TrendCenter Phase I: Final Report

TR-2002-05.01

Todd Heberlein
Net Squared, Inc.
todd@NetSQ.com

Table of Contents

List of Figures

1 Introduction

This report covers the work Net Squared performed on a Phase I SBIR project named TrendCenter. The period of work covers nine months beginning at the end of April 2000 and running through January 2001.

During the course of the contract we decided to combine our work with SANS' Global Incident Analysis Center (GIAC). Our TrendCenter vision was similar to GIAC's plans, so we decided it made sense to combine our efforts rather than develop two different sensor grid communities. Accordingly, throughout this report the term "we" often refers to the combined efforts of Net Squared and GIAC. Throughout the Phase I work Net Squared used GIAC data, and we prototyped our own portal environment. This was not the same portal users saw when visiting GIAC's site however. The TrendCenter portal we developed should be viewed as an experimental prototype from which GIAC will gain ideas and technologies and then transfer them into their own web site.

The name "TrendCenter" is derived from two goals for this project: (1) track the trends in the threat environment, and (2) provide a portal, or a center, that individuals can visit to understand and learn how to respond to the threat environment in which they work.

The vision for TrendCenter was initially inspired by similar organizations that track biological threats (as opposed to the cyber threats that we track) such as the United States' Center for Disease Control and Prevent (CDC) and the United Nations' World Health Organization (WHO). In other words, TrendCenter's goal is to create a Cyber CDC.

While not exactly a mission statement, the following definition of surveillance succinctly captures the capability we envision for TrendCenter:

> Surveillance is the *ongoing systematic* collection, collation, analysis and interpretation of data; and the dissemination of information to those who need to know in order that action may be taken.
>
> Principles of Disease Surveillance
> World Health Organization

Figure 1: Definition of Surveillance

Section 2 provides an overview of the surveillance process and touches on the areas into which we go into greater detail later in the report. Section 3 covers a wide range of analysis issues. This includes an analysis of the value of sensors to an individual site in isolation. This leads into an analysis of how forming a community can greatly increase the value of the sensors. Then we look at a number of different analysis techniques that we can use as part of a surveillance effort. Section 4 goes into a little more detail on the final step in the surveillance process: human interpretation. TrendCenter, in addition to creating technologies that can perform analysis, forms a community of experts that can contribute to the interpretation of TrendCenter's analysis. Section 5 presents a current working prototype we developed to test and demonstrate many of the core TrendCenter concepts. Finally, Section 6 summarizes our work.

2 Surveillance

Surveillance is a process, and Figure 2 captures the critical steps in this process. Throughout this section we briefly describe each of these steps and how they apply to our TrendCenter model.

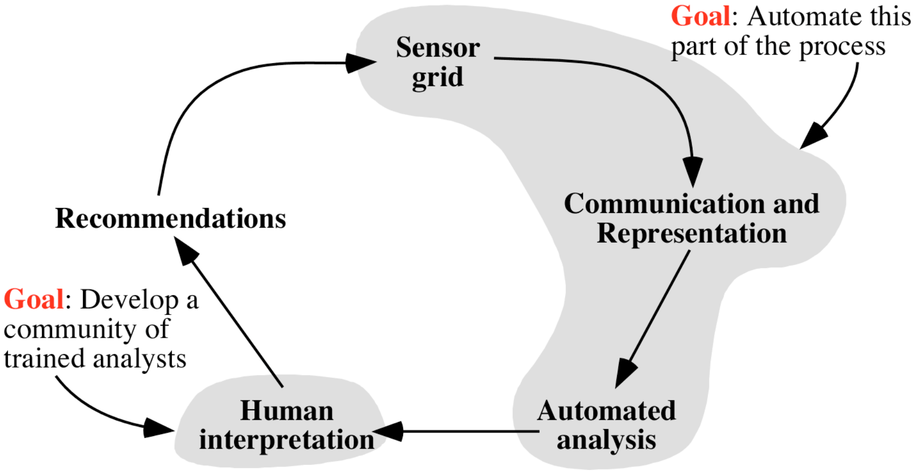

Goal: Automate this part of the process
Sensor grid
Communication and Representation
Automated analysis
Human interpretation
Recommendations
Goal: Develop a community of trained analysts

Figure 2: Intrusion Detection Surveillance Process

We begin at the top with the Sensor grid. Before any surveillance system can begin, it must have in place some means to measure the activity of interest. In the medical world doctors fill out forms about patients and submit them to a central organization such the Center for Disease Control and Prevention (CDC). For our situation various security sensors collect information. For our phase I work we primarily looked at Snort, a network-based intrusions detection system. We chose this sensor because several organizations were submitting their Snort sensor logs to SANS' Global Incident Analysis Center (GIAC), and we were using their data.

After sensors in the field collect the data, it must be communicated and stored in a central repository. Throughout Phase I we examined several approaches for achieving these goals. For example, the Internet Engineering Task Force (IETF) Intrusion Detection Working Group (IDWG) is in the process of developing standard reporting message formats and communication protocols. These designs are changing frequently, and for the first prototype we chose not to go with their approaches. For uploading data we simply used HTTP's built in support for uploading files to web servers. For storage, we used a relatively straightforward set of SQL tables and the MySQL database management system.

Once data has been collected from the field, it is processed through a series of automated statistical analysis algorithms to identify important features in the data. Examples of such analysis include simple statistical summaries, trend analysis, and various data mining techniques. We examined a number of methods, and we searched for other models in the real world. We found Amazon.com provided an excellent model we can emulate. More information on these analysis techniques can be found in Section 3.

After data has been collected and processed by automated analysis, the analysis results need to be interpreted. Human experts must perform this interpretation. TrendCenter, by forming a community of trained analysts from many organizations, will increase the likelihood that the results of the analysis are interpreted correctly. In Section 4 we go into more detail on the need for human interpretation.

The final step of the surveillance process is to communicate the results of the analysis and interpretation to users in the field. The system, network, and security administrators at individual organizations are the ones who affect change in the Internet, and TrendCenter is useless unless it becomes an effective tool for these individuals. To that end, a major part of TrendCenter is network portal to communicate the appropriate information in a timely fashion to the people in the field. In Section 5 we present an early version of an operational TrendCenter portal prototype.

3 Analysis

The core aspect of TrendCenter that makes it unique is our approach to analyzing data from many different organizations. We take a relatively unique view of intrusion detection sensors. Instead of using them to detect attacks, we use them to predict attacks. No single site can do this. Predicting attacks that individual sites might see requires forming an aggregate sensor grid and then analyzing the data feeds from those sensors in special ways.

In Section 3.1 we question whether signature-based intrusion detection systems provide any real value. In Section 3.2 we provide the answer, but then we claim system administrators can be much more effective if they had a crystal ball that could tell them what vulnerabilities at their site will likely be attacked. In Section 3.3 we discuss how a community of sites can form the basis of that crystal ball. Finally, Section 3.4 looks at many techniques we can borrow from Amazon.com to make TrendCenter more effective.

3.1 Intrusion Detection Paradox

Today's popular intrusion detection sensors are primarily signature-based sensors. That is, they look for activity they know is malicious. For example, a sensor may look for the exploitation of the "SITE EXEC" bug in FTP servers. One of the primary benefits this approach provides is that because known threats usually have known countermeasures, a sensor can provide additional value by including the solution to the attack. For example, if a sensor detects an attack exploiting the "SITE EXEC" vulnerability, the intrusion detection system can immediately provide the operator with details on the vulnerability and locations for security patches from the various vendors.

However, there is an important point to remember: securing a system *before* an attack is considerably cheaper than cleaning up *after* an attack. Once an attacker has established a beachhead on a machine, he may modify files, install Trojan Horses (e.g., to capture login passwords), install backdoors to get in later even if the original vulnerability is patched, and break into other local computers that trust the one he has penetrated. The end result is that for systems that have successfully been attacked, the system administrator often has to reinstall the entire operating system, reapply all the patches, and recover applications and user data from older backups that could not have been contaminated by the attacker. All of this is a very expensive proposition.

So why not simply secure the machines before they are attacked? But if we do this, then we are led to a paradox:

> **If intrusion detection systems primarily detect known threats, and those threats already have known countermeasures that should be applied by a proactive system administrator, what is the value of a signature-based intrusion detection system?**

3.2 Costs of Securing Versus Cost of Responding

The answer to the paradox posed in the previous section is that there are simply too many vulnerabilities to secure. With services such as BugTraq reporting roughly a dozen vulnerabilities a week, even the most proactive system administrator cannot apply all countermeasures to all of his systems all of the time. Furthermore, only a small number of these vulnerabilities will be exploited by attacks. Thus, while on a per system basis, being proactive and installing all patches all the time is cheaper than reacting to a successful attack, when looking at a site of thousands of machines, being proactive starts to lose its cost advantage.

More formally, let c_s represent the cost of keeping a single machine completely secure for some period of time (say six months), and let c_a represent the cost of having to respond to an attack against a single machine. Generally, c_s is much less than c_a:

$$\mathbf{c_s << c_a}$$

Now let n represent the number of machines at this site, and let m represent the number of those machines that will be successfully attacked during the same period of time. In general m is much smaller than n. That is, only a small number of machines at any site will be successfully attacked.

$$\mathbf{m << n}$$

Now the total cost for proactively securing all these machines is given by c_s*n, and the cost of responding to all successful attacks is given by c_a*m. Now we must ask ourselves: Is the cost of being completely proactive less than the cost of being reactive? That is, is the following equation true?

$$\mathbf{c_s n < c_a m}$$

While this equation may still be true, the numbers are not entirely convincing. And this opens the opportunity for using intrusion detection systems, even if they primarily detect attacks that we could have prevented.

But what if the system administrator had a crystal ball and could see ahead of time what machines would probably be attacked? If he had this capability he could concentrate is proactive measures on securing that much smaller subset of computers, and our equation becomes:

$$\mathbf{c_s m << c_a m}$$

Our goal is to make TrendCenter that crystal ball.

3.3 Prediction Requires a Community

The key to turning TrendCenter into a crystal ball for system administrators is that we move from securing a site by itself to securing a community. Figure 3 shows two models for system administrators securing their sites. On the left is the typical view: the system administrator sees his site as an isolated island. The problem with this approach is that the first time he sees an attack is when it hits his site. He has no vision beyond the perimeter of his own site. However, on the right is the new view we are creating with TrendCenter: the system administrator sees his site as one member in a larger community. This gives him a vision beyond his immediate perimeter, so he can see the attacks coming before they actually reach his site.

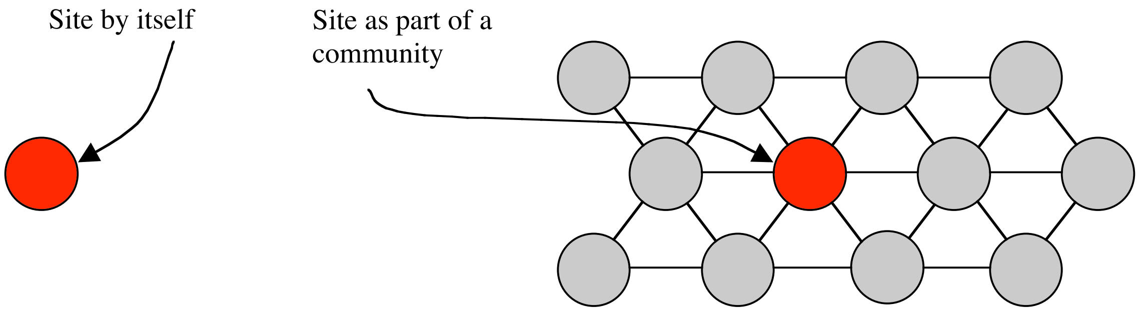


Figure 3: Site As Part of a Community

For example, suppose 1000 sites participate in the TrendCenter intrusion detection sensor grid. Then the chance of any one of those sites being the first victim is only 1/1000, a very small number. Even if you wait for the first 100 sites in the community to be targeted by a specific attack before you take proactive measures to secure your site, there is only a 1/10 chance that you will be attacked before you have a chance to secure your site against the attack.

Thus, by forming a community, a system administrator at any individual site can predict which attacks he might see at his site by observing what attacks the other sites in the community are detecting. In short, each site's intrusion detection system helps the other sites predict what attacks they might see in the near future.

3.4 Amazon Techniques

In the previous section we briefly showed that TrendCenter, by forming a community of sites, could help any individual site predict the attacks they are likely to see. Throughout this section we examine several techniques we have been exploring to refine the effectiveness of TrendCenter's prediction capability.

Interestingly, one of the best existing models TrendCenter can emulate is the online retailer Amazon.com. TrendCenter will track hundreds, perhaps thousands, of vulnerabilities and attacks each year. Simply showing large lists of information to a system administrator would be ineffective. System administrators do not have limitless time to browse all the information that TrendCenter will hold. We must develop strategies to take a large amount of data and present just the most relevant portions to each TrendCenter user.

Similarly, Amazon sells millions of individual products, and simply showing a customer one large list of products would be ineffective. Amazon knows the customers' patience is limited, so they must develop strategies to identify from their enormous inventory a small selection of products that each customer will likely buy.

So Amazon and TrendCenter have very similar problems. We both have an enormous amount of product or information to share with our users, but we can only present a small fraction of that product or information to each user. Amazon has been in this business for many years, and we can learn from their strategies. Sections 3.4.1 through 3.4.4 present several Amazon strategies we investigated for use within TrendCenter.

3.4.1 Top Sellers

Amazon's simplest approach to selecting a small set of books that you might buy from its inventory of millions books is to show you its list of top selling books. The premise is simple: if thousands of other customers think this is a good book, you might as well. Furthermore, generating the list is trivial for Amazon. Amazon also refines the approach by providing top selling lists for specific topic areas such as science fiction and history.

We are exploring an equivalent to Amazon's top sellers lists. In our case, we examine the top attacks used (as identified by sensor signatures) and the top ports targeted (e.g., rejected connections to port 21). This identifies the attacks that are reaching epidemic proportions. The approach also indicates attacks that may be part of self-propagating attacks (e.g., worms) or attacks that have been packaged in simple to use tools (e.g., more "script kiddies" have found a tool easy and/or effective and are using it more often).

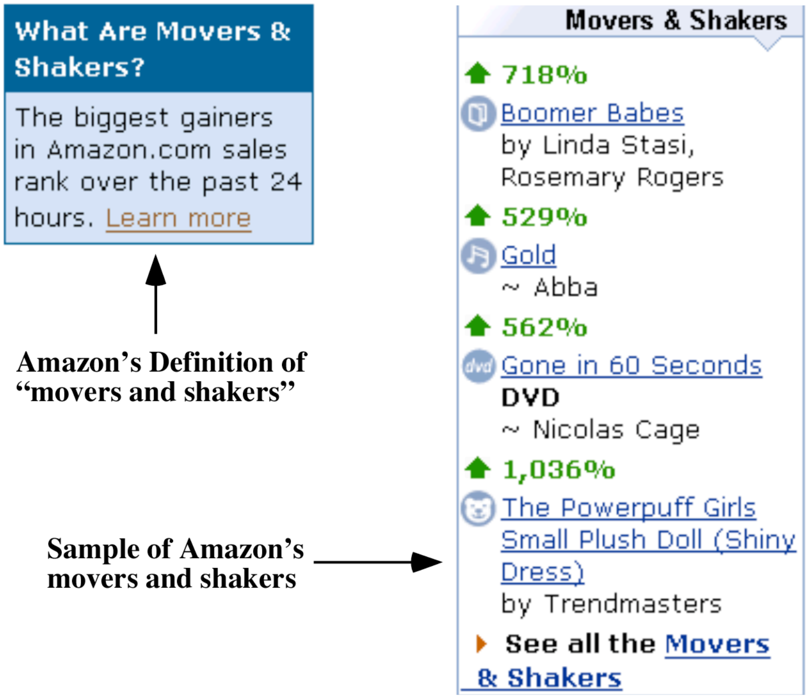
From the individual system administrator's point of view, this list provides a priority of vulnerabilities to fix. For example, if a system administrator only has time to fix 10 vulnerabilities, fix vulnerabilities associated with the attacks on this list first.

Also, as Amazon generates top sellers lists for specific topic areas, TrendCenter could generate attack counts for specific operating systems. For example, if your site is primarily a Linux shop, you may want a list showing only the top ten attacks against Linux systems.

The top sellers approach also adds value to anomaly-based detectors. The simplest example of this is simply counting the number of connections to specific ports that were rejected. This approach does not specifically identify a new attack, but if there is a sudden growth in the number of rejected connections to a particular port (e.g., port 21), then the community could conclude that there is a new attack against that port's service (e.g., FTP).

3.4.2 Movers and Shakers

Whereas the Amazon's "top sellers" approach shows the items that are the most popular today, its "movers and shakers" list shows the items that might make the top sellers list in the near future. Amazon defines movers and shakers as the items with the biggest sales rank gain over the past 24 hours. Figure 4 shows several example items on their movers and shakers list.



What Are Movers & Shakers?

The biggest gainers in Amazon.com sales rank over the past 24 hours. Learn more

Movers & Shakers

718%
Boomer Babes
by Linda Stasi, Rosemary Rogers

529%
Gold
~ Abba

562%
Gone in 60 Seconds
DVD
~ Nicolas Cage

1,036%
The Powerpuff Girls Small Plush Doll (Shiny Dress)
by Trendmasters

See all the Movers & Shakers

Amazon's Definition of "movers and shakers"

Sample of Amazon's movers and shakers

Figure 4: Amazon Movers and Shakers

Moves and shakers might be more formally defines as the items with the highest first derivative of their sales graphs (i.e., the graph's slope). The first derivative provides a means to estimate future values for the sales graph.

For our purposes, movers and shakers are attacks that are relatively rare today but have graphs indicating rapid growth. These are the attacks that might be in the "top sellers", or "top attacks", tomorrow. If a system administrator knows the attacks that will be popular tomorrow, he can prepare for them today. If a system administrator has secured the most popularly exploited attacks already, and he has time, he may want to secure the vulnerabilities exploited by these mover and shaker attacks.

3.4.3 Purchase Circles

Another approach Amazon uses to identify books that might interest a particular customer is to define purchase circles. Purchase circles define hierarchical groupings into which customers can be placed. For example, Figure 5 shows one of Amazon's purchase circle hierarchies. The outer-most group is US government customers. The next refined purchase circle is the US military, and finally the most specific purchase circle is the Air Force. Because of similar backgrounds, locations, and interests, customers in the Air Force might be interested in books that others in the Air Force have purchased.

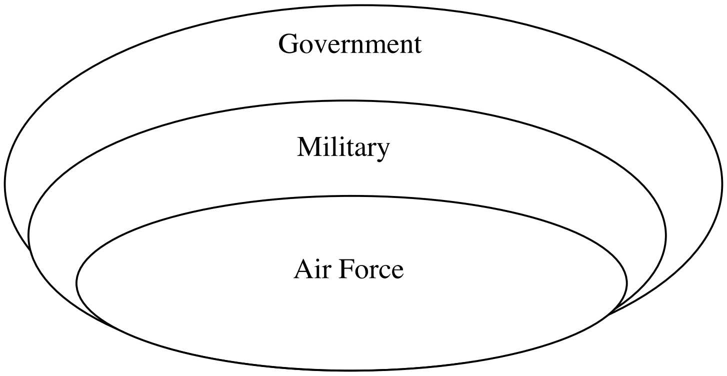


Figure 5: Amazon Purchase Circle Hierarchy

Figure 6 shows Amazon's top four books that are unique to the Air Force as well as the top selling books overall. As the can be seen, the four top selling books are all Harry Potter books. This is virtually identical to the interests of the general population. However, the unique books do seem to give you a better feeling for what makes the Air Force the Air Force.

Rank	Unique To U.S. Air Force	Bestsellers for the U.S. Air Force
1	Air Power: A Centennial Appraisal	Harry Potter and the Goblet of Fire
2	They Also Flew: The Enlisted Pilot Legacy, 1912-1942	Harry Potter and the Chamber of Secrets
3	Victor Padrini: A Novel of the United States Air Force Academy	Harry Potter and the Sorcerer's Stone
4	The Limits of Air Power: The American Bombing of North Vietnam	Harry Potter and the Prisoner of Azkaban

Figure 6: Unique Versus Top Sellers to the Air Force

We can use a similar approach, perhaps illuminating subtle but important attacks. For example, suppose we define a "purchase circle" consisting of members of the power grid. "Best sellers", or "top attacks", for the power grid is simply the top attacks power grid members see, but these are probably similar to what the general population also sees. Much of this probably consists of viruses, worms, and script kiddies – what we call random acts of violence.

But these "top attacks" may not represent the attacks that are the most dangerous to members of the power grid. If a terrorist group or nation state is launching subtle but systematic attacks or probing of power grid members, then these attacks, while low on the "top attacks" list would show up in the "unique to power grid" list. In other words, a site's most popular attacks might not be the most important attacks.

3.4.4 Personal Recommendations

Another approach that Amazon uses to identify books you might likely buy is by building personal recommendations based on past purchase habits. By comparing your past purchases with all the other customers in their database, Amazon can identify customers with similar purchase patterns to you. Then books that these customers have purchased that you have not may be books that you would find interesting as well.

For our purposes, instead of identifying people with similar purchase patterns we identify sites with similar attack patterns. For example, if sites A, B, and C are frequently targets of the same attacks or the same attackers, and later if sites A and B see an attack, then site C should be notified about the attack.

4 Human Interpretation

So far this report has focused on presenting primarily statistical information to the customer. Top attacks, attacks on the rise, customer circles, and site-specific attack predictions are all based on relatively straightforward statistical calculations. However, a community-based portal as envisioned by TrendCenter can go further by tapping into the vast knowledge of its users.

This is particularly critical when it comes to understanding new phenomenon. Why are we seeing the numbers for a particular attack increase dramatically? Has a new easy-to-use tool been released to legions of script kiddies? Or has the attack been embedded into an automated worm? Or was there was an article about the vulnerability, attack, or attack tool posted on a popular news source like Slashdot?

Another problem we face with our approach is that we observer a considerable amount of suspicious activity for which we do not know of a specific vulnerability or attack. For example, our statistical analysis may indicate that there is an upswing in probes to port 21 (FTP). Does this mean there is a new attack tool for a known vulnerability? Or does it mean some group has discovered a new vulnerability that the wider community does not know about? Or is some hacker group just misleading us by faking attacks to port 21? If the increase in the number of probes is associated with a new vulnerability, what is the vulnerability? What applications, versions, and operating systems does it affect? What are the ramifications if the attack succeeds? Will the attacker have root shell access? Or can the attacker simply add or delete files?

The automated and statistical portions of a TrendCenter portal can identify changes to the threat environment, but understanding why the threat environment is changing is beyond the capabilities of the system. Human interpretation is required.

To support human interpretation, a TrendCenter portal needs to support a discussion forum where these questions can be posed, debated, and answered by technically skilled people from around the world.

5 TrendCenter Prototype

In Sections 2 through 4 we have discussed the components associated with a TrendCenter portal. These include sensor grids, methods to represent and store the data, analysis techniques to process the data, and the need to create a community-type environment to provide the critical human interpretation of the changing threat environment. In this section we give a brief tour of a prototype TrendCenter that we have been building in cooperation with SANS' Global Incident Analysis Center (GIAC).

The portal prototype is built using the Apache web server, the PHP scripting language compiled into the Apache server, the MySQL database, and the Linux operation system. Most of the content is database driven. We define an HTML skeleton page that defines the basic layout of the primary pages, and PHP code embedded in the HTML page queries a database to generate most of the content displayed in a user's web browser. This approach keeps the web pages constantly fresh with no work from a web master.

The prototype is not feature-complete, but many of the techniques mentioned previously are present in the web site.

5.1 Front Page

The page shown in Figure 7 is what a user would see when they first connect to the web site. Our goal is to present an overview of relevant content that a user can then drill down into for more details. Since the content is database driven, this page can be viewed multiple times throughout the day, and it will always reflect the freshest view we have of the security environment.

Major components of this front page are discussed in Sections 5.1.1 through 5.1.6.

5.1.1 Overall Threat Indicator

At the top center of the page is the Threat Level warning indicator generated by the SANS' Global Incident Analysis Center (GIAC). The Internet is always at war with hackers, but sometimes that war gets hotter than usual, and this indicator shows how hot the war is at the present moment.

Most of the time the threat indicator is at green, indicating that there is nothing unusual about the current level of network attacks. But if a new attack is rapidly rising, a new vulnerability is discovered that is in large numbers of machines and can easily be exploited by hackers, or if a new virus or email is circulating quickly, this indicator may move to higher warning levels.

We placed this indicator in the most prominent portion of the page because we wanted a system administrator to get an instant feel for how his day might go when he initially connects to the portal in the morning. If he connects and the indicator is green, then his day will probably be fairly standard: install new operating systems, create new user accounts, patch an application, and put out small fires. If the indicator is yellow or red, he knows that many of the normal system administration activities will be put aside in the morning as he assesses his risks to the new Internet threat.

5.1.2 Top Threats

On the top right of the page are three lists of "Top Threats". The first list is the top ports targeted by probes or attacks. This is similar to the "top sellers" list discussed in Section 3.4.1.

We would have preferred to name specific vulnerabilities that are targeted by attacks, but currently with the data available to SANS' GIAC we can only identify the ports. However, where possible we have mapped the server port to a service name, and this can help the user make the jump in their mind to the vulnerability being targeted.

The top probed port is 515, a port used by print spoolers. Probes and attacks to this port are probably trying to exploit a bug in WinCom LPD (BugTraq ID 1701 and CVE-2000-0839). The next most popularly probed or attacked port is 21, the FTP server port. Most of these probes are probably associated with the SITE EXEC attack that gives the attacker shell access on the FTP server (CVE-2000-0573). The third most popularly targeted port is 102, the POP mail server. In 2000 a large number of vulnerabilities in POP servers from many vendors were identified.

What does this information tell a system administrator? While a dozen or more vulnerabilities are announced every week, they should pay closest attention to their LPD, FTP, and POP servers and make sure they are fully patched. In short, this is a priority list for the system administrators.

5.1.3 Low and Slow

Following the top ports probed is the "Low and Slow" list. These are ports that are reportedly probed at a relatively low rate but are reported by a large number of sites. In the past, attackers have tried this approach to stay below the radar of any individual site. Perhaps the most famous example of this is the "Solar Sunrise" series of attacks during February of 1998. Only by aggregating reports from a large number of sites, and then determining which attacks are common to these sites, even if relatively rare in general, can such attacks be easily detected.

5.1.4 On the Rise

The third list in the Top Threats section is the "On the Rise" list. This list shows ports that are not on the top-ten list of probed ports, but they have a high rate of growth indicating they may be on that list in the near future. This is the equivalent to Amazon's Movers and Shakers list described in Section 3.4.2.

5.1.5 Security Relevant News

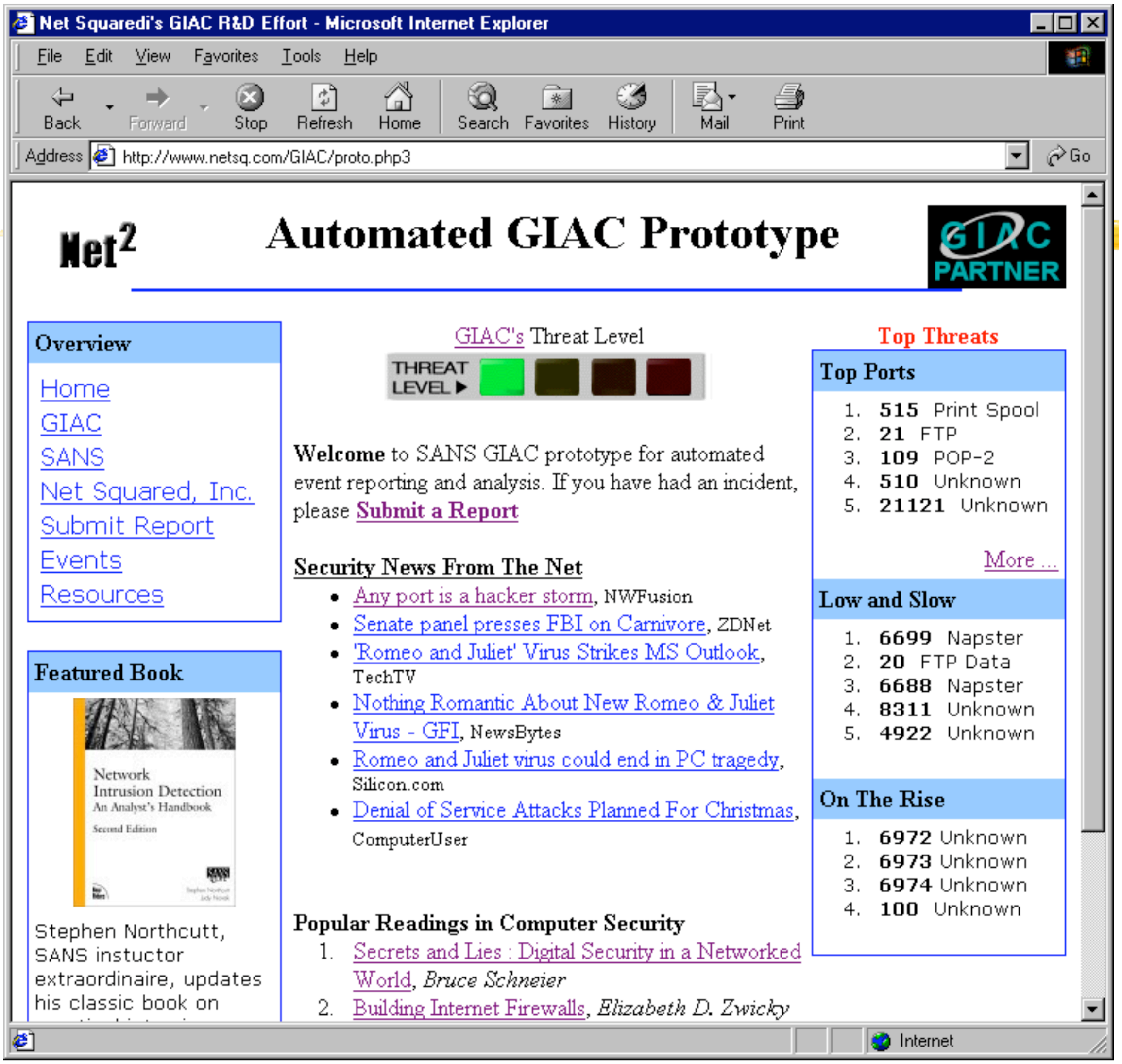
In the center of the front page is a list of "Security News From The Net". This is a list of recent news stories covering some aspect of computer security. We had two goals for including this material on our front page. First, our portal is about providing information to security professionals to help them in their work, and we believe there are many stories each week that can provide just the right nuggets of knowledge to help our user community grow professionally. An article might provide in-depth information on a particular attack, or it may provide step-by-step instructions for cleaning up a system after a particular virus attack. Second, the portal should provide some level of entertainment that makes the portal an enjoyable place to log into every day. Some of the articles are light hearted. And as the saying goes, misery loves company, so when a system administrator reads about another person's "day in sys-admin hell", he cannot help but identify with it.

We have another set of web pages (not shown in this report) that allows a selected set of users to add stories that they have found on the Internet. Because these stories are constantly added to and discovered on the Internet, this portion of the front page is very dynamic. Thus, even if the threat environment does not change hour to hour, the security administrator will be encouraged to check in several times throughout the day to check out the latest stories from around the Internet.

5.1.6 Educating the Community

Finally, on the lower part of the page we have the "Popular Readings in Computer Security" section, and in the lower left of the page we have a "Featured Book". TrendCenter provides this information to help educate the user community. These sections are updated over time, but much less frequently than the "Security News From The Net" section. The reason for the less frequent update is that this material tends to be much deeper (typically books instead of small articles), it is vetted more carefully so that only high-quality material is presented, and the value of the material does not decay as quickly. For example, a news article about a flaw in a web browser may hold the community's interest for a few days at most, but a basic strategy guide to building firewalls holds its value for months or even years.

By helping to educating the user community TrendCenter not only provides a service but it creates value for TrendCenter itself. Once again, as mention in Section 4, understanding why the threat environment is changing requires human interpretation, and creating a user community with greater technical breadth and depth will help in that interpretation process.


Net Squaredi's GIAC R&D Effort - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print
Address http://www.netsq.com/GIAC/proto.php3
Go
Net2
GIAC PARTNER
THREAT LEVEL ▶
Internet

Automated GIAC Prototype

Overview

- Home
- GIAC
- SANS
- Net Squared, Inc.
- Submit Report
- Events
- Resources

Featured Book

Network Intrusion Detection An Analyst's Handbook Second Edition

Stephen Northcutt, SANS instuctor extraordinaire, updates his classic book on

GIAC's Threat Level

🟩 ■ ■ ■

Welcome to SANS GIAC prototype for automated event reporting and analysis. If you have had an incident, please **Submit a Report**

Security News From The Net

- Any port is a hacker storm, NWFusion
- Senate panel presses FBI on Carnivore, ZDNet
- 'Romeo and Juliet' Virus Strikes MS Outlook, TechTV
- Nothing Romantic About New Romeo & Juliet Virus - GFI, NewsBytes
- Romeo and Juliet virus could end in PC tragedy, Silicon.com
- Denial of Service Attacks Planned For Christmas, ComputerUser

Popular Readings in Computer Security

1. Secrets and Lies : Digital Security in a Networked World, *Bruce Schneier*
2. Building Internet Firewalls, *Elizabeth D. Zwicky*

Top Threats

Top Ports

1.	**515**	Print Spool
2.	**21**	FTP
3.	**109**	POP-2
4.	**510**	Unknown
5.	**21121**	Unknown

More ...

Low and Slow

1.	**6699**	Napster
2.	**20**	FTP Data
3.	**6688**	Napster
4.	**8311**	Unknown
5.	**4922**	Unknown

On The Rise

1.	**6972**	Unknown
2.	**6973**	Unknown
3.	**6974**	Unknown
4.	**100**	Unknown

Figure 7: TrendCenter Front Page

5.2 Top Probed Ports

The purpose of TrendCenter's front page is to provide the user with an overview of several types of information without overwhelming him with too much detail. However, the user can drill down for more detail on various types of data. For example, in the "Top Ports" list on the right side of the front page (see Figure 7), the user can click on the "More…" link to jump to another page showing much greater detail on these ports (see Figure 8).

Note: the rankings of the top probed ports are slightly different between the screen shot of the front page and the detailed ports. This difference is because the pictures were taken on different days, and each page's content is generated dynamically by querying a database.

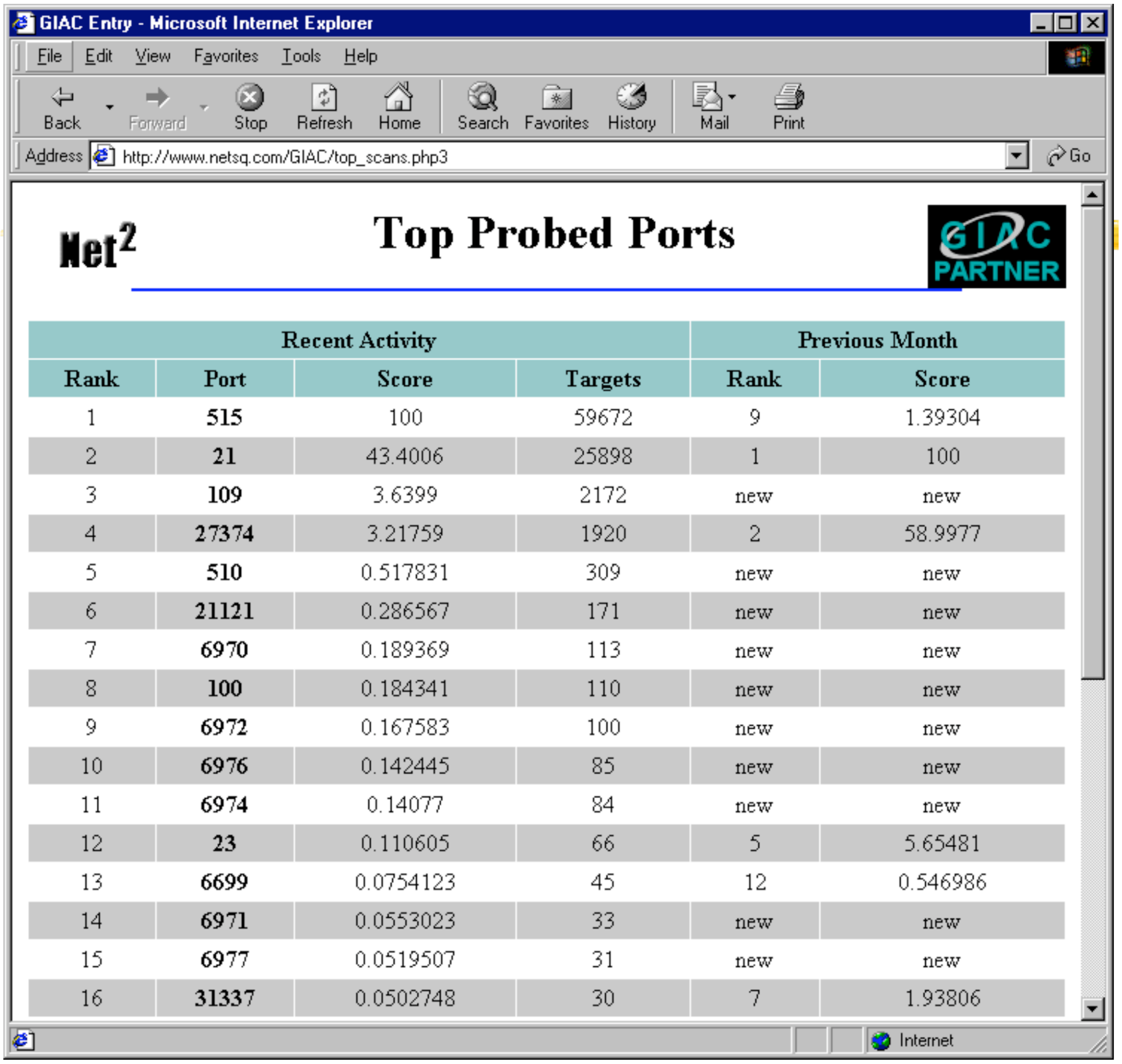
The top probed ports table is divided into two major groups: recent activity for the port and the previous month's activity for the same port. This gives the user a feel for the change in the threat environment.

The port that was probed most recently is port 515. As we mentioned previously, this is probably due to a recently discovered vulnerability in the WinCom LPD server. The probing is given a "Score" of 100. This is a relative score to compare how much one port is probed compared to another. The top probed port is always given a score of 100. Port 109, ranked 3rd, is only given a score of 3.6, so it is only probed 3.6% as often as port 515. In fact, from the Score column we can quickly see almost all the scanning is represented by the first two ports: 515 and 21.

Following the Score column is the Targets column. This shows the number of systems that have been targeted by a probe or attack to this port. For port 515, this number is 59,672. We will probably remove this column in the future.

The "Previous Month" information contains two columns: Rank and Score. These are the same values as the Rank and Score for the "Recent Activity", but they represent older data. As can be seen, port 515 was only ranked 9th in the previous month with a relative low score of 1.39. This big jump in rank and relative score is because the vulnerability is relatively new, and attack tools to exploit it have only recently circulated. On the other hand, attacks or probes to port 21, ranked 2nd with a relative score of 43.4 in recent activity, placed it at number one in the previous month. This is largely due to the fact that the FTP vulnerabilities, and attacks that exploit them, are older. Thus we see one attack is in its ascendancy (those to port 515), while another attack is in decline (those to port 21).

The third ranked port is port 109 (POP), and while it only has a relative score of 3.6, it did not even show up in the previous month's attacks. This may indicate that it is a new attack, and that we will see a lot more of these in the future.


GIAC Entry - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print
Address http://www.netsq.com/GIAC/top_scans.php3
Go
Net2
GIAC PARTNER
Internet

Top Probed Ports

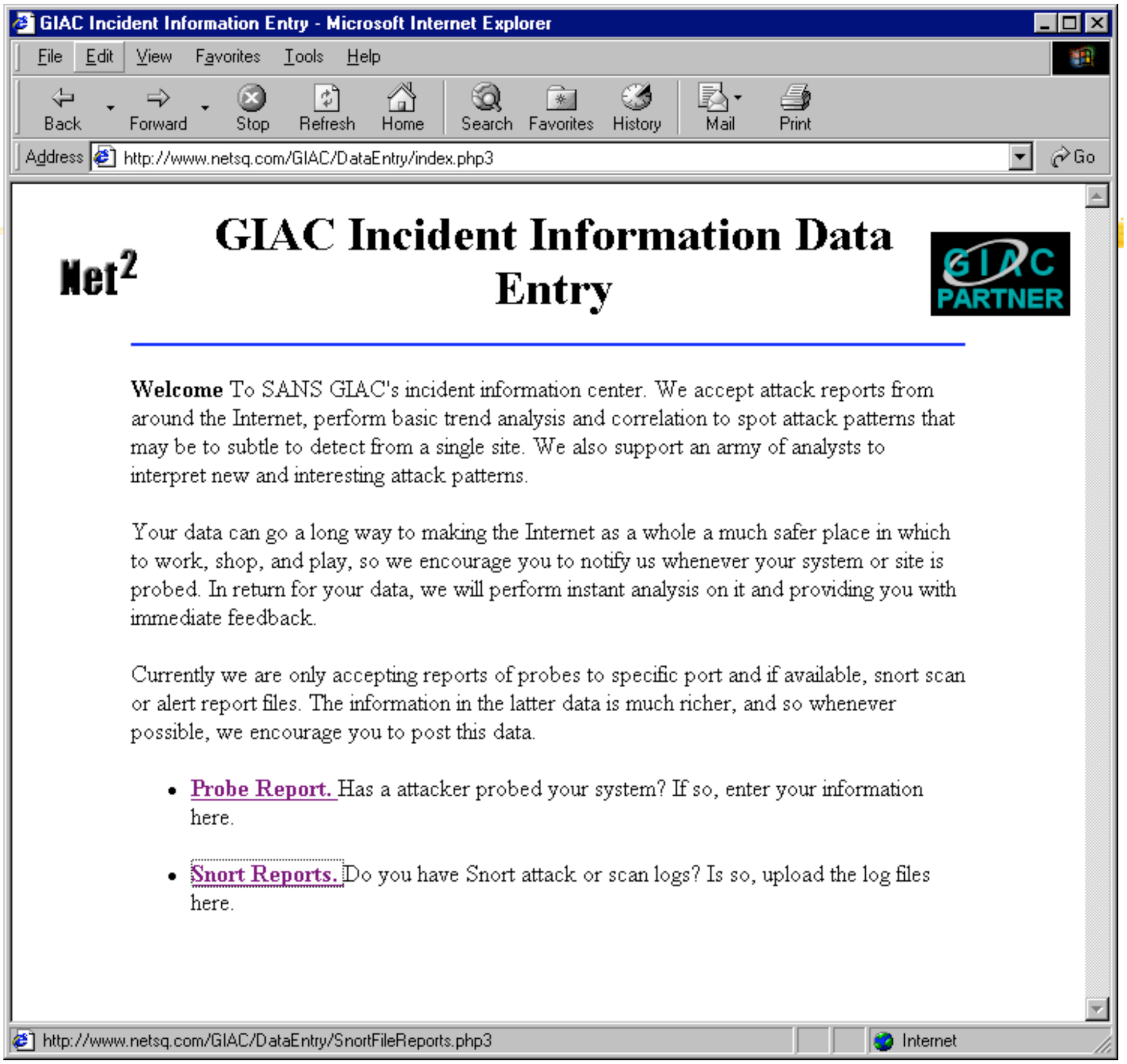
Recent Activity				Previous Month	
Rank	Port	Score	Targets	Rank	Score
1	**515**	100	59672	9	1.39304
2	**21**	43.4006	25898	1	100
3	**109**	3.6399	2172	new	new
4	**27374**	3.21759	1920	2	58.9977
5	**510**	0.517831	309	new	new
6	**21121**	0.286567	171	new	new
7	**6970**	0.189369	113	new	new
8	**100**	0.184341	110	new	new
9	**6972**	0.167583	100	new	new
10	**6976**	0.142445	85	new	new
11	**6974**	0.14077	84	new	new
12	**23**	0.110605	66	5	5.65481
13	**6699**	0.0754123	45	12	0.546986
14	**6971**	0.0553023	33	new	new
15	**6977**	0.0519507	31	new	new
16	**31337**	0.0502748	30	7	1.93806

Figure 8: Details of Top Probed Ports

5.3 Submitting Sensor Reports

On the upper left side of the TrendCenter front page (see Figure 7) is a link labeled "Submit Report". Clicking on this link takes the user to another page allowing him to submit various types of attack reports (see Figure 9). This is where we gather the raw data used in our various forms of analysis. Currently reports must be manually submitted, but we hope to provide automated applications for users that will sanitize and submit attack reports on a regular basis (e.g., daily) without requiring the user to take any specific actions.

Currently only two types of report can be submitted: (1) probe reports to only one service port (e.g., reporting a sweep of port 21 at your site), and (2) Snort reports. Eventually we hope to increase the types of reports to other popular sensors such as ISS RealSecure logs.


GIAC Incident Information Entry - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Forward
Stop
Refresh
Home
Search
Favorites
History
Mail
Print
Address
http://www.netsq.com/GIAC/DataEntry/index.php3
Go
Net2
GIAC PARTNER
http://www.netsq.com/GIAC/DataEntry/SnortFileReports.php3
Internet

GIAC Incident Information Data Entry

Welcome To SANS GIAC's incident information center. We accept attack reports from around the Internet, perform basic trend analysis and correlation to spot attack patterns that may be to subtle to detect from a single site. We also support an army of analysts to interpret new and interesting attack patterns.

Your data can go a long way to making the Internet as a whole a much safer place in which to work, shop, and play, so we encourage you to notify us whenever your system or site is probed. In return for your data, we will perform instant analysis on it and providing you with immediate feedback.

Currently we are only accepting reports of probes to specific port and if available, snort scan or alert report files. The information in the latter data is much richer, and so whenever possible, we encourage you to post this data.

- **Probe Report.** Has a attacker probed your system? If so, enter your information here.
- **Snort Reports.** Do you have Snort attack or scan logs? Is so, upload the log files here.

Figure 9: Submitting Attack Information

5.4 Submitting Snort Logs

When a user clicks on the “Snort Reports” link in the web page in Figure 9, TrendCenter displays another page allowing the user to locate the snort logs on his machine and upload them to TrendCenter (see Figure 10).

The first field in this form is the analyst’s contact information. We want this information for a couple of reasons. First, should the same attacker appear in the submission reports from several different sites, we may want to contact the appropriate analysts about this newly discovered serial attacker. They, in turn, may want to contact the other organizations that were attacked by the same machine in order to coordinate their investigations. In such a situation, TrendCenter would server as a trusted middleman to protect the identities of the individual analysts until each analyst agrees to reveal themselves to each other.

Second, we want to track which analysts are submitting the information because not all analysts are as skilled at setting up their sensors. Some analysts are extremely well trained and have carefully configured their sensors, so we have greater confidence that when they report an attack it is probably not a false positive. However, since anyone can join the TrendCenter community, some people with no training will submit reports that are full of errors. By tracking which analysts are submitting which data sets, we keep data from poorly trained analysts from corrupting the data sets and therefore the information provided by TrendCenter.

The second field in the snort entry form is a site name from which the data was collected. Once again, we have several reasons for wanting this information. First, several analysts may submit logs from the same site. For example, a site may have one analyst submitting logs during the weekdays and another analyst submitting logs during the weekend. In order to perform some types of analysis (e.g., the personal recommendations as described in Section 3.4.4), we need to make sure all the logs from one site are tracked together.

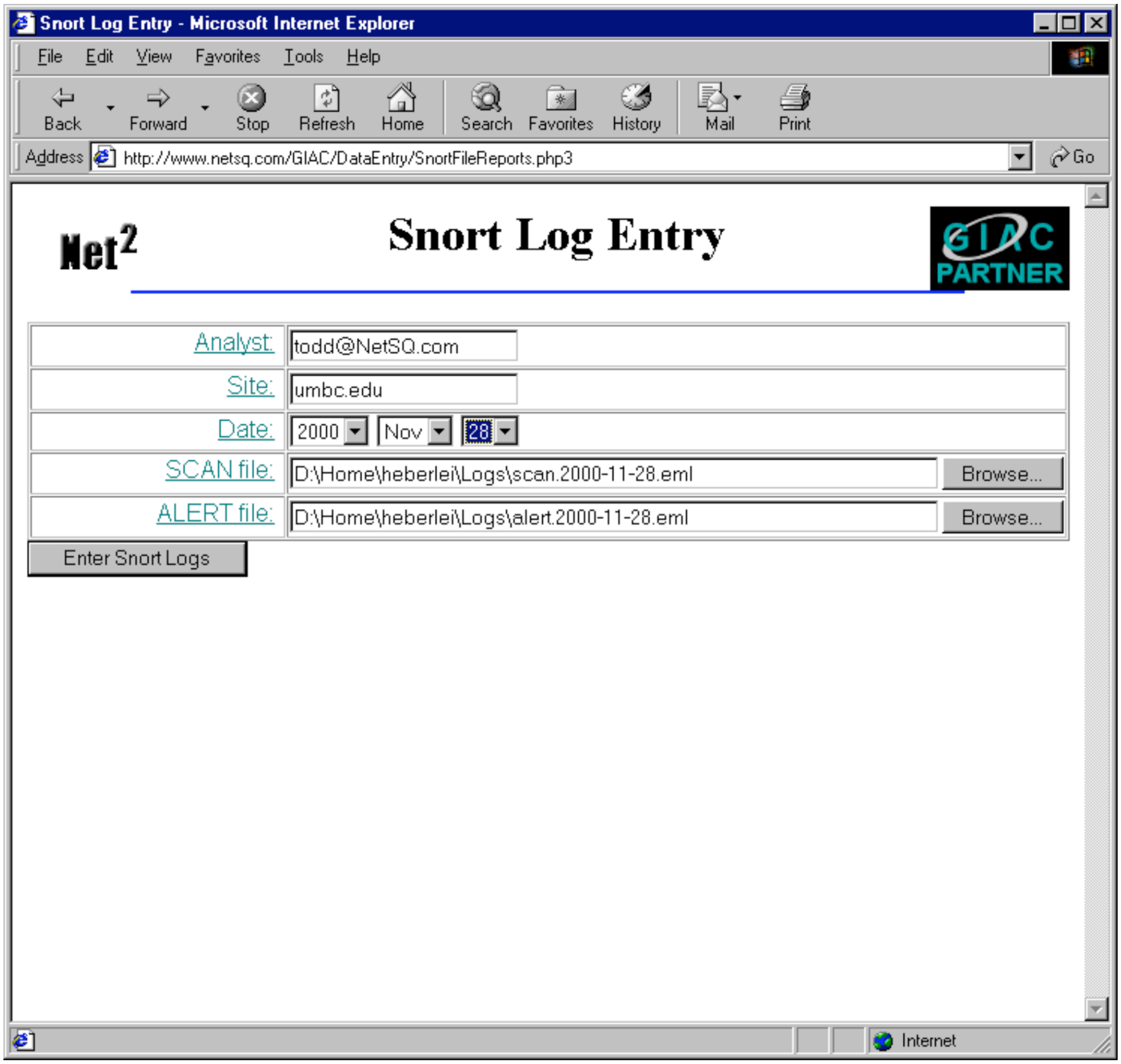
Second, one analyst may actually submit logs from several different sites. We foresee trained analysts providing out-sources intrusion detection support for many small organizations. For example, some health insurance companies may require all doctor offices that bill them to have some data services online, but small doctor offices cannot afford a fulltime system administration staff much less a staff to perform intrusion detection. A trained security specialist could offer his services to these local doctor offices. In such cases, when the analyst submits information, we need to know which organization the data comes from. In fact, in this example I access University Maryland Baltimore County's sensor logs and upload them into the TrendCenter database myself.

The next field in the form is the date of the log file. Generally attack log files contain dates and timestamps for each record, but sometimes they only contain a timestamp. For this reason, we ask each user to tell us the primary day that is covered by the log files. As this approach implies, we assume that the analyst configures his sensor to roll to a new file every 24 hours, and we hope the analyst submits the logs every day.

Finally the last two fields in the form identify the location of the sensor logs. Snort generates two types of sensor logs: scan logs and alert logs. Scan logs are not associated with a particular attack, but if a host appears to be connecting to too many hosts or to too many ports on the same host, Snort starts collecting a log of each connection attempt. The Alert logs are generally messages that are generated when a packet matches a specific rule in the Snort rule base. The Alert logs also contain summary messages about detected scans.

The analyst can either type in the full path to the log file, or, more likely, he can click the "Browse…" button to pull up a file browser and simply select the log file.

Once all the information is filled out, the analyst simply clicks the "Enter Snort Logs" button, and the two log files will be uploaded and processed by TrendCenter.



http://www.netsq.com/GIAC/DataEntry/SnortFileReports.php3

Snort Log Entry

Analyst:	todd@NetSQ.com
Site:	umbc.edu
Date:	2000 Nov 28
SCAN file:	D:\Home\heberlei\Logs\scan.2000-11-28.eml
ALERT file:	D:\Home\heberlei\Logs\alert.2000-11-28.eml

Figure 10: Submitting Snort Alerts

5.5 Custom Response

To encourage analysts to submit their sensor logs to TrendCenter (more logs from more sites help create a better TrendCenter), we provide extra information to these users that other users of TrendCenter will not receive. For example, using the personnel recommendations approach described in Section 3.4.4, we could generate a specific set of attack predictions for a site based on the information in the logs they provide to TrendCenter.

As for this prototype, however, we only generate a custom response based on the log file submitted. For example when the user clicks the “Enter Snort Logs” button in the Snort Log Entry page (see Figure 10), the log files go to TrendCenter, are processed, and a custom page is generated. Figure 11 diagrams this simple approach: submit the logs and receive a custom analysis page.

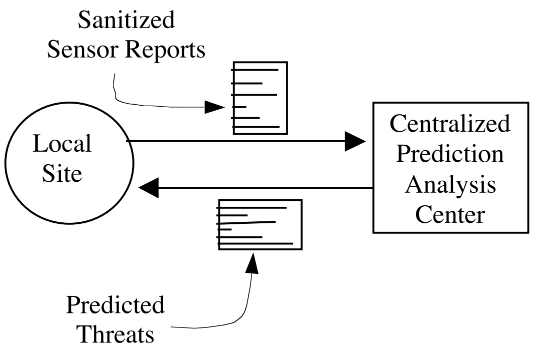
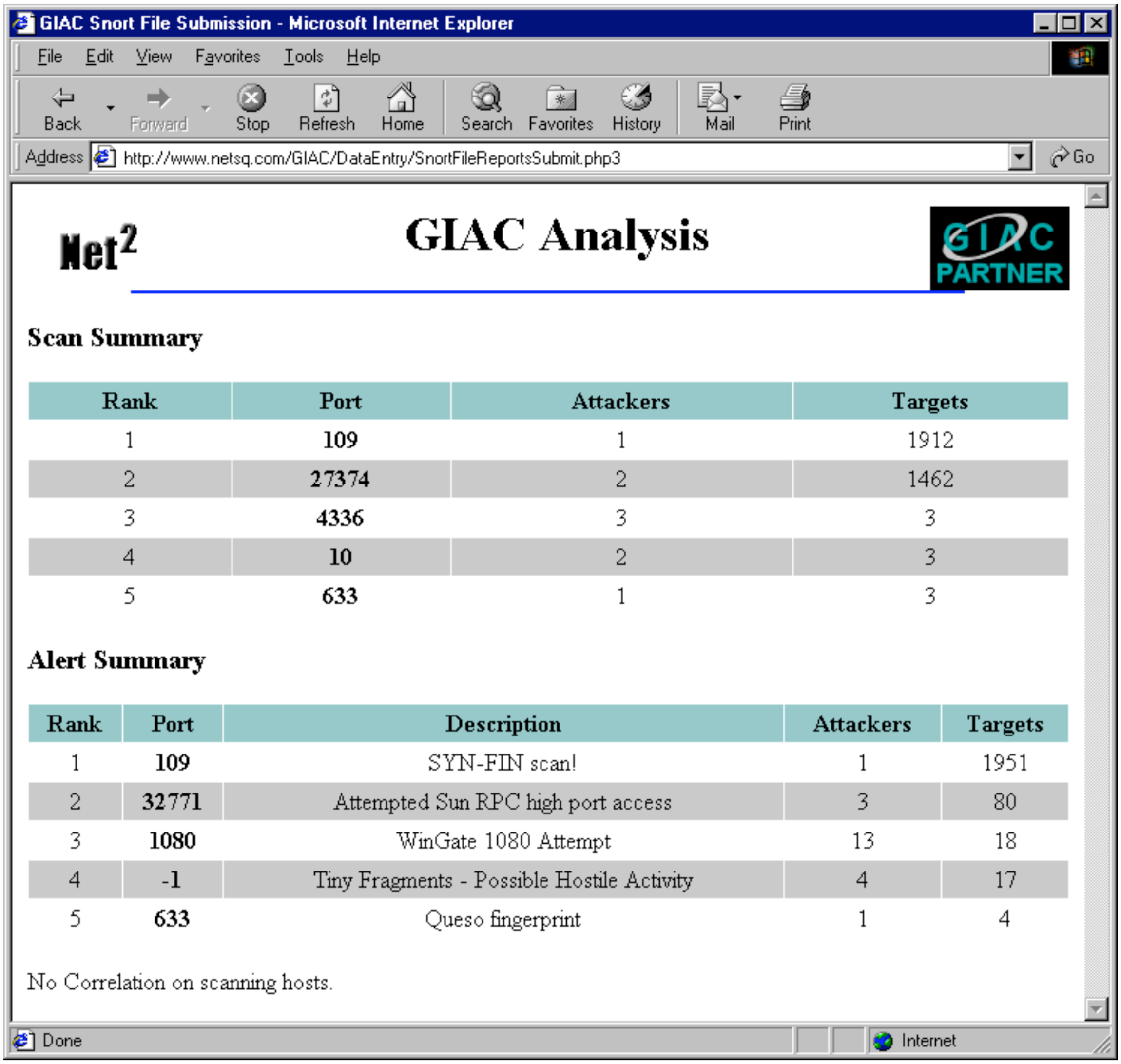


Figure 11: Providing Value to Sites Submitting Reports

In our current implementation we perform three types of analysis and present it to the user. Figure 12 shows an example of such a page from our prototype. The first analysis is a statistical summary of the *scan* logs submitted by the user. In this example, port 109 (POP) was scanned the most, with nearly 2,000 hosts targeted. Next, attackers probed port 27,374, the common port set up by the SubSeven Trojan Horse, on 1,462 machines. The remaining scans were so small, with at most three targets hit for each port, that these may have been false positives.

The second analysis is a statistical summary of the logs in the Snort *alert* file. In our example, the attacker that scanned for POP servers used a technique called a SYN-FIN scan that used to be used to bypass scan detectors (which used to look for packets that only had the SYN flag set). Each instance of the scan matched a rule in the Snort rule base, so 1,951 alert records were generated. Interestingly, this number is larger than the number of probes reported in the scan log file for port 109. The next most frequent attack in the alerts file was attempted RPC access to port 32,771. The common port for Sun's remote procedure call portmapper service is port 111; however, unknown to most users, Sun's operating system often sets up a second portmapper service at port 32,771 (we have no idea why Sun does this). While most firewalls will block attempted connections to port 111, they usually do not routinely block access to ports over 1024, so attackers can bypass the firewall to attack RPC services by contacting the service on port 32,771. Three different attackers tried this attack against a total of 80 machines.

So far, the analysts could do each of these statistical analyses locally without using TrendCenter, but at the bottom of the page we see a third type of analysis that cannot be done locally, namely, cross correlation with other sites. This analysis looks at the source address of the attackers in the Snort log files and searches the database for the same attackers observed by other sites. In this particular case we receive the message "No Correlations on scanning hosts" because no other sites reported seeing the same attackers. This may be due to the fact that we currently have only a handful of sites contributing to our current database. However, as more sites start to contribute their sensor logs, this type of analysis will become more and more fruitful. As multiple sites begin to observe the same attackers, TrendCenter can begin to support coordinated investigations.


GIAC Snort File Submission - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print
Address http://www.netsq.com/GIAC/DataEntry/SnortFileReportsSubmit.php3
Go
Net2
GIAC PARTNER
Done
Internet

GIAC Analysis

Scan Summary

Rank	Port	Attackers	Targets
1	**109**	1	1912
2	**27374**	2	1462
3	**4336**	3	3
4	**10**	2	3
5	**633**	1	3

Alert Summary

Rank	Port	Description	Attackers	Targets
1	**109**	SYN-FIN scan!	1	1951
2	**32771**	Attempted Sun RPC high port access	3	80
3	**1080**	WinGate 1080 Attempt	13	18
4	**-1**	Tiny Fragments - Possible Hostile Activity	4	17
5	**633**	Queso fingerprint	1	4

No Correlation on scanning hosts.

Figure 12: Custom Summary from Attack Submission

6 Conclusions

TrendCenter's goal is to develop a disease surveillance capability for the Internet. It does this technologically through aggregating intrusion detection sensor feeds from many sites and performing various types of analyses on the aggregated data in order to achieve a detailed view of the threat environment. But TrendCenter also accomplishes its goal socially by forming a community of normal system administrators, the agents of change on the Internet, and experts who can help interpret the change in the Internet's threat environment.

While we are inspired by analogous organizations in the biological world (e.g., CDC and WHO), we have found that many of the analysis techniques used by Amazon.com to be very valuable to our approach. We explored most of Amazon's techniques, and we have implemented some of these in our own prototype.

Finally, in conjunction with SANS' Global Incident Analysis Center (GIAC), we developed a prototype TrendCenter portal, complete with real data, to evaluate various visions and approaches for TrendCenter.