Internet Security Threat Data Collection and Analysis using the *Philippine Honeynet Project* Infrastructure

Mark Ryan del Moral Talabis
The Ateneo Information Technology Institute
130 H.V.Dela Costa Street, Salcedo Village
Makati City, Philippines 1227
talabis@gmail.com
ryan@philippinehoneynet.org

ABSTRACT

This study discusses the development of an internet security threat data collection and analysis mechanism using a Gen III honeynet. This honeynet infrastructure is used to gather and collect first-hand data (direct from the internet) on security threats such as worms, hacking attempts and other internet anomalies.

The data, collected over a period of three months, is studied, presented and analyzed using a graphical chart-based analysis approach, utilizing custom tools and techniques developed in the course of the study. The given analysis period produced over thirty individual case-based studies that led to the discovery of new worm signatures, malware variants, and an analysis of other hacking activities. The study also includes a quarter's worth of long-term historical analysis of collected data.

Keywords

Networks, Internet Technologies, IT Security

1. INTRODUCTION

A honeynet is a research tool consisting of a network specifically designed for the purpose of being compromised, with control mechanisms that prevent this network from being used as a base for launching attacks against other networks. Once compromised, the honeynet can be used to observe the intruders' activities, collect tools and determine new trends in network attacks.

A honeynet is nothing more than an architecture. Basically, it is a virtual fishbowl for observing hacker activity. Just like a fishbowl, an environment is created wherein a researcher can watch everything happening inside it. Instead of sand and coral, a researcher can add different operating systems, databases or applications.

Thus, a honeynet deals with gathering information related to security threats. This information has different value to different people in different settings, depending on what a researcher is trying to achieve. The data collected can be utilized for reverse engineering of a new worm or malware variant, analyzing security trends and early warning systems or even profiling hacker groups and techniques.

2. BASIC CONCEPTS: HONEYPOTS & HONEYNET TECHNOLOGIES

2.1 Honeypots

The concept of a honeypot was introduced in 1991. However, until recently there has been no clear and widely accepted definition. Professionals in the field of IT security define a honeypot as tool for law enforcement. Some think of it as an NIDS while others think of it as a deception device. Recently, the following definition was put forward by the Honeypot mailing list, a list consisting of about 5000 different security professionals who are currently involved in honeypot research.

A Honeypot is a security resource whose value is in being probed, attacked or compromised.[1]

This definition is widely regarded as the de facto definition of a honeypot and will be used for the duration of the study.

Honeypots can be divided into two general categories: The low-interaction honeypot and the high-interaction honeypot. Honeynets are a form of high-interaction honeypots.

2.2 Honeynets

The concept of the honeynet first began in 1999 when Mr. Lance Spitzner, founder of the Honeynet Project published the paper "To Build a Honeypot" [2]. In this paper, Mr. Spitzner proposed that instead of developing technology that emulated systems to be attacked, why not deploy real systems behind firewalls waiting to be hacked.

In the most basic sense, a honeynet is a type of honeypot, more specifically, a type of high interaction honeypot. And thus being a high interaction honeypot, nothing is emulated; all services, applications and operating systems are as real as in any production environment. An important characteristic that separates a high interaction honeypot from a honeynet is that a honeynet contains one or more honeypots. It is a network of multiple systems creating an illusion of a production network. It is through this network, specifically through the network access device, is where hacker activity is monitored, recorded and controlled. Based on all of this, we can construct the basic definition of a honeynet:

A honeynet is a network of high interaction honeypots that simulates a production network and configured such that all activity is monitored, recorded and in a degree, discretely regulated. [3]

A honeynet, just like honeypots, works by creating a highly controlled environment. Honeynets as opposed to honeypots though takes the concept one step further. Instead of just one computer or a number of unconnected computers, a network is set up in such a way that everything in the honeynet appears like a normal network. All applications and services are real though all systems running within the honeynet are considered honeypots. No modifications are done to the system such as placing monitoring tools or creating jailed environments like chroot within the host. This kind of setup makes the honeynet the most interactive and authentic of all honeypots.

The key to honeynet architecture is the gateway. The gateway is basically the network access control device that isolates your honeypots from the network. Unlike honeypots in which the actual capture and control mechanism is in the honeypots themselves, a honeynet gateway is the one that capture, controls and collects all inbound and outbound data.

The honeynet gateway provides 2 critical requirements, which are Data Control and Data Capture [4].

Data Control is the containment of activity. The primary purpose of this requirement is the risk mitigation. Risk mitigation requires that all activities should be confined within the honeynet.

The second requirement of honeynets is Data Capture. Data Capture is the monitoring and logging of attacker activities within the honeynet. These activities are what form the basis of the data used in research and analysis. For a more complete data capture and to better piece together activities of the attacker, it is necessary to have multiple mechanisms for capturing these activities. These could be in form of tcpdump logs, IDS logs, Sebek data and firewall logs among others. This is also important so that a failure in one of these mechanisms would still allow you to collect one form of data or another to prevent a total blackout of activity data.

These requirements are important in any honeynet implementation of which there are a number of types based on how they implement the said requirements. The types of honeynets can be summarized into (1) Gen I Honeynets; (2) Gen II Honeynets; (3) Gen III Honeynets; (4) Distributed Honeynets and (5) Virtual Honeynets.

Of the different types, the Gen III honeynet architecture is the basis of this studies deployment. Gen III honeynets introduced a single device that handles the data control and data capture mechanisms of the honeynet called the IDS Gateway.

The changes introduced by the IDS Gateway in data control made Gen II and Gen III honeynets more difficult for attackers to detect or fingerprint.

By making the architecture stealthier, attackers are kept longer and thus more data is captured. There was also a major thrust in improving honeypot layer of data capture with the introduction of a new UNIX and windows based data capture tool called Sebek.

2. DATA COLLECTION MECHANISM AND INFRASTRUCTURE

The data collection infrastructure deployed in this study is a Virtual Gen III Honeynet deployment (see Figure 1). This deployment makes use a Gen III architecture that resides in a single computer using virtualization technology.

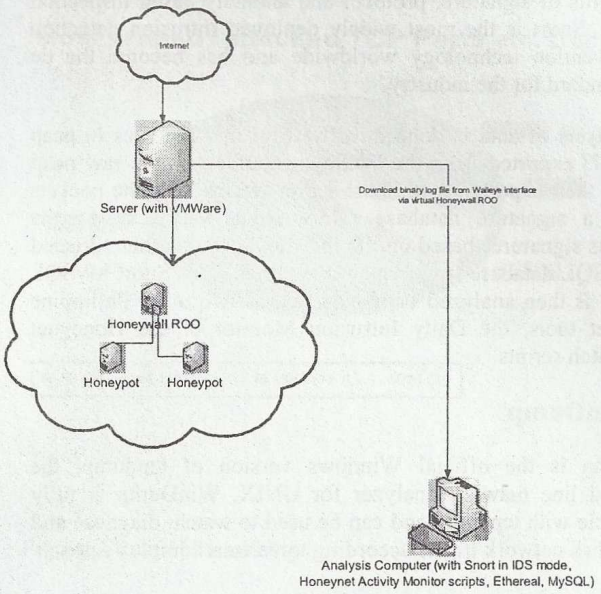

Figure 1 Virtual Gen III Honeynet Deployment

This deployment was made possible by VMWare [5]. VMWare is virtualization software that can emulate most forms of Windows and Linux operating systems. Thus with this software, it is possible to deploy as many operating systems in a single machine. Each deployment will act as if it is a separate computer with separate hardware, software and network identification. This allows for whole honeynets, as long as the server can handle the load, to reside in a single server.

Binary logs are downloaded via a Walleye [6] interface by connecting to the virtual Honeywall server which acts exactly the same as a physical Honeywall. The logs are then processed in an Analysis Computer using the various honeynet analysis tools.

3. ANALYSIS TOOLS AND TECHNIQUES

3.1 Honeywall ROO CDROM

The core tool used by the Philippine Honeynet Project is the Honeywall CDROM Roo. The CDROM Roo is a production honeynet solution developed and maintained by the Honeynet Project Research Alliance. The Roo CDROM contained the core Gen II data control and data capture functionality, but now has remote GUI administration, data analysis integration, Sebek 3.x support, a robust OS base, automated updating, and much more.

The Honeywall ROO installs a minimized version of the Fedora 3 Core Operating System into the hard drive of the honeynet machine. The system is minimized for security reasons (such as no windowing capabilities) but left enough base OS for additional functionality (such as webserver, database, and international keyboard support).

3.2 Snort and Snort-Inline

Snort is an open source network intrusion prevention and detection system utilizing a rule-driven language, which combines the benefits of signature, protocol and anomaly based inspection methods. Snort is the most widely deployed intrusion detection and prevention technology worldwide and has become the de facto standard for the industry.

The analysis of data is done directly from raw log files in pcap format [7] exported from the Walleye interface. These raw pcap files are then exported to a Snort server, which runs the packets through a signature database. The output, which shows the intrusions signatures based on the log files given, is then directed to a MySQL database from which it is mined. This Snort MySQL database is then analyzed or is exported to two other Philippine Honeynet tools, the Daily Intrusion Monitor or the Honeynet Trendwatch scripts.

3.3 WinDump

WinDump is the official Windows version of tcpdump, the command line network analyzer for UNIX. WinDump is fully compatible with tcpdump and can be used to watch, diagnose and save to disk network traffic according to various complex rules.

3.4 Ethereal

Ethereal is a network packet analyzer. A network packet analyzer captures network packets and displays packet data as detailed as possible. One could think of a network packet analyzer as a measuring device used to examine what's going on inside a network cable. Network packet analyzers are used for troubleshooting network problems, debug protocol implementations, learn network protocol internals, and in the case of this study, examine security problems.

3.5 Honeynet Activity Monitor

The Honeynet Activity Monitor is the primary data analysis tool used by the study. This was developed by the Philippine Honeynet Project in the process of this study. The Honeynet Activity Monitor is a set of scripts that generates graphs and a spreadsheet showing all attacks and anomalies extracted from a Snort MySQL database. The scripts are usually run on a daily basis to give the analysts and users a detailed account of what kind of activities happened within the day.

The Honeynet Activity Monitor is composed of 6 scripts that analyzes a different aspect of the data and generates graphs for the data analyzed. These scripts are written in PHP and works in both Windows and Linux as long as the proper PHP scripting engine is installed. The scripts mine data from a Snort MySQL database.

3.6 Honeynet Trendwatch

As with the Honeynet Activity Monitor, the Honeynet Trendwatch is a set of scripts that generates graphs that shows all attacks, anomalies and other relevant data. Only this time, it is geared towards a longer period of time. While the Honeynet Activity Monitor focuses on daily data, the Honeynet Trendwatch focuses more on long-term data. It was initially patterned after the concept of DShield [8] in that it tracks historical data. The data used by the scripts are taken from the data generated by the daily Honeynet Activity Monitor that is exported to a separate Honeynet Trendwatch MySQL historical database. The scripts are also written in PHP and can be run in both Windows and Linux. There are as of this moment, 13 scripts for the Honeynet Trendwatch.

3.8 The Analysis Process

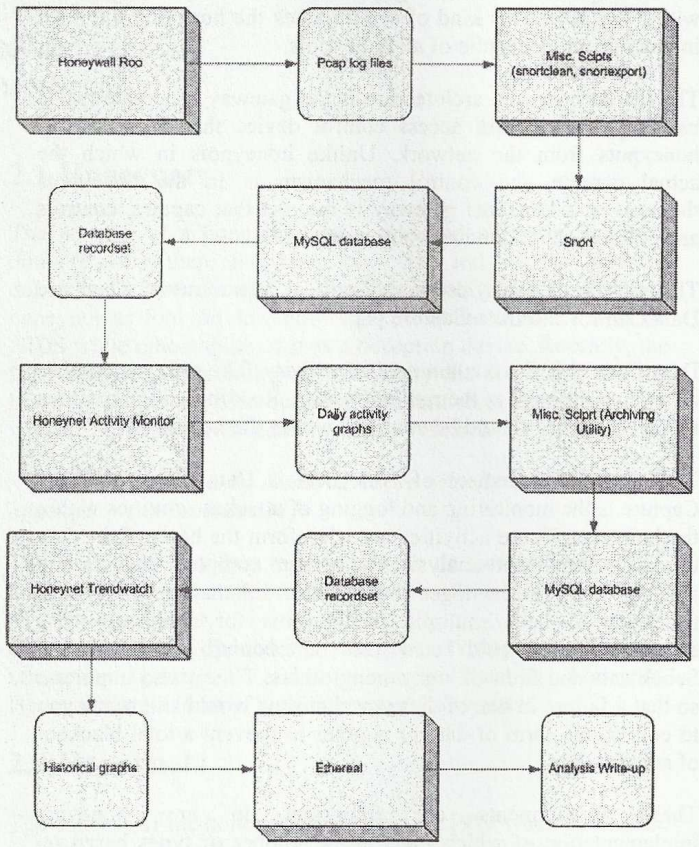

Figure 2 Analysis Process diagram

This diagram (see Figure 2) describes how the study uses the tools stated above that leads to a case analysis. Here is a summary for the process:

1. log files in pcap format at extracted from the Honeywall Roo Walleye interface. This is usually a daily process.
2. The miscellaneous scripts, snortclean.bat and snortexport.bat are used to prepare the Snort database and export the raw log files into Snort
3. Snort then processes the raw pcap files and stores the results in a MySQL database.
4. The database records produced by Snort are then used by the Honeynet Activity Monitor to generate the daily activity graphs.
5. The archiving utility script is used to process the daily data and store it in a historical MySQL database.
6. The records from this historical database are then used by the Honeynet Trendwatch to produce the historical charts.

7. The actual packets are examined using ethereal with the help of the charts produced by the Honeynet Activity Monitor and the Honeynet Trendwatch.

4. RESULTS

4.1 Cases

In the course of the study, more that thirty individual cases were analyzed using the data gathered through the Philippine Honeynet Project Infrastructure. For the sake of readability, the cases will be presented in a summarized form. To read the full cases, the reader is advised to visit the Philippine Honeynet Project website [9] to read the case archive.

What follows are a list of choice cases representative of the entire analysis case. The write-ups that are presented here are the actual unedited analysis sent out as advisories to the security community.

4.1.1 Discovery of new security threats

The analysis that follows became the basis of the discovery of a new worm variant named "Dasher". The study, by using the Philippine Honeynet Infrastructure was the first to report the worm activity. The activity was noticed by researchers due to a sudden increase in activity of one of the honeynet ports. The activity was coming from IPs originating from Korea. The actual packets collected by the honeynet were sent to SANS for further analysis. SANS confirmed that this attack is the signature of a new worm. The Philippine Honeynet Project was referenced by a number of foreign news sources as one of the first organizations to report the new worm activity.

December 13, 2005

Port 1026

Honeynet activity has been bottoming out (Figure 3) for the past few weeks. Activity consists of the usual Welchia / Nachi traffic [31] which seemed to have picked up after the recent bout of awstats / xmlrpc.php attacks [10] that have been very active lately.

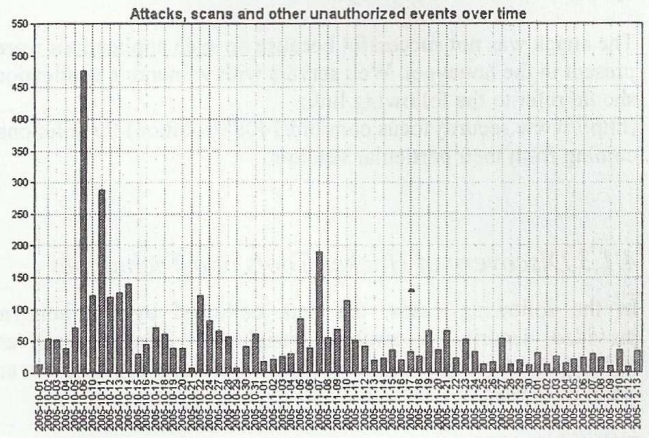

Figure 3 Attacks, scans and other unauthorized events over time 2005-12-13

Aside from this, some peculiar activity during the day was noted, which showed up as shellcode activity directed towards port 1025 (Figure 4) which showed up as having a sudden spike in activity. Source IP seems to indicate a Korean point of origin.

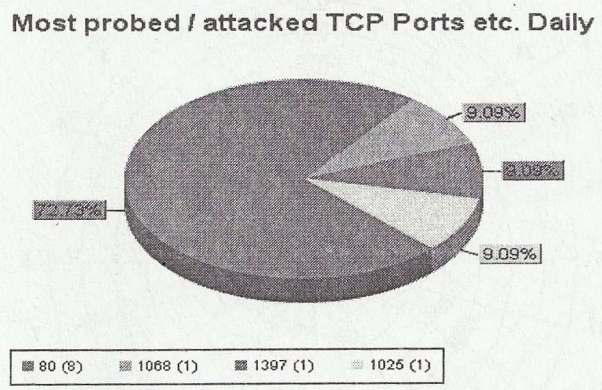

Figure 4 Port 1025 activity

Port 1025 is commonly used by the Microsoft Remote Procedure Call (RPC) service. These scans are most likely RPC and LSA exploit attempts [11] against Windows. In this particular case, the packets captured seem to point to an LSA attack via TCP port 1025.

4.1.2 Detection of new web application attacks

The next analysis reflects the value of the honeynet data collection infrastructure in intrusion detection research. This is particularly true in the case in the detection and analysis of web application attacks which based on our long term analysis, are the one of the most predominant security activity. The honeynet collected multiple variants of web-based attacks directed to popular open source applications like Awstats, XMLRPC for PHP, Coppermine and PhpBB. The analysis that follows is an advisory released to the security community about attacks directed towards Awstats and XMLRPC for PHP

November 11, 2005

Awstats.pl access & configdir command execution attempt and xmlrpc.php

For the past few days, the Philippine honeynet has been receiving a number of attacks comprising of the following snort signatures: (1) WEB-CGI awstats access; (2) WEB-CGI awstats.pl configdir command execution attempts; (3) WEB-PHP xmlrpc.php post attempt and (4) WEB ATTACKS wget command attempt

The signatures form a signature pattern illustrated by one of the attacks (see Figure 5) caught by the honeynet.

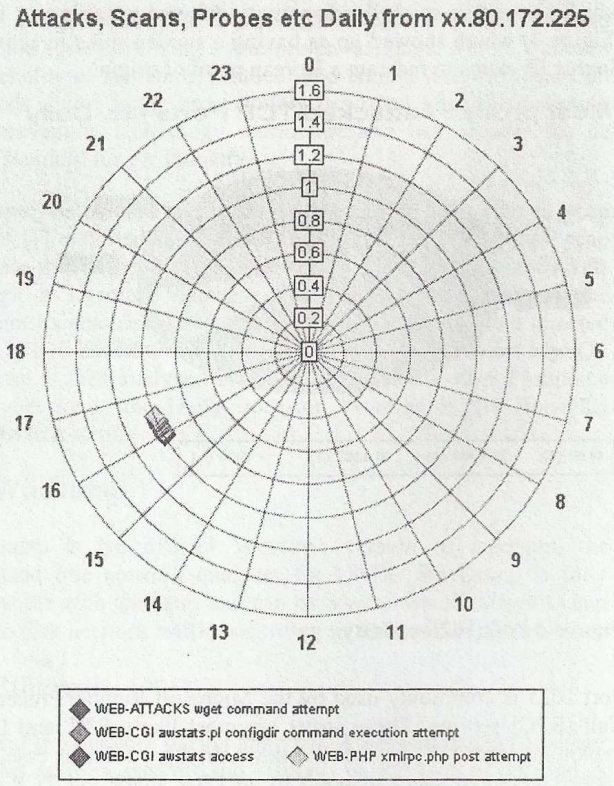

Figure 5 Awstats and xmlrpc.php signature pattern

The actual packets involved in the attack can be viewed in the Philippine Honeynet Project case archive [12]. It can be summarized by the following events:

1. Attacker loops through possible Awstats directories (/awstats/, /cgi-bin/, /cgi-bin/awstats/) where awstats.pl might be found

2. Attacker sends the following URL *"awstats.pl?configdir=|echo;echo YYY;cd /tmp;wget 24.224.174.18/listen;chmod +x listen;./listen 216.102.212.115;echo YYY;echo|"*. This URL tries to exploit the configdir input validation [33] vulnerability that could allow remote command execution

3. If vulnerability is present, the following commands will execute: *Download a file called "listen". Set file permission. Execute the "listen" program.* (most likely the file "listen" is a backdoor program)

4. Attacker proceeds to use the xml-rpc vulnerability to apply the same exploit [13]. A different xml-rpc attack was caught by the Philippine honeynet just recently and can be referenced in our November 5 analysis [14].

The attacks were not successful in the honeynet since Awstats and xml-rpc were not present. Most likely this is a worm or a canned exploit. Administrators should check the appropriate vendors for the necessary patches.

November 5, 2005

xmlrpc.php post attempt

The honeynet captured a peculiar web application activity signature this day. Snort reported it as a WEB-PHP XMLRPC.php post attempt. This is a PHP remote code injection vulnerability caused by an input validation error in XML-RPC for PHP [15]. This vulnerability allows an attacker to execute arbitrary commands or code in the context of the Web server.

The actual packets, captured by the honeynet, of an exploit based on this vulnerability can be referenced in the Philippine Honeynet Project website case archive [14]. Based on the captured packets, the characteristics of the attack are

1. Attack sends post to xmlrpc.php script
2. Changes directory "cd /tmp"
3. Uses wget to download a file "wget 217.160.255.44/cback"
4. Changes the file permission using "chmod +x cback"
5. Runs the file "./cback 202.101.165.61 8080" (note: cback is a trojan)
6. Exits

The attack tries to find the XMLRPC.php file through trial and error using a list of predefined locations based on the applications (e.g. xoops, word press and phpgroupware, etc.) that uses XML-RPC for PHP. Some of the directories attacked were:

```
/xmlrpc/xmlrpc.php
/xmlsrv/xmlrpc.php
/blog/xmlrpc.php
/drupal/xmlrpc.php
/wordpress/xmlrpc.php
/phpgroupware/xmlrpc.php
```

The attack was not successful because no such applications were present in the honeynet. Web servers with vulnerable applications should refer to the following link: (http://www.securityfocus.com/bid/14088/solution) for solutions coming from their particular vendors.

4.1.3 Discovery of Hacker Tools and Techniques

In the following cases, the data collected by the honeynet provided security researchers insights to hacker tools and techniques. Among insights gained were discovery of tools and techniques used in FTP warez and spamming activities.

October 1, 2005

Grimm's Ping

Attempted recon and scanning attempts were the predominant honeynet activity. As usual, Welchia worm activity was present as it has been for some weeks now. To prevent compromise from this worm which had kept on rebooting the system, we decided to patch our Win2k server.

There were two notable priority 1 events. One was an attempted login at around 2100 to our FTP server and an attempted login to our SMTP server at around 1600. After review of the logs, the FTP login though unsuccessful seem to indicate the use of a tool called "Grim's Ping" [16] indicated by the login "Qgpuser@home com" (17).

Grimm's Ping is a software used to scan the network to find vulnerable FTP sites. It is commonly used in warez activity. Further study to determine the exact signature of Grim's ping is ongoing.

December 2, 2005

FTP Probes

Some peculiar FTP related activity (see Figure 6) was noted this day. Based on our November 14 analysis [18], these are FTP probes are prerequisite activities for warez related events. In most cases, these probes are usually looking for anonymous servers to be used for warez storage.

Among the signatures notable for these kinds of activities are (1) POLICY FTP 'CWD' possible warez site; and (2) POLICY FTP anonymous login

The usual pattern in FTP probing goes a little something like (1) Attacker connects to FTP server; (2) Attacker logs in as anonymous and sends bogus password; (3) Attacker does a trial and error search for FTP common directories; and (4) If a directory is found, it creates multiple nested directories

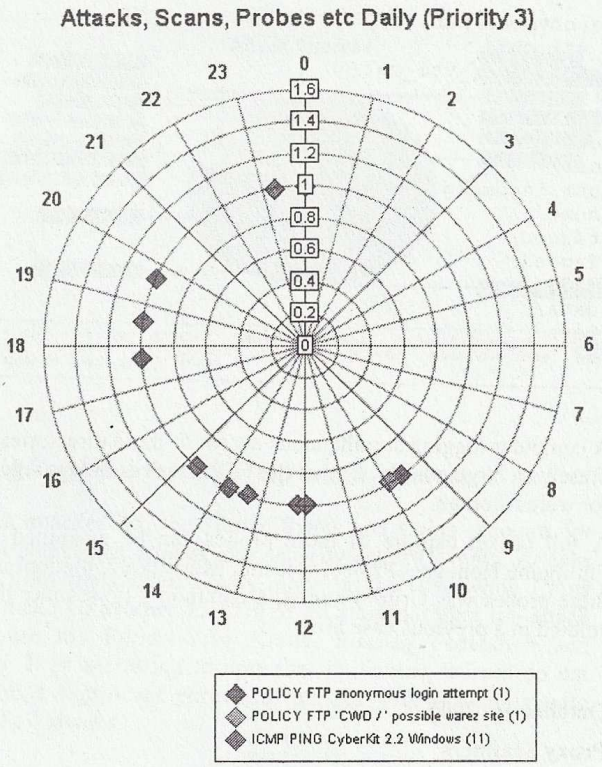

Figure 6 An FTP probe attempts

The usual pattern in FTP probing goes like:

1. Attacker connects to FTP server
2. Attacker logs in as anonymous and sends bogus password
3. Attacker does a trial and error search for FTP common directories
4. If a directory is found, it creates multiple nested directories

The common directories searched by the attacker the following:

```
/pub/
/public/incoming/
/incoming/
/upload/
/_vti_pvt/
/_vti_txt/
/_vti_log/
/wwwroot/
/anonymous/
/public/
/
/outgoing/
/temp/
/tmp/
/anonymous/_vti_pvt/
/anonymous/incoming/
/mailroot/
/ftproot/
/anonymous/pub/
```

```
/anonymous/public/
/_vti_cnf/
/anonymous/_vti_cnf/
/images/
/_private/
/cgi-bin/
/usr/
/usr/incoming/
/home/
/tagged/
/Tagged/
/TaGGeD/
/data/
/Data/
/%/
/  /
```

A computer logging simultaneous access to these directories, presents a large possibility that the FTP server is being targeted for warez storage.

A full packet capture of these probes can be examined in the Philippine Honeynet Project website. Most likely, the tool used in these probes was Grimm's ping, a common FTP scanner that we studied in a previous case [16].

October 10, 2005

Proxy Scanners

A peculiar signature was noted for the past few days. This was the "FTP command overflow attempt" The said signatures formed a pattern which was made up consistently of itself plus TCP port scan and an open port response. Since this rule was made for 3CDaemon server which we don't have, this was either a false positive or some other "attack". Thus we decided to investigate the said pattern further.

The investigation led to a couple of peculiar packets (see Philippine Honeynet Project) which seems to be using the honeynet to relay or "proxy" towards an AOL IP and ebay. This would seem to indicate proxy scanning activity where tools are used to identify and check out proxies for use in spamming or other such related activities.

The attempts were unsuccessful in the honeynet but any administrators noticing these patterns in their perimeter should already be on guard and check if their systems are vulnerable to proxying activities to prevent their systems to be used as such.

December 12, 2005

SPAM / PHISHING from "support@microsoft.com"?

For the past few months, we have been receiving sporadic attempts to relay messages from an obviously bogus "support@microsoft.com", account. This activity was first detected last October in one of our October 19 analysis [19].

The attempts seem to be coming from the CHINANET Guangdong province network in Shenzhen, Guangdong, China though this might possibly be just another relay. Further investigation of the other attempts for the past months will most likely uncover whether this is the original source. The actual packets can be viewed in the Philippine Honeynet Project case archives [19].

4.1.4 Observation of actual break-ins

Aside from hacking attempts and attacks, data from the Philippine Honeynet Project infrastructure also allows security researchers to observe, study, and analyze actual compromises or "breakdowns" which are ordinarily risky or not even possible in production systems. What follows is an actual compromise of one of the honeypots in this study.

October 7, 2005

WINS exploit compromises honeypot

Honeynet activity was very heavy on this day. Several priority 1 and priority 3 events were worth further investigation. The events that were particularly interesting were the WINS and SHELLCODE activity (see Figure 7).

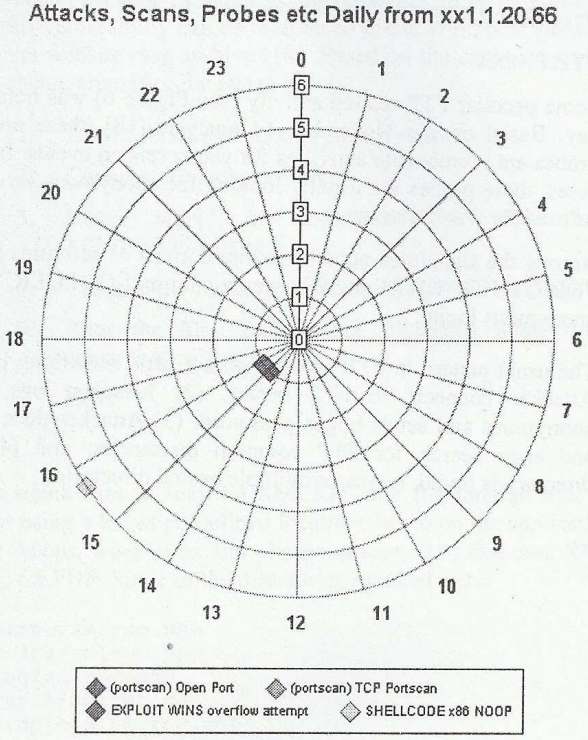

Figure 7 WINS and Shellcode activity

By itself SHELLCODE NOOP event has a large chance of being a false positive but based on the intrusion charts below, it is accompanied by other events such as multiple port scans and a WINS buffer overflow attempt coming from a single source which indicates a high probability of being a real attack. From this we can deduce that the attacker was trying to access a command level interface through a WINS vulnerability [20]. So now having a bit more direction, we were now ready to drill down to the actual packets.

Based on the charts, we extracted the actual packets [21] from the attack. As inferred form the packets, the attacker was successful in his exploit. Our initial findings indicate that the attacker is a worm / bot or an automated tool in which the very short time interval between commands was a clue. The attacker opened an ftp connection provided a username and password and then proceeded to download "date.exe" which most likely is a malware of sorts.

4.2 Long Term Data

The data collection infrastructure implemented in this study also makes it possible for security researchers to analyze security events over a period of time. In this study, data collected over a period of three months was used to report observations about the pervading internet security situation.

4.2.1 Time series average of attacks and unauthorized events

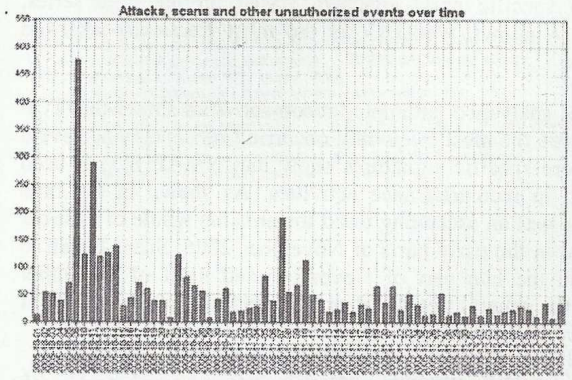

Figure 8 Time series average of attacks and unauthorized events

For the last quarter of 2005 (October to December 2005), security events averaged 55.6 intrusions per day during the period of data collection. These events we're heavily concentrated on the early part of October (Oct 5 to Oct 14) and on the early part of November (Nov 5 and Nov 10). The attacks had begun to drop off in December.

4.2.2 Total distribution of attacks in days of the week

Attacks and other security related events occurred most frequently on Thursdays (Philippine Time) with 25.39% of all attacks and security related events occurred. This was followed by Monday, which totaled 18.71%. Attacks and security related events occur the least frequent on Sundays with 7.95%.

4.2.3 Total distribution of attacks in hours of the day

Attacks and other security related events occurred most frequently between 4 to 7 PM (Philippine Time) with each totaling 13.81% and 13.25% respectively. The next most frequently attacked time period occurs at around 4 AM with 12.88% of all attacks occurring within the said period.

4.2.4 Total distribution of top country sources

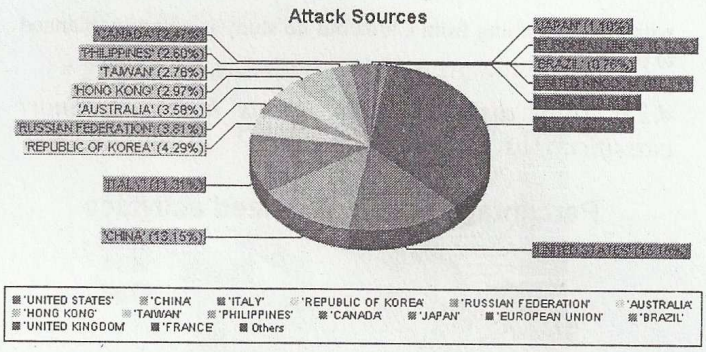

Figure 18 Total distribution of top country sources

Based on attacker IPs, the United States and China accounts for the most sources of attacks and security related events. Attacks coming from the United States make up 15.18% while China sources make up another 13.15% of all attacks. Other top attack sources are the Republic of Korea, Russian Federation and Australia. It is interesting to note that Philippine sources do not amount to a significant portion of attacks with it totaling only 2.80% of all attacks.

4.2.5 Time series distribution of country sources

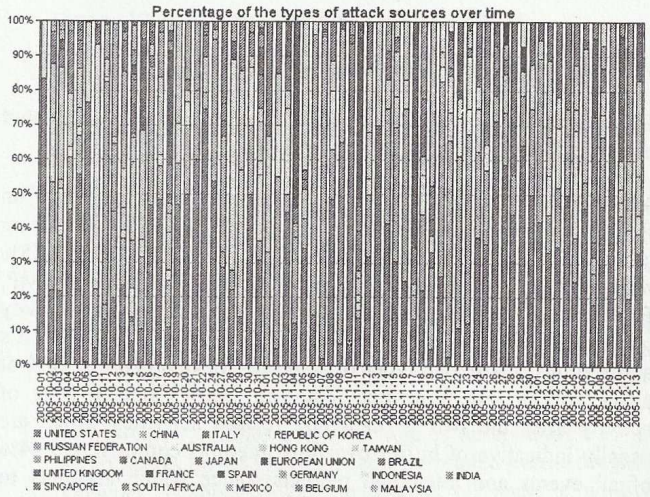

Figure 9 Time series distribution of country sources

The time series data, based on attack sources shows that attacks and security events coming from the United States and China occurs in alternating intervals. US based attacks peaked in the last week of October, the third week of November, and the last week and early part of December. Attacks based on China occurred at the intervals between the high US concentrations. Attacks from Korea and Italy, though having a significant amount have an infrequent distribution characterized by very heavy occurrences at very specific points in the time series as compared to the high amount, high distribution of the US and China. It is interesting to note, that though attacks from Taiwan are not that large, they are well distributed in the time series. A possibility is a correlation

with attacks coming from China but no study is currently planned to prove so.

4.2.6 Total distribution of events based on Snort classifications

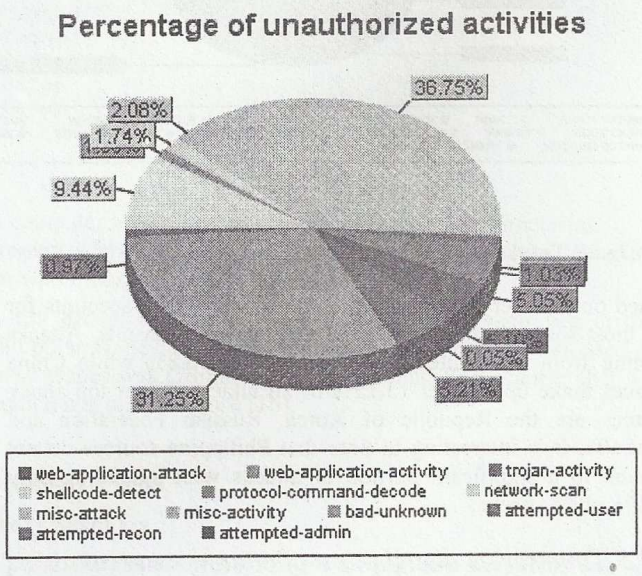

Figure 10 Total distribution of events based on Snort classifications

Miscellaneous activity and web-application activity, based on Snort classifications, makes up the most frequently occurring security related events comprising each of 36.75% and 31.25% respectively. Miscellaneous activity generally includes ICMP, FTP and SMTP activities among others. Web application activity generally comprises events directed to port 80 and it's applications. These two types of activities make up almost 70% of all activity in the honeynet. The next most frequent type of activity after the two is shellcode related activities which are usually indicative of buffer overflow attempts comprising 9.44% of all events and attempted reconnaissance which accounts to about 5.05%.

4.2.7 Time series distribution of events based on Snort classifications

Time series analysis of the types of activities show us that web application activity has been predominant in the early part of the quarter, particularly in the whole of October and the early part of November. Miscellaneous activities has been fairly well distributed but had a noticeably sharp increase in the latter part of November and the early part of December. Attempted recon and attempted admin activities, though small, occurs frequently in the time series as compared to the first two types. Shellcode related activities are infrequently distributed in the time series though they occur in large amounts when they do occur. What follows is a breakdown of the most common activity types and their corresponding attacks signatures.

4.2.8 Time series distribution of signatures categorized by top snort classifications

The top security related activity, based on Snort classifications are miscellaneous activities. The most common and most frequent attacks related to this activity are ICMP, FTP and SMTP related events. Most notable were the ICMP PING Cyberkit 2.2 Windows, which accounts for the most and is the most frequently occurring security event in this classification if not the whole honeynet data set. ICMP PING Cyberkit 2.2 Windows is actually a false positive and indicative of the probing attempts of the Welchia / Nachi worm which we indicated in one of our Honeynet Activity Monitor analysis [22]. Another consistent security event, though much lower in amount are FTP probe attempts, indicative of warez activity.

The second most frequent activity type was web application activity. This, together with the more specific web application attack signatures will be explained more in depth later when we delve into specific ports attacked. Based on the time series, Microsoft Windows WebDAV is the most frequently attacked application.

Time series analysis of attempted recon activities shows that TCP port scans are the most frequently occurring reconnaissance event by far. More specific signatures show that NMAP [23] is used most often in reconnaissance activities. As shown in the time series there was an ascending trend towards the whole of October and peaked in the early parts of November 2005. Scans began to drop off in the late part of November and most of December following the general trend of decrease seen in general number of security events.

4.2.9 Total distribution of top ports correlated with unauthorized events

The most attacked port is port 80, which makes up an overwhelming 70.24% consistent to the data shown above indicating the wide distribution of web application activity. The next most frequently attacked port is port 42 or the WINS service, which amounts to 11.36% of all total ports attacks. Port 21, the FTP service also makes up a significant 7.87% of all activities. Other notable ports are 1900, the UPnP service and 3306, which is commonly attributed to MySQL.

4.2.10 Time series distribution of top ports correlated to unauthorized events

Time series data of the ports attacked shows the overwhelming frequency of distribution of port 80 attacks for the last quarter of 2005. Except for some days in latter part of November and early part of December, which other ports where predominantly attacked, most of the days in the last quarter typically involved port 80. Port 42, the WINS service was not as distributed but had very large concentrations on specific days typically in the first half of the quarter. There were also overwhelming concentrations of port 1900 attacks on the latter part of the quarter though interspersed in varying intervals between November 24 and December 4 2005. There seems to be a notable increase in port 21 activities in the latter part of November and the early part of December.

4.2.11 Time series distribution and total distribution of signatures correlated with top ports

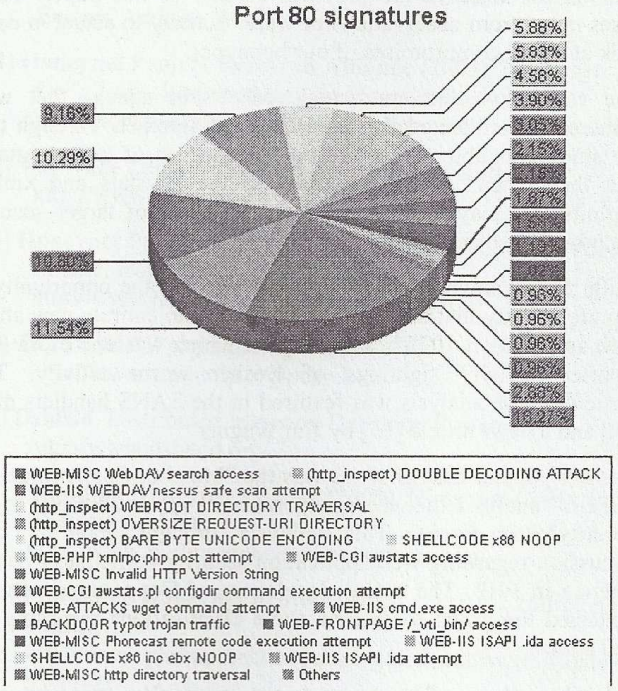

Figure 11 Port 80 totals

Port 80 events are the most frequently occurring events in the honeynet. Attacks directed towards WebDAV. Attacks as indicated by our Honeynet Activity Monitor reports show these as scans that determine the existence of Microsoft Windows WebDAV application and attacks exploiting the buffer overflow vulnerability of the said application. Together, the WebDAV scans and attacks consists of 18.27% and 10.80% respectively totaling to almost 30% of all port 80 attacks. Other attacks are those that try to exploit common IIS based Double Decoding (11.54%), Oversize URI (9.16%), Webroot Traversal (10.29%) and Unicode Encoding (5.88%) vulnerabilities. These set of attacks make up another 30% of all port 80 attacks. Aside from WebDAV and IIS related attacks, there are also notable ones directed towards specific web applications like XMLRPC.php, Awstats, wget, Cacti, and WebCalendar.

Analysis of the time series data on port 80 attacks shows that aside from a highly concentrated denial of service attack on October 6 2005, WebDAV attacks and scans were the most frequently occurring attack in the duration of data collection. The mechanism used in the denial of service attack on October 6 2005 involved a double decoding attack and web root directory traversal which accounted for the significant number of attacks attributed to the two signatures. WebDAV related attacks were predominant in the month of October. Beginning November, there was an increase in attacks directed towards specific web applications most notable of which was the Awstats and the XMLRPC.php library which are typically code injection attempts exploiting the input validation vulnerabilities of the said applications. The end of November and the most of December shows a drop in port 80 attacks.

4.2.12 Total distribution of top signatures

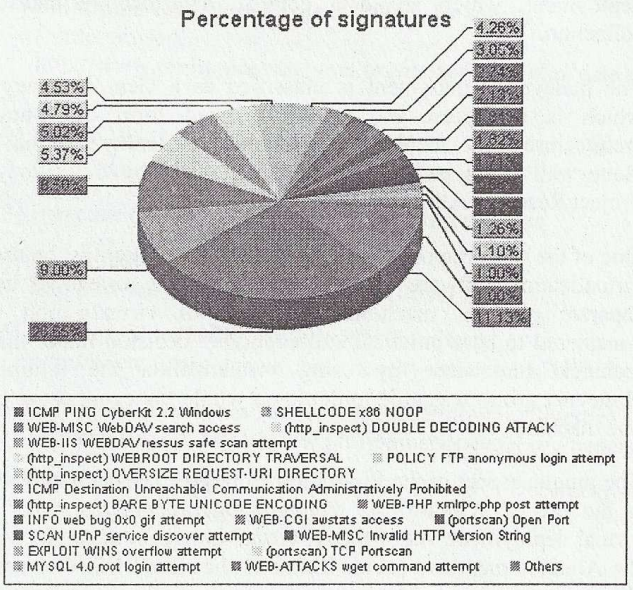

Figure 12 Total distribution of top signatures

So far, the largest percentage of security related events comes in the form of ICMP PING CyberKit 2.2 at 26.65%, which we described as the preliminary recon activity of the Welchia / Nachi worm though we cannot discount the possibility that this could be other recon activities from some other sources. The next security related event was shellcode activity at 9.00%, an event that are indicative of diverse buffer overflow attacks on different services. The most specific attack that have the highest occurrence are WebDAV based scans and attacks at 8.50% and 4.79% respectively. We have isolated most of the activity coming from the Welchia / Nachi worm. IIS URL based attacks that exploit the double decoding, Webroot directory traversal and oversize request URI vulnerabilities follow with 5.37%, 4.79% and 4.26% respectively. Another event of note is anonymous FTP probes, which makes up 4.53% of all security related events caught by the honeynet. Notable spikes on attacks on specific applications, services or libraries this quarter involve the following: WebDAV, IIS, FTP server, XMLRPC.PHP, Awstats, WINS, MySQL, Cacti WebCalendar, UPnP

5. CONCLUSION

5.1. Data Collection Infrastructure

This study has successfully deployed a honeynet based on the standards and framework of the global Honeynet Project Research

Alliance. This is the only officially recognized honeynet implementation in the Philippines which follows the framework set by the Honeynet Project in the United States. This honeynet adheres to the three major requirements of a honeynet deployment, which are data control, data capture and data collection.

The honeynet deployment is classified as a Gen III honeynet, which is the latest and most advanced form of honeynet architecture. The core of the honeynet deployment is the Honeywall ROO, a bridging gateway made by the Honeynet Project Research Alliance.

One of the most important aspects in this deployment is the use of virtualization software to create the honeynet instead of using separate physical machines. The use of virtualization was considered to be a practical and economic decision rather than a technical one since by using virtualization the Philippine Honeynet group was able to deploy a whole honeynet using only one machine.

The unique aspect in the Philippine Honeynet Project deployment is the use of Windows as the base Operating Systems of our virtual deployment. As of today, virtual honeynet deployments of the Alliance members prefer Linux as their base OS. Our use of Windows was dictated by practicality than anything else since the machine we were running the honeynet was being used for other tasks which required Windows to be installed.

5.2 New IT security tools and Techniques

This study has released 2 open source tools called the Honeynet Activity Monitor and the HoneyTrends historical graph generator. The Honeynet Activity Monitor is a front-end GUI used to extract and present data from Snort intrusion databases. The HoneyTrends meanwhile is a historical analysis tool for the Philippine Honeynet Project historical security events database.

Both tools produce charts that illustrate different aspects of honeynet activity. The Honeynet Activity Monitor primarily deals with daily activity while the HoneyTrends deals with much more long term data. Both tools have the capability to correlate different information elements that exists in the database.

The Philippine Honeynet team has been contacted by Snort, considered the de facto standard in Intrusion Detection, to request the Honeynet Activity Monitor to be included in their website. In a related event, the Philippine Honeynet Project has been recognized by Snort as an official Snort-related project.

These tools are freely available to anyone and we have been contacted by various individuals and organizations regarding these two tools. Currently, the Philippine Honeynet team are talking to people in Bridgestone, Eastern Telecoms and PICA of Hong Kong who are interested in the Honeynet Activity Monitor. Even Lance Spitzner, the founder of the Honeynet Project has suggested integrating our tools into the Honeywall ROO bridging gateway.

5.3 Collection, documentation and analysis of real world security cases

The study collects all data gathered by the honeynet infrastructure. From these daily sets of data, the researchers look out for cases that they would be relevant to the current IT security environment. From the start of data collection, the study has already documented over thirty cases as presented in a summarized form in the previous section of this paper. These cases range from observations of warez activity to actual in depth look at actual compromises of our honeypot.

The cases collected are actual real world attacks that were gathered and collected firsthand from the Internet. Through this, the study was able to capture unique variants of known attacks like the "listen" malware variant in the Awstats and xmlrpc exploits that may have escaped the attention of larger security organizations like antivirus companies.

Aside from known attacks, the study also got the opportunity to capture activity that led to the discovery of an entirely new attack such as the "Port 1025" case analysis, which was one of the first reported activity sightings of Dasher worm activity. This particular case analysis was featured in the SANS handlers diary [24] and a news article [25] by Jim Wagner.

Another popular case analysis was the "Mambo, Coppermine and PHPBB" analysis that the group released as an advisory in the Securityfocus forum. This analysis brought about a long discussion regarding web application security particularly coding practice in PHP. The actual developers of Coppermine actually contacted the group to give us some clarifications regarding the said attack.

5.4 Creation of a long-term security events and attack database

Data collected daily by the Philippine Honeynet infrastructure is archived into a historical database which can be used to generate time series data. Among the data included in this historical database are the date and time of attack, Snort signatures and classifications, source and destination ports, country sources, and attacker IPs.

By the end of this study, the Philippine Honeynet Project infrastructure has already collected three months worth of security data. This three month data set was used to generate a simple form of time series analysis shown in the previous section using the HoneyTrends tool. The data was first presented in the Open Web Application Security Project (OWASP) chapter meeting last November 2005. A subsequent report called "Philippine Internet Security Report" [26] which was based on the same data and was released in the first week of January 2006 and was published by a local newspaper [27] here in the Philippines.

6. ACKNOWLEDGMENTS

The author would like to acknowledge the members of the Philippine Honeynet Project namely Ms. Mida Guillermo, Mr. John Ruero, Mr. William Yu, Dr. John Paul Vergara and Mr. Carlo Monteverde

For without their assistance, this paper would not have been possible.

7. REFERENCES

[1] **Security Focus (2000)**, Honeypot Mailing List, http://www.securityfocus.org

[2] **Spitzner, L (2003)**, Tracking Hackers, Addison Wesley

[3 **Honeynet Project Research Alliance (2003)**, Know your Enemy, Addison Wesley

[4] **Honeynet Project Research Alliance (2003)**, Honeynet Definitions, Requirements, and Standards, www.honeynet.org/alliance/requirements.html

[5] **VMWare (2001)**, VMWare Virtualization Software, http://www.vmware.com

[6] **Honeynet Project Research Alliance (2004)**, Know Your Enemy: Honeywall CDROM Roo, http://www.honeynet.org/papers/cdrom/roo/

[7] **TCPDUMP**, Programming with PCAP, http://www.tcpdump.org/pcap.htm

[8] **Dshield**, Distributed Intrusion Detection System, http://www.dshield.org/

[9] **Philippine Honeynet Project (2005)**, Honeynet Case Analysis Archive, http://www.philippinehoneynet.org/data.php

[10] **Philippine Honeynet Project (2005)**, awstats / xmlrpc.php attacks, http://www.philippinehoneynet.org/dataarchive.php?date=2005-11-21

[11] **Linklogger (2005)**, TCP Port 1025, http://www.linklogger.com/TCP1025.htm

[12] **Philippine Honeynet Project (2005)**. awstats.pl access & configdir command execution attempt and xmlrpc.php, http://www.philippinehoneynet.org/dataarchive.php?date=2005-11-11

[13] **Juniper (2004)**, Cacti RRD Remote File Inclusion, http://www.juniper.net/security/auto/vulnerabilities/vuln1883.html

[14] **Philippine Honeynet Project (2005)**. xmlrpc.php post attempt, http://www.philippinehoneynet.org/dataarchive.php?date=2005-11-05

[15] **SANS (2005)**, XML-RPC for PHP Vulnerability Attack, http://isc.sans.org/diary.php?storyid=823

[16] **CA (2005)**, Grim's Ping, http://www3.ca.com/securityadvisor/pest/pest.aspx?id=42920

[17] **Philippine Honeynet Project (2005)**, Grimm's Ping, http://www.philippinehoneynet.org/dataarchive.php?date=2005-10-01

[18] **Philippine Honeynet Project (2005)**, FTP warez, http://www.philippinehoneynet.org/dataarchive.php?date=2005-11-14

[19] **Philippine Honeynet Project (2005)**, SMTP relays, http://www.philippinehoneynet.org/dataarchive.php?date=2005-10-19

[19] **Philippine Honeynet Project (2005)**, Proxy Scanners, http://www.philippinehoneynet.org/dataarchive.php?date=2005-10-05

[20] **Secunia (2004)**, Microsoft Windows WINS Server Buffer Overflow Vulnerability, http://secunia.com/advisories/10835/

[21] **Philippine Honeynet Project (2005)**, Honeynet Compromised!, http://www.philippinehoneynet.org/dataarchive.php?date=2005-10-07

[22] **Philippine Honeynet Project (2005)**, WebDAV attacks (http://www.philippinehoneynet.org/data.php)

[23] **Insecure (1998)**, NMAP, http://www.insecure.org

[24] **SANS (2005)**, SANS handlers diary, http://isc.sans.org/diary.php?date=2005-12-17

[25] **Wagner, Jim (2005)**, No Friendly Reindeer, http://www.internetnews.com/security/article.php/3571741

[26] **Philippine Honeynet Project (2006)**, Philippine Internet Security Report, http://www.philippinehoneynet.org/papers

[27] **Oliva, Erwin (2006)**, Web attacks in RP up in December - Internet Security Group, INQ7