Understanding the World of your Enemy with I-CAT (Internet- Categorization of Attacks Toolkit)

Peter Mell
NIST, Computer Security Division
5-26-99

Abstract

Security professionals need to understand the attacks and vulnerabilities utilized by hackers to penetrate and shut down computer systems. However, security companies that collect such knowledge share very little of it with the general security community. The result is that security professionals must use the Internet as their source of computer attack and vulnerability information. While the Internet is a rich source of such information, the data is disorganized, distributed, and unverified. Because of the difficulty of collecting attack and vulnerability information on the Internet, security organizations spend an enormous amount of money sifting through the data. We are trying to mitigate this problem with the Internet- Categorization of Attacks Toolkit (I-CAT). I-CAT is a tool that allows security professionals to quickly glean attack and vulnerability information off the Internet. I-CAT has three main services: attack description lookup, statistics on the popularity of attacks, and measurements of current trends in attack publication. These three services enable security professionals to determine what kinds of attacks are available, what specific attacks are popular, and how to find information on the published attacks.

1.0 Introduction

Many computer security professionals need to possess and use attack scripts. Attack scripts are executable code or manual instructions that enable a user to compromise a computer system. Intrusion detection companies obtain attacks in order to figure out how to detect them. Vulnerability scanner companies obtain attacks in order to figure out what vulnerabilities they are exploiting. Penetration experts obtain attacks in order to perform penetration testing. Law enforcement and system administrators obtain attacks in order to understand what attacks could have been used to penetrate a compromised system.

While many security professionals feel the need to obtain attack scripts, few are willing to share what they obtain. Security companies do not share in order to keep a competitive advantage. Others do not share because they fear the liability of giving away dangerous tools. Whatever the reason, very little sharing of attacks occurs in the computer security industry.

The result is that computer security professionals must obtain attacks scripts from the Internet. A variety of hacker and security web sites exists that publish attack scripts for the public to download. Rootshell[1] has over 750 attack scripts. Fyodor's Playhouse[2] has over 350 attack scripts. The Legacy[3] has over 500 attack scripts.

While large collections of attack scripts exist on the Internet, they are either disorganized, poorly documented, or not searchable. The result is that security researchers spend an enormous amount of time and money finding attack scripts. Because of the time and money involved, small companies and research groups can not effectively find the attack scripts that they need.

2.0 Existing Sources of Attack and Vulnerability Information

It may seem strange that it is difficult for security professionals to find attacks on the Internet because the Internet is such a rich source of attack and vulnerability information. Mailing lists like Bugtraq[4] and NT Bugtraq[5] provide daily information on the latest vulnerabilities. Attack script web sites like Rootshell and Fyodor's playhouse contain scripts that enable one to launch and understand the latest

computer attacks. Emergency response teams like CERT[*6] provide advisories to warn system administrators when to patch their systems. However, each of these sources have drawbacks which hamper their use by security professionals.

2.1 CERT (Computer Emergency Response Team)

CERT maintains a public database of vulnerability information that is very useful to system administrators seeking to protect their networks. CERT educates people about the existence of significant vulnerabilities and provides details on how to protect sites and apply patches. However, CERT does not release detailed information on vulnerabilities which makes it almost useless to security companies who need to collect attack scripts. Services similar to CERT are offered by:

- Federal Computer Incident Response Capability (FedCIRC)[7]
- the Forum of Incident Response and Security Teams[8] (FIRST)
- Computer Incident Advisory Capability (CIAC) run by the department of energy[9]
- the Australian Computer Emergency Response Team[10] (AUSCERT)

2.2 Bugtraq and NT Bugtraq

Bugtraq and NT Bugtraq are widely read mailing lists that discuss newly found vulnerabilities in computer systems. They fully disclose all details in order to allow security professionals to understand the vulnerabilities. The problem with using these mailing lists is that they are high volume and the reader must wade through many documents discussing different facets of each vulnerability. While these lists are the best source of vulnerability information on the Internet, the security community needs the information in a more compact and organized format.

2.3 Attack Script Web Sites

Many security and hacker web sites exist on the Internet that provide public access to attack scripts. Rootshell and Infilsec[11] are examples of white hat attack databases while the Legacy and Fyodor's Playhouse are examples of hacker related sites. While these sites offer useful information, it is often hard to find what one needs. Even though some sites may organize attacks by operating system or provide a simple text search capability on the attack descriptions, security professionals need more sophisticated search techniques. Researchers need to be able to search for attacks by the goal of the attack, the target platform, the launching platform, the transmission method used, and many other features.

2.4 VulDa: Vulnerability Database

IBM, like many security companies, has developed a large database of attack and vulnerability information. VulDa[12] contains over 3.5 Gigabytes of compressed data that is available to the user using keyword searches and text mining techniques. While VulDa is very impressive, IBM is not ready to openly share it to the security community. VulDa is just one example of many corporate attack and vulnerability databases that can not be released to the general security community forcing security professionals to use existing resources on the Internet.

2.5 CERIAS[†] Vulnerability Database

CERIAS is making available to the security community a vulnerability database that describes vulnerabilities at a great level of detail. It is a heavyweight solution to the problem of sharing attack and vulnerability information in that it requires a substantial amount of work to analyze and enter a vulnerability into the database. If the security community would take ownership of this database then the workload could be distributed and a very rich and detailed attack and vulnerability database would be

[*] CERT is the Computer Emergency Response Team

[†] CERIAS is the Center for Education and Research in Information Assurance and Security run by Purdue University

available to the security community. Until then, we must make the best of what is available on the Internet.

3.0 Our Approach

We are working to make the attack scripts on the Internet more accessible to researchers until that day when the security community begins to share their data. We feel that the security community needs an interim solution to be used until a database like the CERIAS database is adopted by the security community and fully populated with all known vulnerabilities.

Our solution is I-CAT, the Internet- Categorization of Attacks Toolkit. I-CAT enables a user to find attacks published on the Internet. It can be used to find attacks that have a particular set of characteristics or attacks that could penetrate a particular type of host. In addition, I-CAT monitors which attacks scripts are most popular and takes statistics on attack publication on the Internet.

3.1 The I-CAT Attack Database

I-CAT contains descriptions of over 320 attack scripts published in the last 16 months. Each attack script is categorized by over 70 characteristics. The broad categories of statistics include: script goal, target type, transmission method, attacker platform, and requirements on the attacker to launch the attack.

We designed I-CAT such that each attack script could be categorized within five minutes. This small amount of time needed per attack enables the maintainers to easily keep up with the pace of Internet attack publication that by our estimate is no more than 30 to 40 new attacks per month.

For the initial I-CAT development we used only attacks published on Rootshell. Rootshell is one of the most important attack script publication sites and it publishes the majority of publicly available attack scripts. While it is difficult to remotely measure the popularity of a web site, we know that Rootshell has close to 30,000 people on its mailing list and receives hundreds of thousands of queries to its search engine every month.

We categorize attack scripts using the I-CAT form shown in Figure 1. The example attack shown is smurf. The figure reveals that smurf is a remote denial of service attack that uses ICMP and effects almost all network devices. The source code for the attack was published on Rootshell on 10/30/97 under the name “smurf.c”.

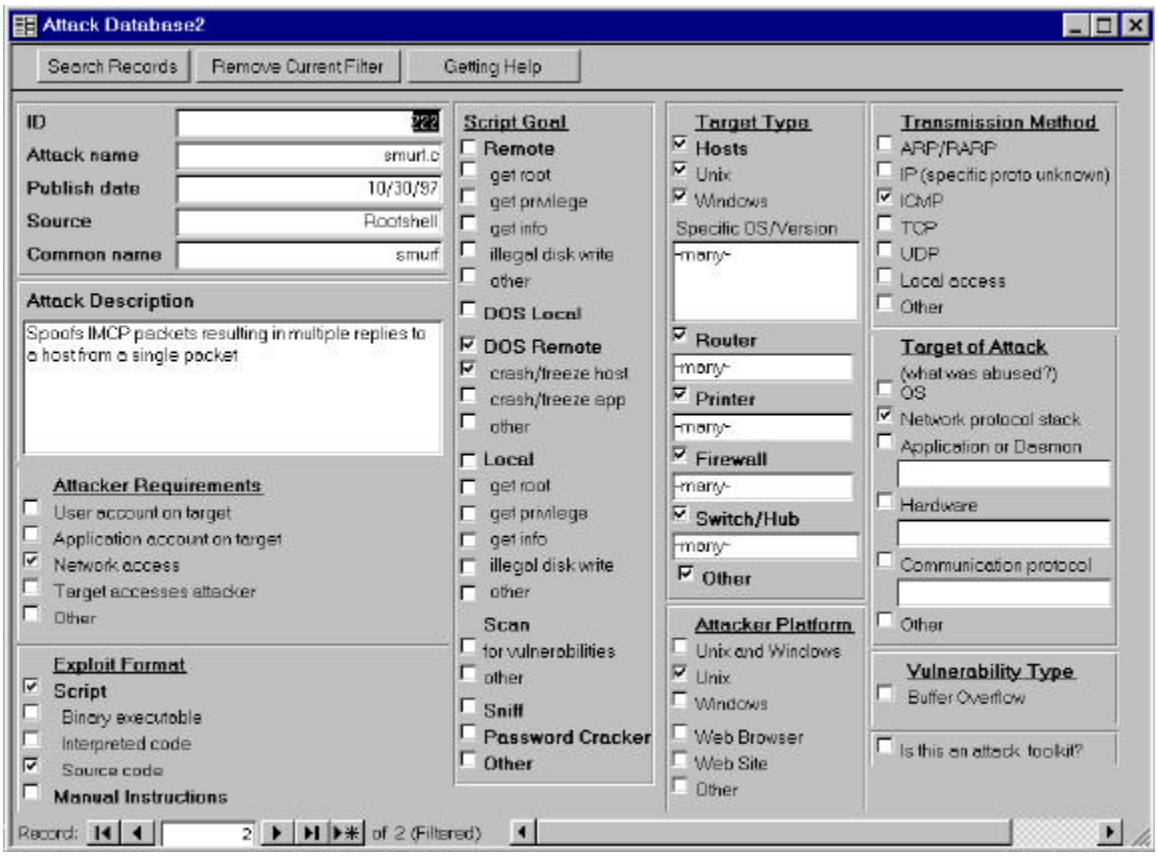

Figure 1: An Example I-CAT Attack Description Entry

Most of the fields in the attack characterization form are self explanatory, however we need to make a few clarifications:

1. The check boxes are completely dependent. Theoretically, a single attack could cause all check boxes to be marked.
2. We define an “attack toolkit” to be an attack script that exercises more than one vulnerability.
3. The check box “target accesses attacker” is for attacks where the attacker can only hurt the target if the target voluntarily visits the attacker’s host, application, or web site.
4. The “get root” check box is for any attack that gains complete control of the target host. This check box is thus applicable to both Windows and Unix machines.

3.2 Measuring the Popularity of Attacks

I-CAT also measures the popularity of attack scripts published on the Internet. This is done by measuring what people search for in the search engines of attack script web sites. Rootshell has a cgi script which allows one to see the last 50 search requests made to the attack database with a date and time stamp included. I-CAT harvests this information periodically and determines the popularity of attacks. While we wish that an attack script web site would publish what attacks people download, this information is not yet available and we are stuck using our indirect measurement method. Even though are measurements are indirect, we believe that they are significant. Our reasoning is that if one types the name of an attack into a search engine then they must be either planning to use the attack or else concerned that the attack will be used against them.

3.3 Measuring the Availability of Attacks

I-CAT also measures statistics on Internet attack publication. In this case, we measure not the attack scripts that people use but the population of attack scripts that is available for people to use. For example, we measure the percentage of attacks that are remote exploits, the percentage of attacks that use ICMP, and the percentage of attacks that compromise routers. I-CAT provides measurements on what is publicly available to the hacker community which is important for anyone trying to demonstrate the availability of attack scripts on the Internet.

4.0 Results

We now show how to use I-CAT to find attacks that meet user defined criteria. We give two examples in which we use I-CAT to answer the following two questions:

1. What is the set of attacks that meet a particular user-defined set of criteria?
2. What is the set of attacks that will penetrate a particular type of host?

Besides being useful for looking up attack scripts, I-CAT can be used to understand statistics on the popularity of attacks and statistics on trends in attack publication. To demonstrate this, we list the top 20 attack database search terms for December 1998 and give some details on the attacks found therein. Then, we give statistics on particular trends in attack publication for 1998.

4.1 Example Usage of I-CAT to Find Attacks that Meet Particular Criteria

Assume that we want to find attack scripts that enable a web site to attack its visitors. Furthermore, suppose that we want the subset of these attacks that gives the web server complete control over the victim. Finding these attacks on the Internet would be non-trivial but with I-CAT it is easy. In Figure 2, we show the I-CAT attack search screen. It allows the user to create arbitrary AND/OR expressions between all fields shown. In addition, it allows the user to perform regular expression matching in any of the text fields.

In Figure 2 we form the search query by checking the “Remote” box to indicate that we want remote penetration attacks. We check the “get root” box to indicate that we want only attacks that give complete control over the machine. We check the “target accesses attacker” box to indicate that we want only attacks that are launched when the target approaches the attacker in some way. We check the “web site” box under attacker platform to indicate that we want only attacks that can be launched from a web site. Last, we press the “apply filter” button to search the I-CAT database.

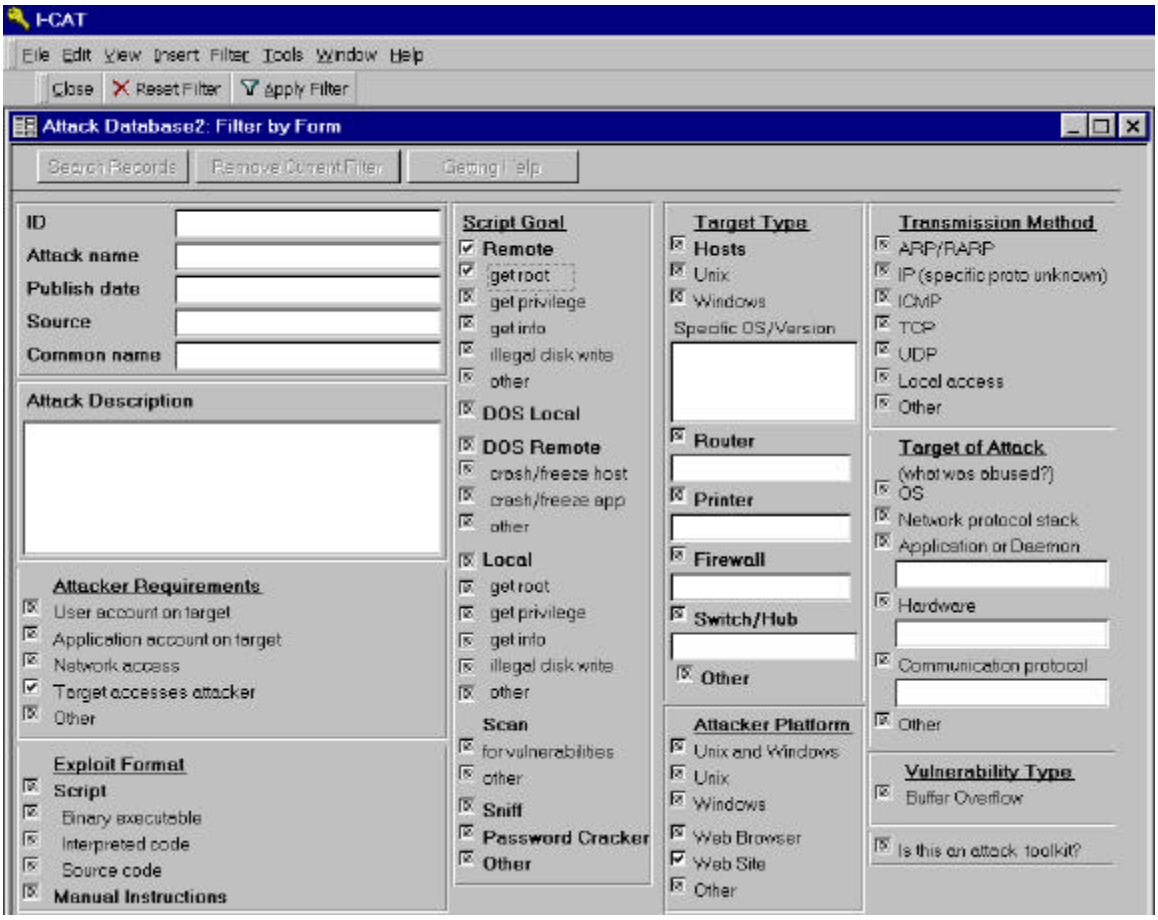

Figure 2: I-CAT Search Form, Example 1

I-CAT then returns a list of all attacks that meet the specified criteria. Figure 3 shows the first attack in this list. The actual attack scripts can easily be found since the location of the attack is pinpointed by the fields titled “source”, “publish date”, and “attack name”. In this case, a user would load the Rootshell web site and type “nsover.txt” into Rootshell’s search engine to retrieve the attack script.

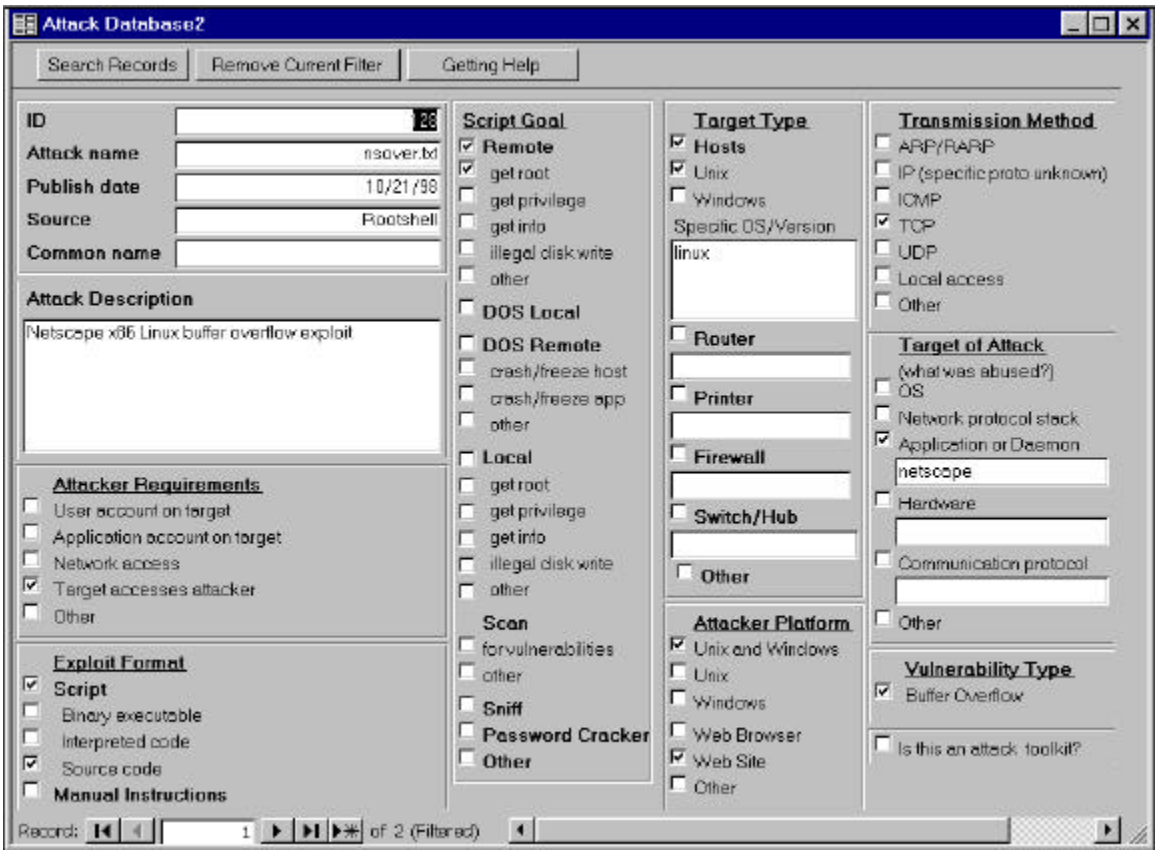

Figure 3: I-CAT Search Results, Example 1

4.2 Example Usage of I-CAT to Find Attacks that are Capable of Violating a Particular Host

Assume that we want to find the set of attacks that could penetrate one's multi-purpose mail server and web server. The first step is to write down the operating system that is running on the host and any applications that offer network services. A typical list might be as follows[*]:

Operating System: Linux
Applications: IMAP server, Apache web server, Telnet, FTP

For each application, we query I-CAT separately to see if there exists attacks that can remotely penetrate the host. However, it is possible to include all applications into the same query. In Figure 4, we query I-CAT for remote penetration attacks that effect Linux operating systems and violate the IMAP application. We check the "Remote" box to specify a remote penetration attack. We check the "Hosts" and "Unix" boxes to indicate that we want only attacks that work against Unix hosts. We perform regular expression matching on the "Specific OS/Version" to obtain attacks that work against Linux and attacks that work against many operating systems. We check the "Application or Daemon" box to indicate that the attack should violate an application. Lastly, we do a regular expression search on the "Application or Daemon" text box to find attacks that violate IMAP.

[*] Unix hosts have a services file which lists all network services offered by the host

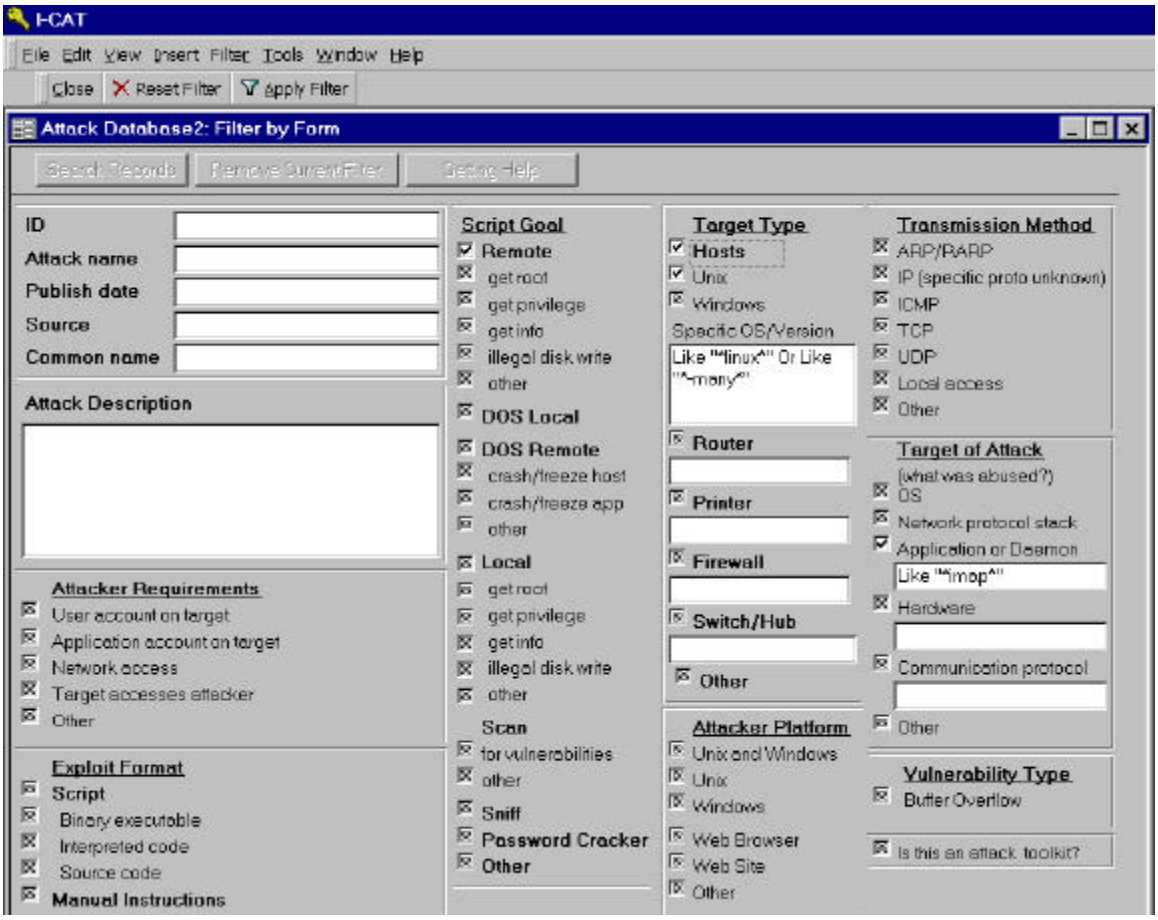

Figure 4: I-CAT Search Form, Example 2

After pressing the “Apply Filter” button, I-CAT returns the screen shown in Figure 5. The result is two different attack scripts that will remotely penetrate a Linux host running an IMAP server. A security expert would then look up the attack on the Internet using the fields “source”, “publish date”, and “attack name”. Afterwards, the I-CAT user formulates queries for the other applications running on the target host.

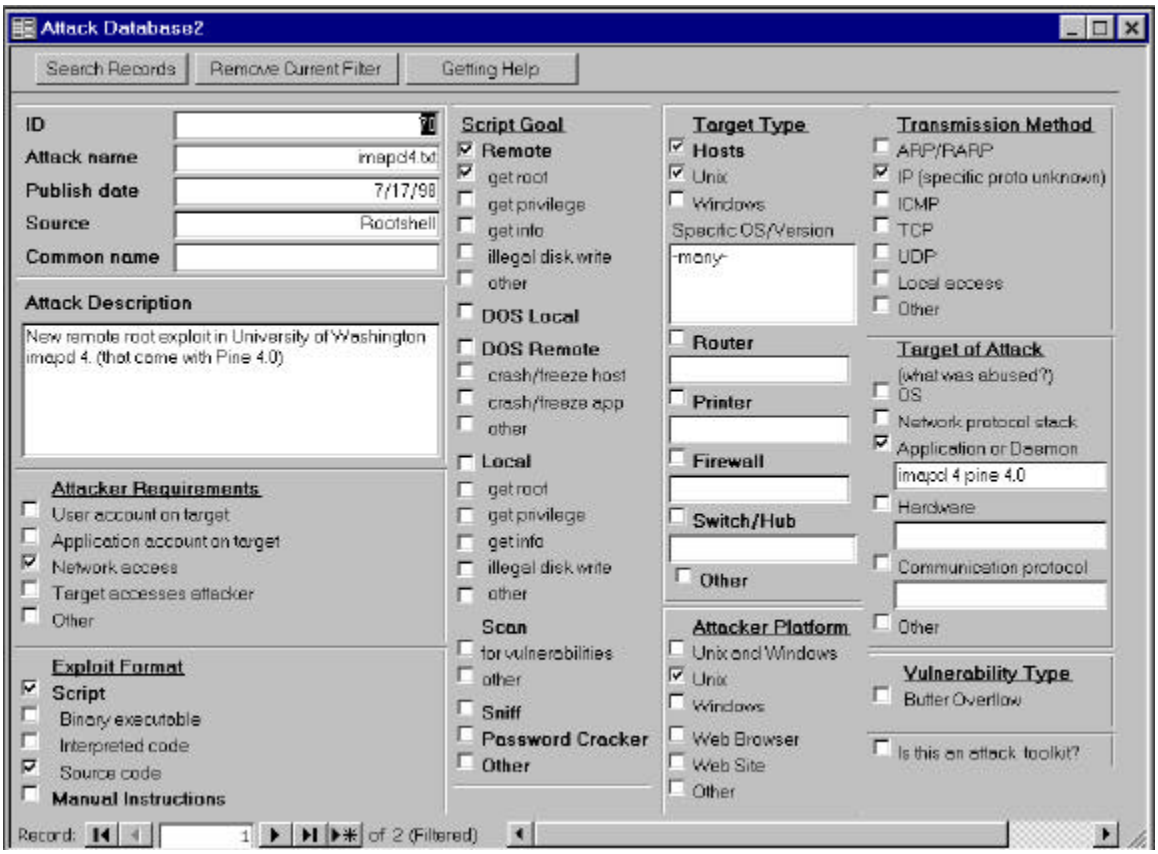

Figure 5: I-CAT Search Results, Example 2

4.3 Measurements on the Popularity of Attacks

Every month, I-CAT publishes the top 200 most requested search terms typed into the Rootshell attack database search engine. The top 200 terms represent around 60% of the sample set. Note that we are measuring the popularity of an attack, not the usage of an attack. Attack usage statistics can be obtained from incident response teams like CERT.

In December of 1998, 33407 queries were sampled which represent 20 percent of the queries made that month. I-CAT estimated by looking at the date and time stamp of each query that Rootshell received 170,000 queries in December. Below are the top 20 search terms for December 1998. Operating systems are italicized. Names of attacks are bolded. Names of applications are underlined.

1. *linux*	(2.29%)
2. *windows nt*	(2.25%)
3. *windows*	(1.45%)
4. icq	(1.40%)
5. sendmail	(1.40%)
6. **back orifice**	(1.36%)
7. **smurf**	(1.32%)
8. **teardrop**	(1.31%)
9. imap	(1.27%)
10. “ “	(1.17%)*
11. *solaris*	(1.10%)
12. *red hat*	(1.04%)

* 1.17 percent of the time the user of Rootshell pressed return in the search box without typing anything.

13. *windows 98* (0.91%)
14. **netbus** (0.81%)
15. **nuke** (0.75%)
16. scanner (0.75%)
17. *freebsd* (0.72%)
18. *irix* (0.69%)
19. **mscan** (0.68%)
20. **nestea** (0.66%)

Care must be taken in interpreting the search terms that are operating systems. People may be searching for attacks that are effective against a particular operating system. Alternatively, people may be searching for attacks that they can launch from a particular operating system. Since Linux is a very popular platform for which to write attack scripts, its prominence at the top of the list is not unexpected.

The three applications that appeared on the top 20 are ICQ, Sendmail, and IMAP. ICQ is an advanced chat program that lets Internet users find their friends when they are online. It is an important application because of its popularity and because America Online uses it. Unfortunately, ICQ has had multiple vulnerabilities published about it in 1998. The ICQ protocol was insecure in that it let anyone pretend to be any other user. By taking on the identity of people's friends, unethical people could convince others to receive and execute trojan horses. Sendmail is a very old but widely used application whose enormous number of features have caused it to be consistently insecure. In 1998 it has again proved to have multiple vulnerabilities. Sendmail is proof that vulnerabilities may not disappear over time because as soon as the current vulnerabilities are patched, developers release new features with new vulnerabilities. IMAP is a protocol that people use to retrieve their mail. Unfortunately, a recent version of an IMAP server contained buffer overflow problems which allow a remote user root access of the machine running the server.

Attacks that penetrate a particular application are typically named after the application. Attacks that have unique names are typically denial of service attacks, trojan horses, and scanners. The attacks listed by name on the top 20 list can be categorized as follows.

Trojan Horse:	Back Orifice, Netbus
Denial of Service:	Smurf, Teardrop, Nuke, Nestea
Scanners:	Mscan

Back Orifice:	Allows an attacker to remotely control a Windows 95 host
Net Bus:	Allows an attacker to remotely control a Windows NT host
Smurf:	Uses a network that accepts broadcast ping packets to flood the target with ping reply packets
Teardrop:	Freezes vulnerable Windows 95 and Linux hosts by exploiting a bug in the fragmented packet re-assembly routines
(Win)nuke:	Freezes a Windows 95 host by sending it out of band TCP data
Nestea:	Variant on teardrop that freezes windows and Linux hosts

4.4 Statistics on Internet Attack Publication

Very few statistics have been taken on the nature of the attacks that are available on the Internet. However, I-CAT offers users this type of hard data. I-CAT in a quantitative way can demonstrate the insecurity that abounds in almost all major operating systems and applications.

Below are some sample statistics taken on the 237 attacks characterized by I-CAT in 1998.

Statistic:	29% of attacks can launch from Windows hosts
Lesson:	One does not need to understand Unix to be dangerous anymore

Statistic:	20% of attacks are able to remotely penetrate network elements
Lesson:	Attacks that give remote users access to hosts are not rare
Statistic:	3% of the attacks were web sites attacking those who visited the site
Lesson:	Surfing the web is not a risk free activity
Statistic:	4% of attacks scan the Internet for vulnerable hosts
Lesson:	Automated tools that find hosts which are easily compromised abound
Statistic:	5% of attacks are effective against routers and firewalls
Lesson:	The Internet infrastructure components themselves are vulnerable to attack*

5.0 Future work

I-CAT currently consists of attacks only from Rootshell. We wish to expand I-CAT to use data from multiple attack script web sites. This will be especially useful from the standpoint of taking statistics on Internet attack publication. If other attack script web sites yield substantially different statistics then we will know that our statistics are the result of the publication bias of Rootshell and not representative of Internet attack publication.

We plan on following the advisories published by organizations like CERT and to provide a mapping between CERT advisories and the attack script database. Since CERT publishes advisories on only the most important attacks and vulnerabilities, our mapping will enable security companies to analyze the important attacks that exist in the hacker world.

We attempted to find trends in attack publication over time but since our database contains only a sixteen month history we were unable to conclusively identify trends. However, we plan to publish the trends that occur in attack publication over time as I-CAT encompasses more attacks over a larger time period.

While it has been interesting to measure for what people search on Rootshell, we need to find more measures of attack popularity. Ideal sources would be the attack download statistics from attack script web sites and the incident statistics from computer emergency response teams.

6.0 Conclusion

I-CAT has proven to be a useful tool that enables security researchers to understand the attacks published on the Internet. By providing lists of the most popular attacks, I-CAT enables security experts to focus on the greatest threats. By providing statistics on trends in Internet attack publication, I-CAT enables security experts to understand what types of attack scripts are available and in what quantity. While I-CAT is not a silver bullet to security experts attempting to find attack and vulnerability information, it is a tool that gives security experts a new window for viewing the world of our enemy.

7.0 References

[1] Rootshell, http://www.rootshell.com
[2] Fyodor's Playhouse, http://www.insecure.org
[3] The Legacy, http://www.jabukie.com/hacking.html
[4] Bugtraq, http://geek-girl.com/bugtraq
[5] NT Bugtraq, http://listserv.ntbugtraq.com/archives/ntbugtraq.html
[6] Carnegie Mellon University Software Engineering Institute's Computer Emergency Response Team (CERT), http://www.cert.org
[7] Federal Computer Incident Response Capability, http://www.fedcirc.gov
[8] Forum of Incident Response and Security Teams, http://www.first.org

* Most of these attacks are denial of service attacks and scanning attacks as opposed to penetration attacks.

[9] Department of Energy's Computer Incident Advisory Capability run out of Lawrence Livermore Laboratory, http://ciac.llnl.gov
[10] Australian Computer Emergency Response Team, http://www.auscert.org.au
[11] Infilsec, http://infilsec.com
[12] D. Alessandri, M. Dacier. *VulDa: A Vulnerability Database*. 2nd Workshop on Research with Security Vulnerability Databases, January 21-22, 1998 at Purdue University.