Analysis Module would be responsible analyzing messages in the database with respect to all other information being maintained about the messages being captured from chat room conversations. This ability could open up opportunities such as discovering a user's involvement in a criminal network that is not readily apparent from reading messages as they appear. An additional functionality potentially offered by this portion of the Analysis Module would be the ability to reclassify the message types used by the Collection Module for parsing incoming messages.

Study conducted by Eiman M. Elnahrawy on text categorization [10] provides a advantageous look into a possible approach to the database analysis portion of the Analysis Module. In the study, Elnahrawy discusses the advantages of this processes as well as mentioning current uses such as classifying emails and internet searches. The current keyword approach to chat message analysis using tools such as GREP is not appropriate for making assessments about conversations in chat rooms where context is of essence. The paper discusses three particular algorithms which were compared: the Naïve Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) methods.

This study determined the KNN and Naïve Bayes perform better than SVM method. However, KNN is slower than the Naïve Bayes which is likely due to the size of the data set used during experimentation. This algorithm lends itself well to the task of classifying the chat messages being collected in the Storage Module. The algorithm chosen will require a more extensive study using a data set significantly larger than the 20,000 messages in Elnahrawy's study. By implementing this within the Analysis Module, the tool has the potential to grow beyond the original design of the system.

6.4 Alert Module

The final subsystem in the DAIT, the Alert Module, will be used in order to notify the agents responsible for the particular crime being sought. In the case of identity theft or stolen credit card numbers, the crime would fall into the interstate commerce category and would most likely find itself in the lap of the FBI. However, as the system expands, the crimes for which the tool is searching will evolve and may in time come to involve local law enforcement in the jurisdiction of the offending user. This module of the tool will accept a notice alert about the activity discovered in the Analysis Module, retrieve information from the Storage Module for the appropriate agent based on the crime type and location, and send the information about the discovery to the appropriate mailing list for the crime. This module would have the ability to send email to computers and text messages to cell phones and pagers with the information the mailing list deemed necessary.

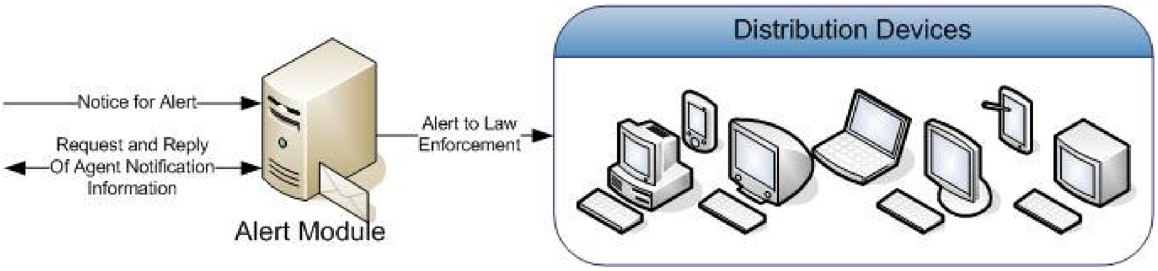


Figure 6 - Alert Module

The layout of the module will flow as illustrated in Figure 6 on the previous page. Hopefully in time the function of the system could evolve to the point where it could notify the patrol car nearest to the criminal incident in question for a speedy response.

6.5 Locator Module

The Locator Module subsystem will begin operation after the two components of the Analysis Module review the data being captured from the chat room. When a potential offense is located, the Analysis Module will pass the user's ID to the Locator Module. The Locator Module will then use any information, if available, in the Storage Module (server names or IP addresses for example) to attempt to find the user's physical location. This arrangement is illustrated in Figure 7.

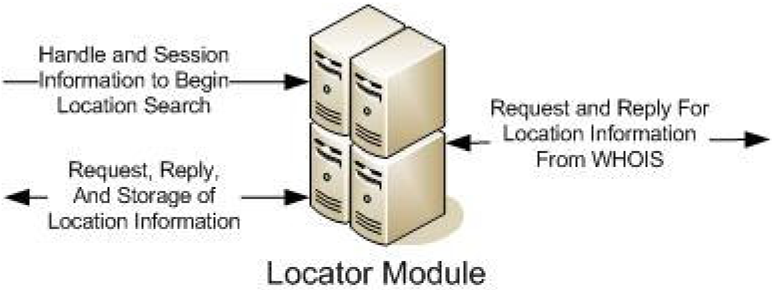

Handle and Session Information to Begin Location Search
Request, Reply, And Storage of Location Information
Request and Reply For Location Information From WHOIS
Locator Module

Figure 7 - Locator Module

The Locator Module will make use of the WHOIS Protocol [6] specified in RFC 3912 as found at www.ietf.org. This protocol provides for the ability to look up location and contact information for servers and IP addresses connected to the World Wide Web. Using the server or addressing information provided by IRC clients at logon such as *(~njvc@1CCED6DD.204D9C25.F48CD9E7.IP)*, the Lookup Module will attempt to retrieve location information either on the user or on the service provider of the user depending on what information is given. Lookups for this module will be sent using libraries such as the Asp.NET Whois lookup or the PHP Script library for Whois. Due to reliance on login information, the Locator Module may not necessarily return usable information every time it is implemented. As explained in the WHOIS specification sections 1, 4, and 6, there are certain shortcomings in the protocol being used in which case, some of the offerings by certain types of chat rooms as discussed in the following subsections may prove useful. However, if this module is successful for even a portion of the total requests it makes, it is reducing the man-hours necessary for the investigation and thereby accomplishing the task for which it was designed.

6.4.1 Paid-Account Rooms

A paid-account chat room could be likened to those owned by ISPs such as AOL or those owned by internet dating sites. These rooms are inhabited by users who pay to use the services offered by the room owners. In rooms such as this, no Lookup Module would be necessary to locate the offending user. Owners of these rooms require their users to pay in order to access the services, which mean the user would typically need to have a credit card and a valid mailing address on file with the chat room provider. When criminal activity is detected in a paid-account chat room, the investigating officers will need to take the evidence before a judge and obtain a subpoena for the account information on the chat room owner’s servers. When this type of room is under investigation, a priority alert should be sent so the investigating agent may begin the process to obtain a subpoena for the location information of the user. Although this applies to the Analysis and Alert Modules, it warrants mentioning here due to the relevance to user location.

6.4.2 Static-Account Rooms

A static-account chat room would be similar to a room operated by AOL for their AIM users or members using Yahoo! chat rooms. These users do not pay to use this service, however, when they create an account or username, it belongs to them until such time as it is removed or they give up their claim. These providers will not necessarily require users to provide location information, but may record location information on logon. If a room without user location information is being investigated with this tool there are two possible courses of action. The first course would be to use the Lookup Module in accompaniment with any information being provided by the room upon login. The other possibility would be to do an IP traceback on the user’s messages as discussed in “Network Support for IP Traceback [3],” by Stefan Savage. The complete text is cited for later reference, however the summary of the solution is:

> *In this paper, we describe a general purpose traceback mechanism based on probabilistic packet marking in the network. Our approach allows a victim to identify the network path(s) traversed by attack traffic without requiring interactive operational support from Internet Service Providers (ISPs). Moreover, this traceback can be performed “post mortem”—after an attack has completed.*

The caveat to this solution would be the need for an overhaul of networking infrastructure in order to facilitate the proposal by Mr. Savage.

6.4.3 Dynamic-Account Rooms

Dynamic-account rooms are of the same nature as those IRC rooms being discussed elsewhere in this document. These are rooms where a user does not necessarily have to use the same username or nickname every time they log on. This provides a challenge not only attempting to locate the user’s physical location, but also to link the user to other conversations in which they have been involved. Luckily, IRC has a feature that will assist the tool in this area. This assistance comes in the way of the server message when a user joins a room. As shown before, the message displayed below not only displays the user’s nickname, but their username and potential server address as well.

Nava (~njvc@1CCED6DD.204D9C25.F48CD9E7.IP)

This aids pursuit by the way of offering the details the Lookup Module would need in order to attempt to locate this person. With rooms lacking this feature, another route of discovery would be the aforementioned traceback solution implemented at the time of the messaging in order to assure it is still the same user sending the messages.

By making use of some convenient built-in features of the selected chat room protocol and with the added utility of the WHOIS protocol, the Lookup Module has the potential to reduce the amount of time officers would usually be required to use in order to manually do the same tasks. The apparent difficulty presented in tracing the location of a user when there is insufficient server or account information would not be changed regardless of an automated approach being implemented in the search. Using a module such as the one described here would be a benefit to law enforcement not only with respect to the timeliness of investigations, but also financially in the way of reducing man-hours spent investigating the location of a chat room user.

7 Initial Implementation

This thesis is proposing a general tool not targeted at any specific investigation group. The remainder of this paragraph and list suggests and incremental design which borrows from the rapid deployment model within the Extreme Programming paradigm. So, in order to implement the DAIT, the design team must approach this in __ tasks in order for the design and implementation to go smoothly and to gain the largest benefit from an incremental implementation. These steps are as follows:

1. Additional Research
2. Database Schema Design
3. Collection, Parsing, and Storage
4. Keyword Analysis
5. Alerting Investigators
6. Locating Users
7. Return to Research

7.1 Additional Research

As will be mentioned again in Section 9.3, some additional research will need to be conducted before implementation of the DAIT may begin. This research will primarily come in the form of requirements solicitation from the investigation agency for which the specific implementation is being created. But, before interviewing and planning a design, first and foremost the design team must study precise handling requirements for digital evidence so as not to corrupt possible evidence during collection. These requirements will be a very important aspect of how the DAIT implementation will need to treat incoming data.

The research with the investigating agency will contribute heavily to later designs. The design teams needs to solicit information on particular key words, phrases, or regular expressions that matter most to the investigating organization. These will be used during analysis in order to flag suspect messages, so the criteria must be selected which will provide the most useful information to the organization

The final segment of research necessary is gathering information and requirements on investigator responsibilities and information necessary for alerting them. This information will be necessary to include for the Alert Module implementation so as to alert the correct investigator for the specific message type being flagged.

7.2 Database Schema Design

A database schema will need to be designed to hold the large quantities of data which will be captured by this system. In Section 6.2 a possible schema for storage of message components was suggested and in Appendix F another schema was demonstrated which also included the relationships between alerts and investigators. These schemas will need to be refined for the precise requirements of the agency for which the DAIT implementation is being designed. Additional expansions to the schema are as follows:

- Additional information for key words, phrases, and regular expression search criteria. If stored in the database, this allows for changing search criteria without having to modify the Analysis Module.
- Additional information message classification. If stored in the database, this allows for changing the parsing algorithm without having to modify the Collection Module.

- Alert and analysis storage after the Analysis Module reviews the incoming messages.
- Due to differing opinions from different groups, additional information will need to be stored as requested by the agency for which the specific DAIT implementation is being designed.

Once this information has been added to the schema for the storage module, the data will need to be normalized so as to reduce redundancy in tables. Redundancy offers potential for inconsistency in the database unless additional measures are taken which will create an additional workload on the designers. Once the schema is ready to be used, a suitable DBMS will have to be chosen. This decision will be specific to the needs of the agency and suitable research must be completed in order to choose a DBMS which meets the requirements of the design and the contracting organization.

The final consideration needed for design of the Storage Module is permission sets to be applied to users accessing the database. These permission sets, sometimes referred to as views, are used in order to restrict access to tables within the database. For the purpose of this design, views will be used in order to provide a layer of security to the data stored within the tables. One consideration for this should be the table holding evidence. This table should only allow the DAIT to insert data and other to ready the data. No one should have the ability to update or delete (except for archival purposes) already existing information. Other tables should have similar security measures implemented based on the business rules of the particular organization.

7.3 Collection, Parsing, and Storage

Collection and parsing of information is covered under the umbrella of the Collection Module and should be launched at the same time as the Storage Module because collecting information does no good without a means to keep the data, so the database designed in the previous section must be implemented along side the Collection Module discussed here.

For collection of IRC chat room information, the protocol in RFC 1459 was discussed in earlier sections. Implementing this protocol in an object-oriented language such as Java or C++ would provide an already library of functions to use for internet connection. By not using an already implemented chat room client the team is free to create a tool integrating all aspects of the Collection Module rather than having to link out of the box tools.

The other consideration for the Collection Module is parsing the data based on the message classifications stored in the Storage Module. By implementing a program which parses based on a set on rules stored in the database, the Collection Module will have a very dynamic operation. Now, when classification patterns are discovered which need to be implemented in the parsing algorithm, the Collection Module can be updated without recompiling the entire tool. The completion of this portion will provide data in a smaller and more usable format for post-mortem analysis. By launching this portion of the tool immediately, time spent analyzing data would be reduced, already contributing to the final goal of the DAIT.

7.4 Keyword Analysis

The Analysis Module, responsible for detecting suspect messages, should be the next portion of the tool implemented. This portion would further reduce the load on investigators by automatically analyzing the data in real time rather than investigators being forced to collect data, then analyze it manually post-mortem. This tool would be implemented to run automatically just before a new message is introduced for storage. Using the key word, phrase, and regular expression rules gathered from the investigation organization, the analyzer could check the message for matches to these rules and then add an alert record to the database with the message if a rule match is found within the message. Again, by keeping these rules in the database, the keyword analyzer could have a dynamic behavior without the need to recompile the tool every time a change is made.

7.5 Alerting Investigators

The next logical step for implementation of the DAIT would be the Alert Module. After the messages are collected, parsed, stored, and analyzed, the investigating agents would need to refer to the stored alert data to check on the status of new messages. This task would be repetitive and time consuming. By implementing the Alert Module next, the agents would be advised as to when there is a suspect message waiting for them, and therefore not unnecessarily check the database for information.

The Alert Module, by associating particular crimes with particular investigators, will reduce the amount of unnecessary seen my investigating agents and thereby reducing the amount of time spent looking at data thus freeing time which could be used to pursue other responsibilities.

7.6 Locating Users

The final portion of the tool to implement on the initial launch will be the Locator Module. Upon implementation of this module successful searches will reduce the amount of time agents spend attempting to determine the physical location of a suspect.

This portion of the implementation would involve selection of a code library appropriate to use with the code from other modules and selection of a WHOIS server or set of servers from which to attempts to retrieve data. This module comes with a caveat however; the information used for WHOIS can be spoofed or faked by a malicious user. So, the Locator Module may return bad results. But, for every good result returned, the investigator is saved the time required to locate a user.

7.7 Return to Research

Once the DAIT is implemented the work does not stop. In order to achieve maximum return on potential of the system, there are upgrades and ongoing data maintenance which needs to be addressed. These items will be noted in Section 9.3 for tasks to complete upon implementation of the DAIT system. The benefit of performing these tasks will be a tool which will continue improving performance based on the needs of the organization for which it was implemented, thereby reducing the time investment needed by agents and freeing up time which could be used to investigate other crimes.

8 Summary

8.1 Benefits of Automation

The tool described within this thesis would, in-essence, be a ad hoc group of many tools working in unison to achieve the common goal of detecting, alerting law enforcement to, and locating malicious users in internet chat rooms. The particular benefits afforded by this design are:

- **Automation** – Investigators would no longer be required to look at chat text waiting for an indicator of a crime.
- **Individual Module Benefit** – Each module of the DAIT would individually be responsible for removing a portion of an investigator's workload.
- **Incremental Implementation** – The agency implementing this tool would be able to benefit from the completion of each stage of the tool rather than having to wait for the completion of the entire system before beginning use.
- **Reallocation of Resources** – Investigators previously tasked with the job of reviewing chat logs or monitoring chat rooms would now be free to undertake new assignments which offer the potential to increase productivity of the agency using the DAIT.

This combination of tools would lighten the load of already fully-tasked investigators of law enforcement agencies by doing the same job that would currently take a person or persons working around the clock all year to accomplish. This would reduce the work-load by 8766 man-hours invested in looking at chat room text waiting for something to happen. By reducing the review load on investigators and broadening the searchable area by running multiple instances, this tool set would allow for more efficient detection and location of suspects in chat room crimes.

8.2 Conclusion

An architecture design for an automated IRC investigation tool has been presented. Using this design, a working implementation of the tool can be put into operation which will benefit investigating agencies currently using a manual or post-mortem approach to their investigations. This tool will reduce the number of man-hours invested in monitoring activity within IRC chat rooms or waiting for text searches of large chat logs to complete. By using the incremental implementation approach suggested, the agencies will begin receiving the benefits of the tool even before completion of the final product, thus slowly reducing investigator workload so as to begin integrating other responsibilities such as offline investigations.

8.3 Future Work

Tasks before implementation:

- Research proper digital evidence handling techniques and practices
- Interview agency for key word, phrase, and regular expression requirements
- Collect agent information requirements
- Collect data analysis requirements
- Revise database schema to include new information
- Choose a design lifecycle which supports incremental implementation such as the Extreme Programming example in Appendix H.
- Begin Documentation
    - Software Requirements Specification

  - Software Design Specification
  - Software Test Plan

Tasks after implementation:

- Method of differentiating between human users and bots. Bots are small automated programs with one of their primary uses being identity theft [11, 12].
- Further research on message classification using large data sets using research by Elnahrawy [10] as a guide. Then implement as the second half of the Analysis Module if the benefit is found to have an appropriate cost to benefit relationship.
- Review performances based on current implementation and modify search criteria accordingly.

As the pre-implementation tasks are completed, these will guide the developer toward a product which will most benefit the investigation agency for which the DAIT is being implemented. These tasks are presented in the order which best suits the order of implementation so as to allow incremental versions of the tool to be built while requirements elicitation is ongoing.

The post-implementation tasks are three tasks which could be used to enhance overall performance of the DAIT system after the initial launch. The first task, determining if a user is a bot or human, will prove useful for determining if the users suspected of committing crimes in chat rooms are doing so manually or have bot-nets implemented for a more hands-off approach.

The second task, data reclassification, will open up new avenues of investigation through demonstrating new ways to look at messages. These reclassifications may begin to offer research areas to improve not only the investigation benefits, but routs through which to enhance human knowledge as well.

The third and final task, performance reviews, should be done on a regular basis to be determined by the investigation agency and performance of the DAIT. When reviewing the search criteria, considerations should be made to those criteria which appear to be generating excessive work through flagging of false-positive results. The investigation agency should consider removing criteria causing this issue and potentially introduce new criteria if common threads are found in correctly alerted records. Through These three tasks the overall performance of the DAIT could increase the number of detections made while reducing workload-increasing false-positives.

References

[1] US v. Charbonneau, 979 F.Supp. 1177 (S.D. Ohio 1997)

[2] NH judge throws out paedo chat-log evidence. 12 Apr, 2004. The Register. <http://www.theregister.co.uk/2004/04/12/judge_paedo_chatlog/>

[3] Savage, Stefan. "Network Support for IP Traceback." IEEE/ACM Transactions on Networking. June 2001: p226-237.

[4] Breeden, Bob. "Investigating 'Internet Crimes Against Children' (ICAC) Cases in the State of Florida." Proceedings of the 2006 ACM Symposium on Applied Computing SAC '06. April 2006

[5] Bill of Rights Transcript. The National Archives. <http://www.archives.gov/national-archives-experience/charters/bill_of_rights_transcript.html>

[6] RFC 3912 WHOIS Protocol Specification. 8 April, 2007. Internet Engineering Task Force. <http://tools.ietf.org/html/rfc3912>

[7] RFC 1459 IRC Protocol Specification. 8 April, 2007. IRChelp. <http://www.irchelp.org/irchelp/rfc/rfc.html>

[8] Dawson, Ed. "Slang of the Crimeware Hackers." Australian PC Authority. 12 March 2007

[9] Meehan, A. Manes, G. Davis, L. Hale, J. Shenoi, S. "Packet Sniffing for Automated Chat Room Monitoring and Evidence Preservation." Proceedings of the 2001 IEEE Workshop on Information Assurance and Security. June 2001.

[10] Elnahrawy, Eiman M. "Log-Based Chat Room Monitoring Using Text Categorization: A Comparative Study."

[11] Levy, Elias. Arce, Iván. "A Short Visit to the Bot Zoo." IEEE Security & Privacy. 2005.

[12] http://honeynets.org/papers/bots/

[13] Magid, Larry. "Help Children Know the Risks of Chat Rooms." PCAnswer.com. April 2007.

[14] Rules Prof. Conduct (1995): Rule 1-400. Advertising and Solicitation. 8 April, 2007. The State Bar of California. < http://www.calbar.org/pub250/9/s0009-a.htm>

[15] grep – GNU Project – Free Software Foundation (FSF). 12 April, 2007. The GNU Project. < http://www.gnu.org/software/grep/>

[16] Seay, Gregory. "Web Unit Gets Tough." Hartford Courant. 15 April, 2007.

[17] Locke, Mandy. "Volunteer jailbait helps snare Internet predators." The News & Observer. 9 February, 2007.

[18] Extreme Programming: A Gentle Introduction. 20 April, 2007.
Don Wells. < http://www.extremeprogramming.org/index.html>

Appendices

Appendix A – Laws

Fourth Amendment

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

California Rule of Professional Conduct 1-400

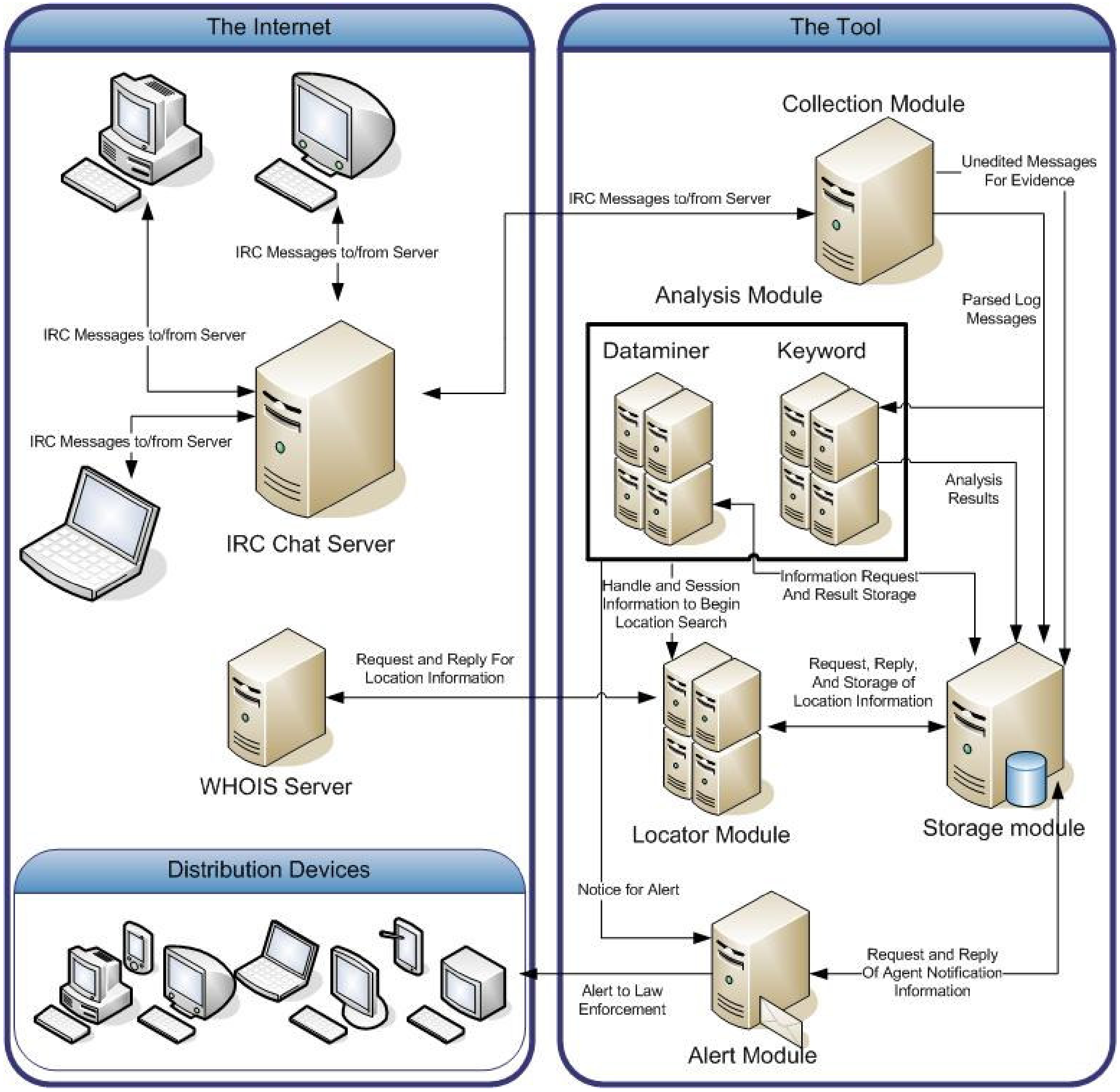
A solicitation shall not be made by or on behalf of a member or law firm to a prospective client with whom the member or law firm has no family or prior professional relationship, unless the solicitation is protected from abridgment by the Constitution of the United States or by the Constitution of the State of California. A solicitation to a former or present client in the discharge of a member's or law firm's professional duties is not prohibited.

For full text see: http://www.calbar.org/pub250/9/s0009-a.htm

US Code – Title 18, Part 1, Chapter 119, §2511

Interception and disclosure of wire, oral, or electronic communications prohibited. For full text see: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=browse_usc&docid=Cite:+18USC2511

Appendix B – Detailed Module Diagram



The Internet
The Tool
Collection Module
Unedited Messages For Evidence
IRC Messages to/from Server
IRC Messages to/from Server
IRC Messages to/from Server
IRC Messages to/from Server
Analysis Module
Parsed Log Messages
Dataminer
Keyword
IRC Chat Server
Analysis Results
Information Request And Result Storage
Handle and Session Information to Begin Location Search
Request and Reply For Location Information
Request, Reply, And Storage of Location Information
WHOIS Server
Locator Module
Storage module
Distribution Devices
Notice for Alert
Request and Reply Of Agent Notification Information
Alert to Law Enforcement
Alert Module

Figure 8 - Detailed Module Diagram

Appendix C – Capture Sample

****Excerpts from logs captured from the #ccpower channel on the irc.hackcrew.cc server.****
****Actual credut card information has been masked to protect the victims.****

```
**** BEGIN LOGGING AT Wed May 10 11:13:20 2006

May 10 11:13:20 * Now talking on #ccpower
May 10 11:13:20 * Topic for #ccpower is: Â«Your Ultimate Internet
Security GuideÂ» | Forum: http://www.hackcrew.cc  | RULES: @/+v verify
first, | For cashout join #Cashout |  Report/View rippers #rippers -/-
For Help and Chat Channel join #hackcrew -/- For WesternUnion Join
#WesternUnion Notice: Only English !
May 10 11:13:20 * Topic for #ccpower set by FY5HT1 at Thu May  4
16:15:32 2006
May 10 11:13:22 * ^|MUST|KILL|^
(~__MUST_KI@FDDA220A.B45C5666.C2C615BD.IP) has left #ccpower (Client
parted)
May 10 11:13:25 * cosminel has quit (hackcrew.cc)
May 10 11:14:51 _dEAN   n0Mb3r
May 10 11:14:53 _dEAN   u are cashier?
May 10 11:14:55 * AlbaHack gives channel operator status to Brainstorm
May 10 11:14:56 * nowdolegro (~gabriel@E2C3E814.3649F9F6.A451C01D.IP)
has joined #ccpower
May 10 11:15:05 * AlbaHack sets mode +a #ccpower Brainstorm
May 10 11:15:07 * Bel| (~bell@470B4855.4331DFC0.85D3F3C9.IP) has joined
#ccpower
May 10 11:15:13 * AlbaHack sets mode +q #ccpower Brainstorm
May 10 11:15:18 * AlbaHack slaps Brainstorm around a bit with a large
trout
May 10 11:15:22 * _dEAN slaps n0Mb3r around a bit with a large trout
May 10 11:15:24 _dEAN   + me
May 10 11:15:24 _dEAN   :)
May 10 11:15:25 * ^|MUST|KILL|^
(~__MUST_KI@FDDA220A.B45C5666.C2C615BD.IP) has joined #ccpower
May 10 11:15:26 * Geohacker (~asscb2005@F5D7F41.1086CC46.EEA7CD36.IP)
has joined #ccpower
May 10 11:15:27 * Brainstorm slaps AlbaHack around a bit with a large
trout
May 10 11:15:28 * n0Mb3r gives voice to _dEAN
May 10 11:15:30 |SounD|  ===========
May 10 11:15:31 |SounD|  Cardname: ***** *******
May 10 11:15:31 |SounD|  Address1: *** ****** **
May 10 11:15:31 |SounD|  Address2:
May 10 11:15:31 |SounD|  City: Tylertown
May 10 11:15:31 |SounD|  State: MS
May 10 11:15:31 |SounD|  Zip: 39667
May 10 11:15:31 |SounD|  Phone: **********
May 10 11:15:31 |SounD|  CC Number: ****************
May 10 11:15:31 |SounD|  CC Month: 12
May 10 11:15:31 |SounD|  CC year : 2007
May 10 11:15:32 |SounD|  Cvv2: ***
May 10 11:15:33 |SounD|  Bank Account Number: *******
May 10 11:15:34 |SounD|  Bank Routing Number: *********
May 10 11:15:35 |SounD|  MMN: **********
May 10 11:15:36 |SounD|  SSN: ***-**-****
May 10 11:15:37 |SounD|  DOB: */*/**
May 10 11:15:38 |SounD|  ===========
May 10 11:15:46 * Geohacker slaps AlbaHack around a bit with a large
trout
May 10 11:15:50 * Geohacker slaps AlbaHack around a bit with a large
trout
May 10 11:15:56 * Brainstorm sets ban on
*!*@FDDA220A.B45C5666.C2C615BD.IP
May 10 11:16:05 |SounD| ?
May 10 11:16:06 * Brainstorm has kicked ^|MUST|KILL|^ from #ccpower
(dont kick the boss)
May 10 11:18:43 * AlbaHack gives channel operator status to Geohacker
May 10 11:18:45 AlbaHack      Bitch :P
May 10 11:18:46 AlbaHack      hahahaha
May 10 11:19:22 Geohacker     Geohacker Confirm WU in 10 Minutes
May 10 11:19:27 Geohacker     MTCN: **********
May 10 11:19:28 Geohacker     Amount: $823.56
May 10 11:19:28 Geohacker     Fee:         $49.44
May 10 11:19:28 Geohacker     Total:     $873.00
May 10 11:19:28 Geohacker     Test Question: Pet's name?
May 10 11:19:28 Geohacker     Answer: *****
May 10 11:19:28 Geohacker     Sender: ******* ****
May 10 11:19:32 |SounD| albahack
May 10 11:19:35 |SounD| albahack
May 10 11:19:36 |SounD| albahack
May 10 11:19:38 AlbaHack      ?
May 10 11:19:48 Security      Anyone got stable Shell with Tcl /msg
AlbaHack !
May 10 11:19:48 Security      Anyone got stable Shell with Tcl /msg
AlbaHack !
May 10 11:19:59 Geohacker     Todays COnfirmed Result
May 10 11:20:00 |SounD| chk pvt
May 10 11:20:02 |SounD| chk pvt
May 10 11:20:35 ^|MUST|KILL|^ _ACTION selling fresh maillist
(Yugoslavia-Yemen-Vietnam-Venezuela-Vanuatu-Uzbekistan-Uruguay-United
Kingdom-United Arab Emirates-Ukraine-Uganda-Turks Caicos Islands-
Turkmenistan-Turkey-Tunisia-Trinidad and Tobago-Tonga-Togo-Thailand-
Tanzania-Taiwan-Syria-Switzerland-Sweden-Swaziland-Sudan-St. Lucia-St
Kitts Nevis-Sri Lanka-Spain-South Africa-Slovenia-Slovak-Singapore-
Seychelles-Saudi Arabia-Salvador-Russian Federation-Romania-H
May 10 11:21:16 |SounD|  User Account
May 10 11:21:17 |SounD|  -------------------
May 10 11:21:17 |SounD|  Account open in : **
May 10 11:21:17 |SounD|  Online ID : *********
May 10 11:21:17 |SounD|  Passcode : ******
May 10 11:21:17 |SounD|  ATM PIN : ****
May 10 11:21:17 |SounD|  Email : ********@*******.com
May 10 11:21:17 |SounD|  Billing Address
May 10 11:21:17 |SounD|  Cardholder name : ******** *******
May 10 11:21:17 |SounD|  Address1 : ******** **
May 10 11:21:17 |SounD|  Address2 :
May 10 11:21:18 |SounD|  City : New York
May 10 11:21:19 |SounD|  State : NY
May 10 11:21:20 |SounD|  Zip : 22111
May 10 11:21:21 |SounD|  Phone : *** *** ********
May 10 11:21:22 |SounD|  Credit Or Debit Information
May 10 11:21:22 |SounD|  ------------------------------
May 10 11:21:24 |SounD|  Credit Card Number : ****************
May 10 11:21:25 |SounD|  Exp Date : 07-2007
May 10 11:21:26 |SounD|  C V N : ***
May 10 11:21:27 |SounD|  Bank Account Number : *******
May 10 11:21:28 |SounD|  Bank Routing Number : ********
May 10 11:21:29 |SounD|  Security Question
May 10 11:22:39 |SounD|  -------------------
May 10 11:22:40 |SounD|  Account open in : AR
May 10 11:22:40 |SounD|  Online ID : **********
May 10 11:22:40 * AlbaPoweR (~KHG@6D1872F6.9B77F742.6A7D5018.IP) has
joined #ccpower
May 10 11:22:40 |SounD|  Passcode : *******
May 10 11:22:40 * Security gives channel operator status to AlbaPoweR
May 10 11:22:43 |SounD|  ATM PIN : ****
May 10 11:22:44 |SounD|  Email : *********@excite.com
May 10 11:22:45 |SounD|  Billing Address Cardholder name :
May 10 11:22:46 |SounD|  Address1 : *** ****** ** ***
May 10 11:22:47 |SounD|  Address2 :
May 10 11:22:48 |SounD|  City : pine bluff
May 10 11:22:49 |SounD|  State : AR
May 10 11:22:50 |SounD|  Zip : 71601
May 10 11:22:50 |SounD|  Phone : ***-***-****
May 10 11:22:53 |SounD|  Credit Or Debit Information
May 10 11:22:53 |SounD|  Credit Card Number : ****************
May 10 11:22:54 |SounD|  Exp Date : 02-2010
May 10 11:22:55 |SounD|  C V N : ***
May 10 11:22:56 |SounD|  Bank Account Number : ************
May 10 11:22:57 |SounD|  Bank Routing Number : *********
May 10 11:22:58 |SounD|  Security Question
May 10 11:22:59 |SounD|  Mother Maiden Name : ******
May 10 11:23:00 |SounD|  S S N : *********
May 10 11:23:00 |SounD|  D O B : **/**/1981
May 10 11:23:02 |SounD|  Driver License Number : *********
May 10 11:23:02 |SounD| □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
May 10 11:23:06 |SounD| > User Account
May 10 11:23:06 |SounD|  -------------------
May 10 11:23:06 |SounD|  Account open in :
May 10 11:23:07 |SounD|  Online ID : *********
May 10 11:23:07 * \-- has quit (hackcrew.cc)
May 10 11:23:08 |SounD|  Passcode : *******
May 10 11:23:09 |SounD|  ATM PIN : ****
May 10 11:23:10 |SounD|  Email : *******@aol.com
May 10 11:23:11 |SounD|  Billing Address
May 10 11:23:12 |SounD|  -------------------
May 10 11:23:13 |SounD|  Cardholder name : **** *** *********
May 10 11:23:14 |SounD|  Address2 :
May 10 11:23:15 |SounD|  City : Lutz
May 10 11:23:16 |SounD|  State : FL
May 10 11:23:17 |SounD|  Zip : 33559
May 10 11:23:18 |SounD|  Phone : ***-***-****
May 10 11:23:19 |SounD|  Credit Or Debit Information
May 10 11:23:20 |SounD|  ------------------------------
```

Appendix D –Preliminary IRC Message Classification

Room Participant Changes
HANDLE (~*USERNAME@SERVER*) has joined *CHANNEL*
HANDLE (*USERNAME@SERVER*) has joined *CHANNEL*
HANDLE (~*USERNAME@SERVER*) has left *CHANNEL* (*REASON*)

Mode Changes
Modes {a, q, m, i, n, p, C}
HANDLE sets mode *[+/-]MODE CHANNEL*
HANDLE sets mode *[+/-]MODE CHANNEL HANDLE*
HANDLE sets ban on *HANDLE/WILDCARD@SERVER*

Moderator Give Commands
HANDLE gives channel operator status to *HANDLE*
HANDLE gives voice to *HANDLE*

Exit Messages
HANDLE has quit (*MESSAGE)*
HANDLE has kicked *HANDLE* from *CHANNEL* (*MESSAGE*)

Topic and Entry Messages
Now talking on *CHANNEL*
Topic for *CHANNEL* is: *TOPIC*
Topic for *CHANNEL* set by *HANDLE* at *DATE/TIME*
HANDLE has changed the topic to: *TOPIC*

Moderator Remove Commands
HANDLE removes ban on *HANDLE/WILDCARD@SERVER*
HANDLE removes voice from *HANDLE*
HANDLE removes channel operator status from *HANDLE*
HANDLE removes user limit

User Changes Handle
HANDLE is now known as *HANDLE*

Unclassified Message Types
HANDLE slaps *HANDLE* around a bit with a large trout
HANDLE Back!
HANDLE MESSAGE
Disconnected (*MESSAGE*).

Appendix E – Collection Module Pseudo Code Algorithm

```
Connect to IRC server and channel using protocol specified in RFC 1459
Connect to database of choice
Store session information
Create list of connected users
While connected to channel
    Read incoming message
    If administrative message
        Set LogonID to denote a nonuser message
    Else if message from user
        If nickname is not in user list
            Search database for username /nickname combo
            If user exists in database
                Retrieve information to add to user list
                Store user’s ID
            Else
                Add user’s information to database
                Retrieve new ID
                Store user’s information in list
            End if
        Else
            Get user’s ID from list
        End if
    End if
    Add message to messages table and keep message ID
    Send message to Keyword Analysis Engine
    Search for hyperlink in message text
    If link exists in message
        If link exists in links table of database
            Increment count attribute
            Get the link ID
        Else
            Add link to links table in database and retrieve new ID
        End if
        Add link ID and message ID to the database
    End if
    If administrative message
```

```
        Switch on (message classification)
            Case: user sign on/off
                Retrieve nickname, username, and server
                If user does not exist in list
                    Add Handles entry
                End if
                Add new record in LogonInfo
                Add new record in LogTime
            Case: exit message
                Find user in users list
                Add new entry in LogTime
            Case: user changes handle
                Find and update info in user list
                If new information does not exist in Handles table
                    Add new entry in Handles table
                End if
                Add new entry in LogTime for previous handle exit
                Add new entry in LogTime for new handle
            Case: interactions
                Set Actor as first word in message
                Find second handle in message
                Find action
                If action does not exists in Actions table
                    Add action record
                End if
                Increment action count
                Add record to Interactions table
        End switch
    End if
End while on disconnect from server or signal from user
Close database connection
```

Appendix F – Database Schema

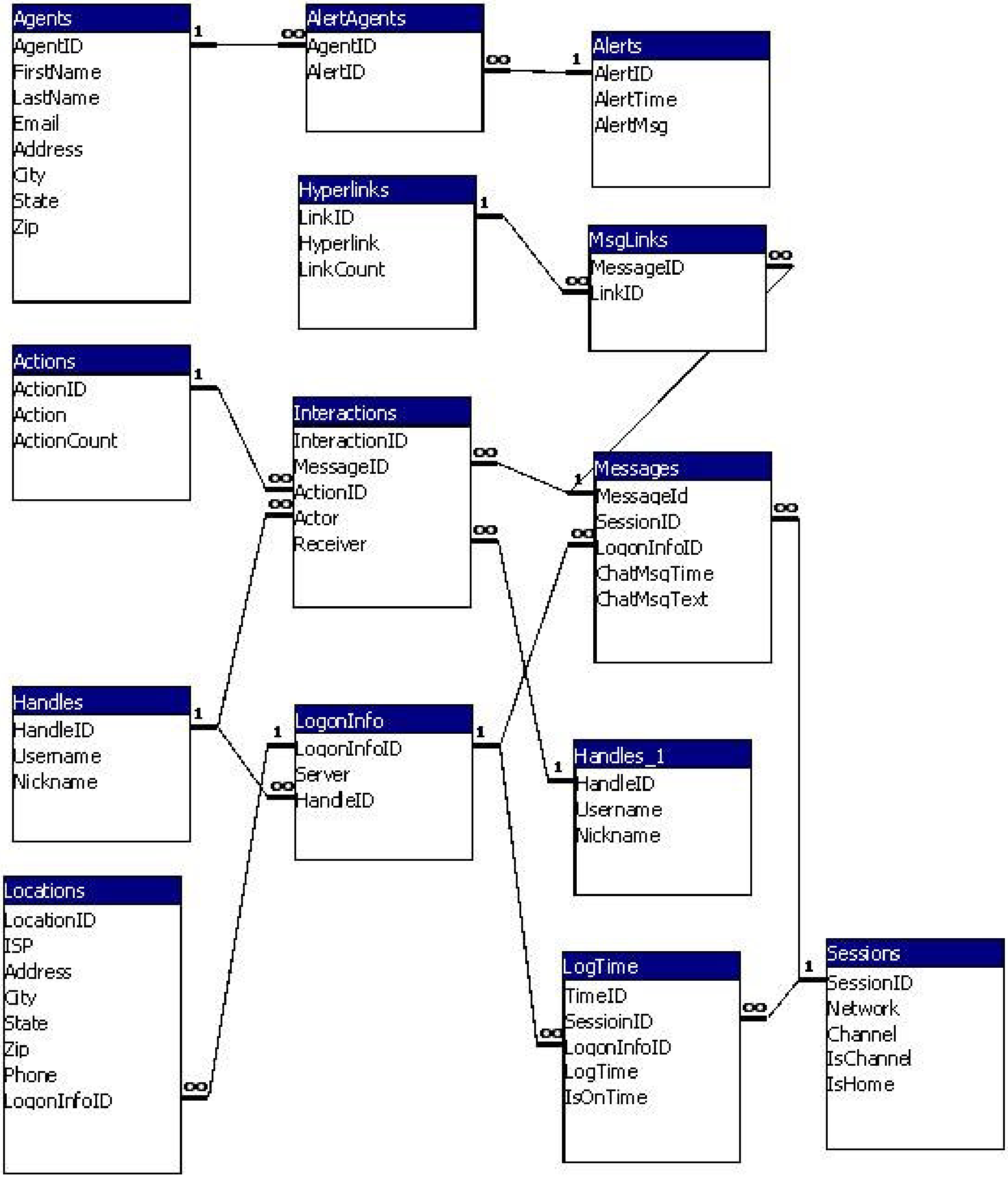


Figure 9 - Detailed Example Database Schema

Appendix G – Database Schema Description

DATA DICTIONARY

Session

*SessionID	Unique number to identify the logging session
Network	Network the message occurred on
Channel	Channel or IM the message occurred on
IsChannel	True for a channel message and false for private message
Is Home	True if the channel is the landing page for the server

Handles

*HID	Unique number to identify handles records
Nickname	The nickname a user goes by
Username	The name provided by each user at logon

LogonInfo

*LogonInfoID	Unique number to identify a user in the database
HandleID	Unique identifier of a record in the handles table
Server	Gathered by network and provided on channel logon

LogTime

*TimeID	Unique identifier associated with a user's logon/logoff
SessionID	Identifier for a related record in the session table
LogonInfoID	Identifier for a related record in the LogonInfo table
LogTime	Time the user joined or left the channel
IsOnTime	True if the user joined the channel

Message

*MessageID	Unique identifier for each message
SessionID	Identifier for a related record in the Session table
LogonInfoID	Identifier for a related record in the LogonInfo table
ChatMsgTime	Date/Time that a message was sent
ChatMsgText	Text sent with each message

HyperLinks

*LinkID	Unique identifier for links in chat
HyperlinkLink	Link text identified in message
LinkCount	Number of times the link has appeared in logging

MsgLinks

*MessageID	Message in which the link was found
*LinkID	Hyperlink text found within the message

Actions

*ActionID	Unique Identifier for each action
Action	The action recorded between two users
ActionCount	Number of times the action has appeared in logging

Interactions

*InteractionID	Unique identifier to observer actions between users
ActionID	Action text found within the message
MessageID	Message in which the action was found
Actor	HandleID of the person initiating the action
Receiver	HandleID of the person receiving the action

Locations

*LocationID	Unique identifier for each lookup
ISP	ISP of the user
Address	Physical street address
City	City of user
State	State of user
Zip	Zip Code of user
Phone	Phone number of user
LogonInfoID	Identifier to link location to a particular user

Alerts

*AlertID	Unique identifier for each alert
AlertTime	Date and time the alert was sent
AlertMsg	Message of the alert

Agents

*AgentID	Unique identifier for each agent
FirstName	First name of agent
LastName	Last name of agent
Email	Email address of agent
Address	Physical street address
City	City of user
State	State of user
Zip	Zip Code of user

AlertAgents

*AgentID	Identifier of agent to whom the alert was sent
*AlertID	Identifier of the alert sent to this agent

* - Single or set of attributes that are unique to the table. (Primary Key)

Appendix H – Extreme Programming

This software design methodology is an ideal approach for incremental implementations of a large system where software requirements are likely to change as the development progresses. The map illustrated in Figure 13 on this page may be found on http://www.extremeprogramming.org [18].

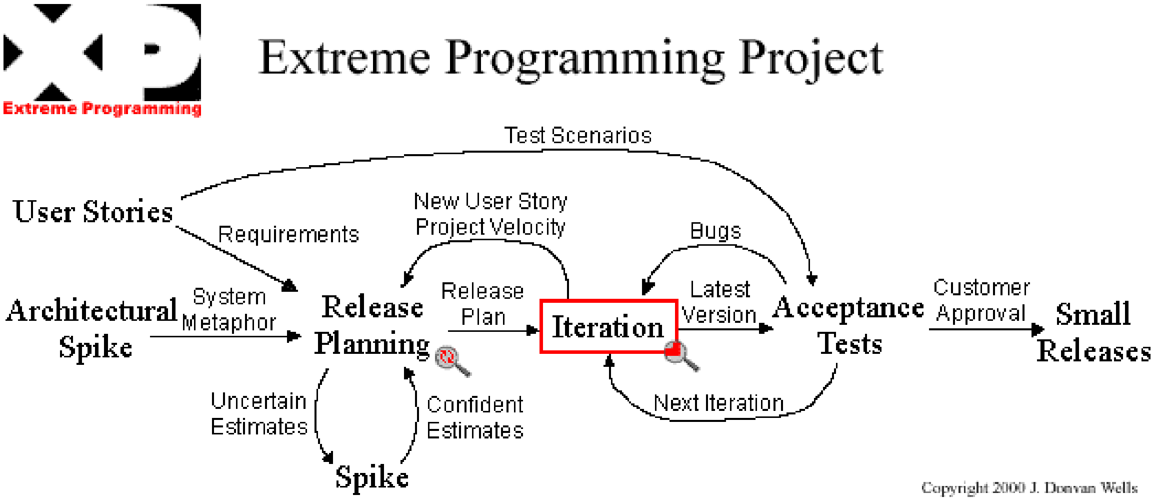


Figure 10 - Extreme Programming Model

Appendix IExample Data through the Tool

The diagrams represented in this section demonstrate three different messages potentially received by the DAIT. This represents the text of the message only, no user or date information is to be implied. These represent three cases:

1. A credit card number in the message
2. A user joining the channel
3. A general message

Message with Credit Card Number

The message: “Credit card: 1234-2345-3456-4567”

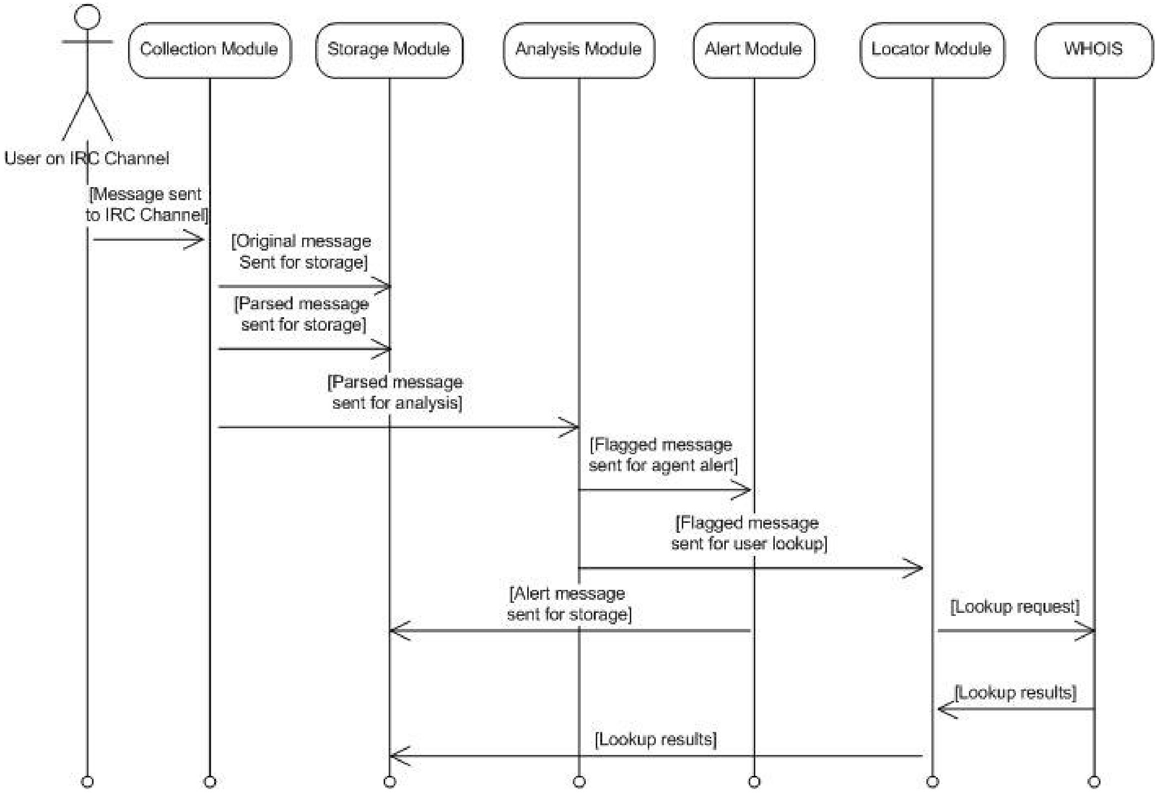

User on IRC Channel
Collection Module
Storage Module
Analysis Module
Alert Module
Locator Module
WHOIS
[Message sent to IRC Channel]
[Original message Sent for storage]
[Parsed message sent for storage]
[Parsed message sent for analysis]
[Flagged message sent for agent alert]
[Flagged message sent for user lookup]
[Alert message sent for storage]
[Lookup request]
[Lookup results]
[Lookup results]

Figure 11 - Use Case with Credit Card Number

Message with User Joining the Channel

The message: “User1 (user@123.234.345.456.IP) has joined channel”

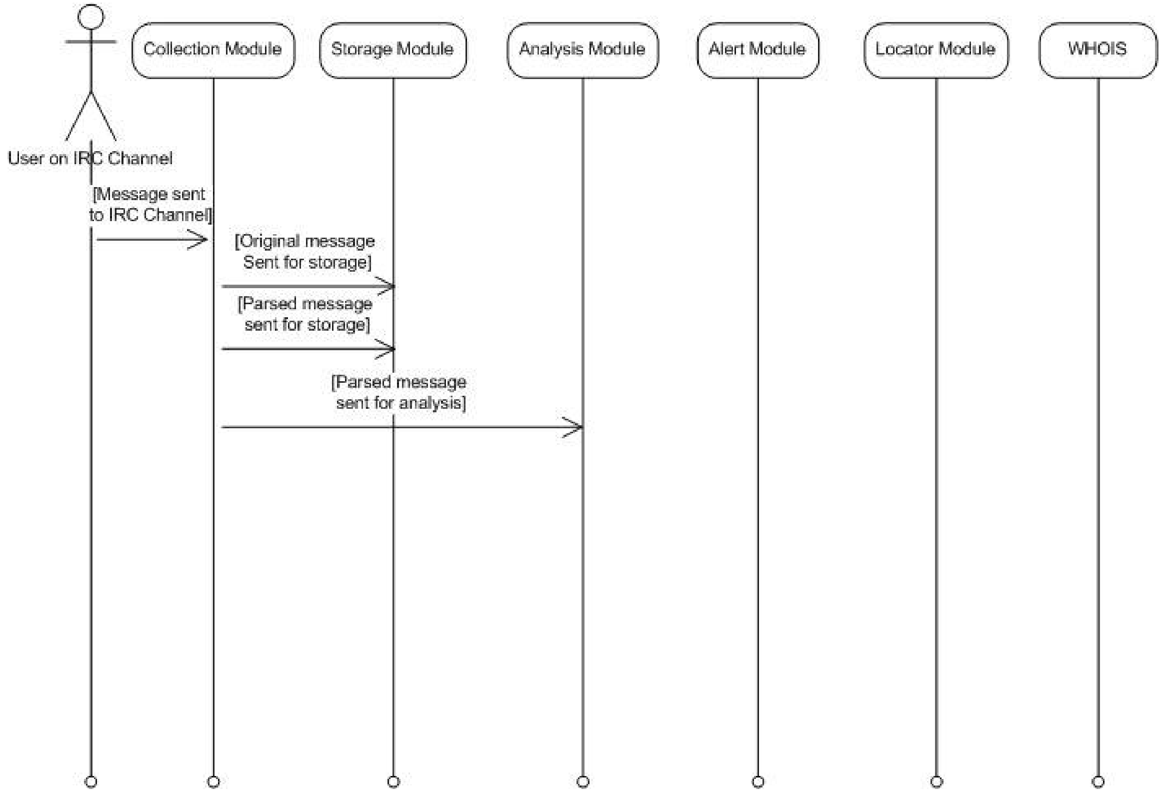
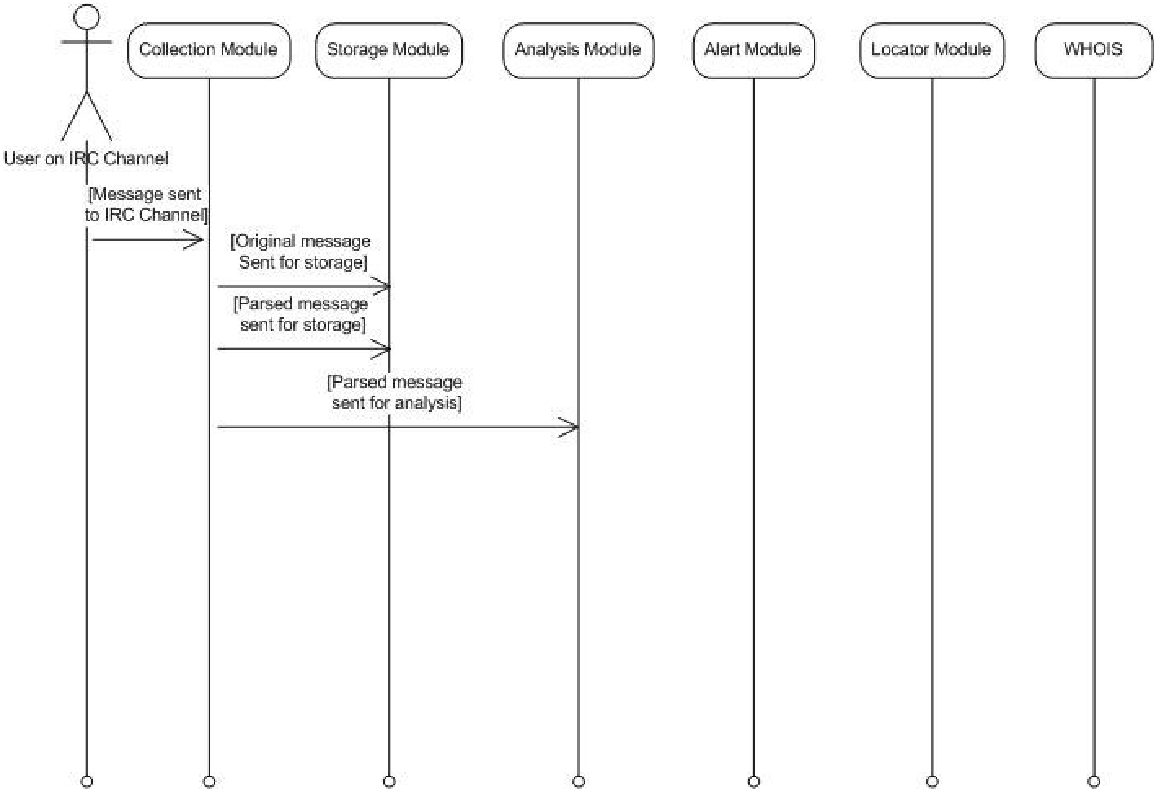

User on IRC Channel
Collection Module
Storage Module
Analysis Module
Alert Module
Locator Module
WHOIS
[Message sent to IRC Channel]
[Original message Sent for storage]
[Parsed message sent for storage]
[Parsed message sent for analysis]

Figure 12 - Use Case with Login Message

Message with General Text

The message: "Hello room. How is everyone today?"



Collection Module
Storage Module
Analysis Module
Alert Module
Locator Module
WHOIS
User on IRC Channel
[Message sent to IRC Channel]
[Original message Sent for storage]
[Parsed message sent for storage]
[Parsed message sent for analysis]

Figure 13 - Use Case with General Text Message