(IN)SECURE

OPEN. INFORMATIVE. TO THE POINT. Issue 4 - October 2005

STRUCTURED TRAFFIC ANALYSIS

by Richard Bejtlich, founder of TaoSecurity

PDA ATTACKS, PART 2: AIRBORNE VIRUSES - EVOLUTION OF THE LATEST THREATS

by Cyrus Peikari, M.D., CEO of Airscanner Mobile Security

TABLE OF CONTENTS

Welcome to (IN)SECURE 1.4 the digital security magazine

Welcome to yet another issue of (IN)SECURE. The book contest we held in the previous issue was a great success and here are the winners: Dominic White, Dr. Gary Hinson, Ronaldo Vasconcellos, Joey Ortiz, Adrian St. Onge and Frantisek Holop.

To all of you that sent us insightful comments we thank you, (IN)SECURE will grow to be better because of it. Expect more contests in the future and keep that feedback running, there's always place for us to improve. If you're interested in writing for (IN)SECURE feel free to contact us, we take a look at all submissions and ideas.

Enjoy the magazine!

Mirko Zorz
Chief Editor

Visit the magazine website at www.insecuremag.com

(IN)SECURE Magazine contacts

Feedback and contributions: Mirko Zorz, Chief Editor – editor@insecuremag.com

Advertising and marketing: Berislav Kucan, Director of Marketing – marketing@insecuremag.com

Distribution

(IN)SECURE Magazine can be freely distributed in the form of the original, non modified PDF document. Distribution of substantively modified versions of (IN)SECURE Magazine content is prohibited without the explicit permission from the editor. For reprinting information please send an email to reprint@insecuremag.com or send a fax to 1–866–420–2598.

Corporate security news

Secure Mobile Access Solution for PDAs and Smart Phones

Aventail announced Aventail Mobile, the market's most complete SSL VPN for mobile devices, providing easy access to critical applications from virtually any mobile device with complete security and granular access control. Aventail Mobile supports all major operating systems, including Blackberry, Palm, Windows Mobile, DoCoMo, and Symbian.

The Aventail Secure Mobile Access Solution will be available in Q4 2005. WorkPlace Mobile will be included in the overall Aventail Smart SSL VPN platform. Connect Mobile is an add-on feature priced from $995.00.

(ISC)2 Information Security Scholarship For 2006

The International Information Systems Security Certification Consortium, Inc. announced it is accepting applications for the Annual (ISC) ^{2} Security Scholarship Program for 2006, which offers $50,000 in financial assistance to post-graduate students who are conducting information security research projects.

One-year scholarships of up to $12,500 apiece will be awarded to four qualifying full-time post-graduate students pursuing an advanced degree in information security at any accredited university worldwide. The scholarships may be consecutively renewed if all criteria have been met, or for multiple research projects. The deadline for submission is Nov. 30, 2005.

To obtain the (ISC) ^{2} Information Security Scholarship form, prospective candidates should go to www.isc2.org/scholarship.

Symantec Completes Acquisition of WholeSecurity

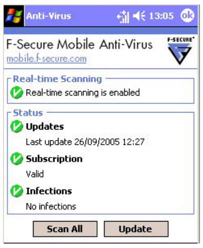
Symantec Corp. announced the completion of its acquisition of WholeSecurity, Inc., a leading provider of behavior-based security and anti-phishing technology. WholeSecurity solutions protect PCs from threats such as worms, Trojan horses, keystroke loggers, and phishing attacks. WholeSecurity's products leverage behavioral technology to protect users from these threats, whether they are known or unknown, on both managed and unmanaged PCs. The acquisition was announced on Sept. 22, 2005.

Survey Shows that Companies Don’t Secure Data “On the Move”

Senforce Technologies Inc. announced the findings of research it conducted by surveying 56 public and private sector organizations. The Senforce survey revealed that while 87 percent of critical business data is found on endpoint PCs, 56 percent of those asked think their current wireless network security strategy is reactive or inadequate. Other key survey findings include:

- 82 percent of new PC procurements are notebooks versus desktop PCs.
- 74 percent of those notebooks are wireless-enabled.
- 92 percent are concerned about notebooks moving in and out of the network perimeter.
- 43 percent have deployed production wireless networking infrastructure with defined policies.
- 63 percent prefer non-proprietary wireless networking hardware solutions.

F-Secure Mobile Anti-Virus Extended to Cover Windows Mobile OS



F-Secure Mobile Anti-Virus
mobile.f-secure.com

Real-time Scanning
✓ Real-time scanning is enabled

Status
✓ **Updates**
Last update 26/09/2005 12:27
✓ **Subscription**
Valid
✓ **Infections**
No infections

F-Secure Corporation announced that it is to extend the support for its award-winning F-Secure Mobile Anti-Virus to Microsoft’s Windows Mobile operating system. The solution brings new levels of protection for Windows Mobile users.

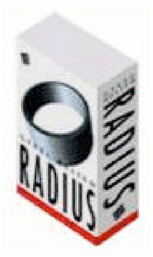
F-Secure Mobile Anti-Virus and F-Secure Mobile Security can be purchased from selected mobile operators or directly from the F-Secure eStore at: http://www.f-secure.com/estore/. A free 30-day trial is also available in the mobile-device optimized F-Secure Mobile Portal: mobile.f-secure.com.

Study Reveals Spyware Encounters are Increasing at Work

Trend Micro, Inc. announced key findings from a study that reveals that more than 87 percent of corporate end users are aware of spyware, and yet 53 percent of survey respondents demand greater education from IT to better understand the threat. The findings indicate that awareness does not translate to knowledge, and as a result users are looking to their IT departments to play a more protective role. Some of the findings include:

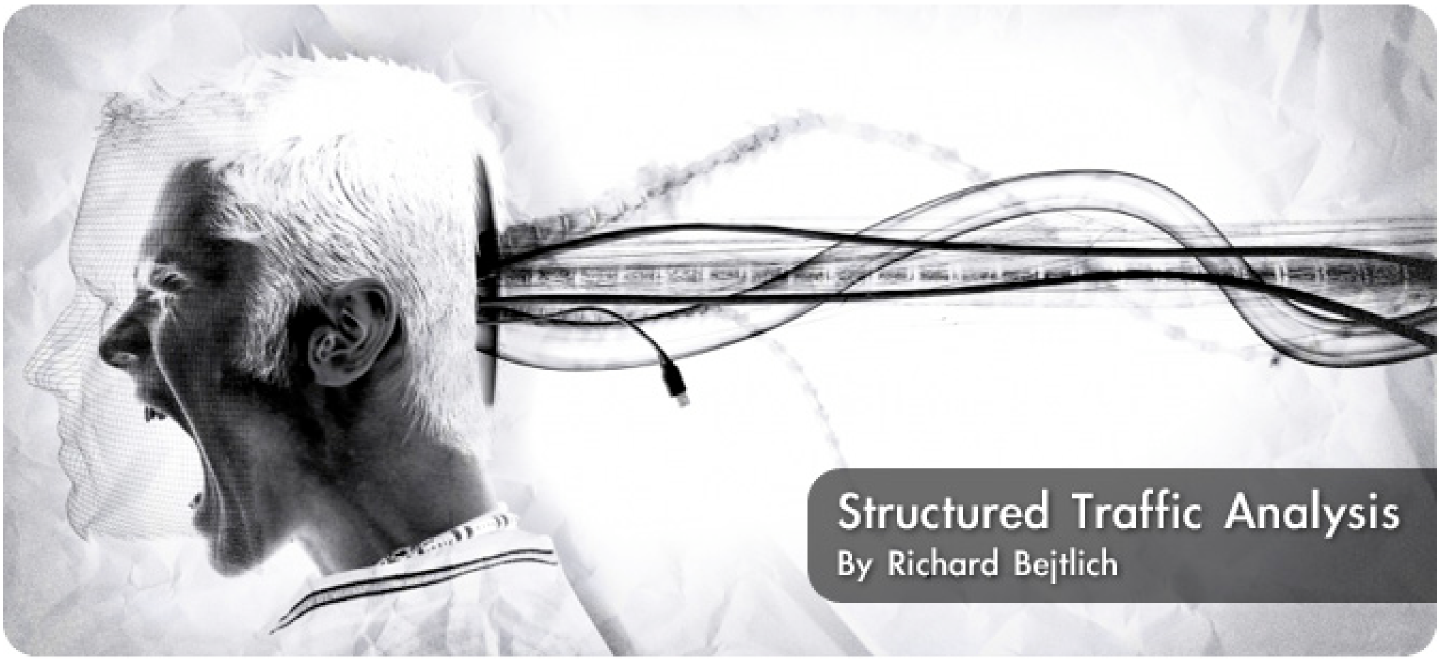
- Viruses and Spyware are perceived as being more serious threats to corporate security than spam.
- 26% American SMB workers, and 21% American enterprise workers stated that they had fallen victim to spyware while at work.
- Among U.S. based respondents, the top five consequences of being victimized by spyware were lower computer performance, malicious downloads, violation of privacy and more.

Funk Software Ships Endpoint Integrity Solution Based On TCG’s Trusted Network Connect Standards



Funk Software announced general availability of its suite of endpoint integrity products, including new Steel-Belted Radius/Endpoint Assurance and a new version of its secure, low-overhead 802.1X supplicant Odyssey Client.

The client lets enterprises take a proactive approach to network security and significantly strengthen the protection of their networks – using a standards-based approach that permits interoperability across a wide variety of vendor offerings and ensures scalability to future technologies.



Structured Traffic Analysis

By Richard Bejtlich

One day while working on the latest security problem, a member of the network administration team walks toward your desk carrying a CD-ROM. She drops the disc in front of you and asks “Would you mind looking at this traffic? Joe thinks there might be a security incident here.”

You respond by casually slipping the CD-ROM into your laptop, responding “No problem! I’ll get right on that.” As the network administrator walks away, you load the 2 GB trace file into Ethereal. After listening to your hard drive groan and thrash, you eventually see millions of packets staring at you. You stare back and wonder “Now what?”

If this scenario sounds familiar, you are not alone. Thousands of network analysts around the world think packet analysis begins and ends with Ethereal. The open source protocol analyzer is indeed an excellent tool, but it isn’t necessarily the program with which one should begin a network investigation. This is especially the case when performing a zero-knowledge assessment, where the analyst is given little or no information regarding the enterprise reflected by the network trace.

The purpose of this article is to introduce Structured Traffic Analysis (STA). STA is a top-down approach to examining network traces that builds upon the author’s work on Network Security Monitoring (NSM) [1]. STA has been successfully used to analyze traces for NSM, network incident response (NIR), and network forensics (NF). STA is not by itself sufficient to perform NSM, NIR, or NF, but the STA methodology applies any time an analyst must make sense of a network trace. After reading this article, you may share the sentiments of a student in one of the author’s recent classes who said “I’m embarrassed I ever used Ethereal to start network analysis!”

STA consists of the following steps. The process assumes you are an analyst supporting a client, although the methodology applies to traffic you collect yourself. The four forms of NSM data – statistical, session, full content, and alert – each play a role.

1. Make a new directory on the analysis platform specifically for the investigation at hand.
2. Copy the trace into the analysis directory and change the trace permissions to read-only.
3. Hash the trace and copy the hash elsewhere.
4. (optional) Run Capinfos on the trace to acquire initial statistical data.
5. Run Tcpdstat on the trace to obtain basic statistical data.
6. Run Argus on the trace to extract session data.
7. (optional) Run Ragator on the Argus file to collapse redundant session records.
8. Run Racount on the Argus file to count session records.
9. Run Rahosts on the Argus file to list all IP addresses.
10. Run Ra on the Argus file to enumerate source IP, destination IP, and destination port combinations.
11. Run Ra on the Argus file to observe session data directly.
12. Run Tcpflow on the trace to rebuild full content data of interest.
13. (optional) Run Snort on the trace to generate alert data.

The next section will explain each of these steps in detail. In this example, the trace provided by the client is called **sample**.

1. Make a new directory on the analysis platform specifically for the investigation at hand.

```
$ mkdir 2005-041-santini_air
$ cd 2005-041-santini_air
```

Create a new directory for every trace to be analyzed. Name the directory using a convention that makes sense for you, such as YYYY-MM-DD-CASE_NUMBER-CLIENT_NAME.

Do not rename the trace file.

2. Copy the trace into the analysis directory and change the trace permissions to read-only.

```
$ cp /cdrom/sample /home/analyst/2005-041-santini_air

$ ls -al
total 96
drwxr-xr-x  2 analyst  analyst    512 Jul 17 16:00 .
drwxr-xr-x  5 analyst  analyst   1024 Jul 17 16:03 ..
-rwxr-xr-x  1 analyst  analyst  93506 Jul 16 20:38 sample
$ chmod 444 sample
$ ls -al
total 96
drwxr-xr-x  2 analyst  analyst    512 Jul 17 16:00 .
drwxr-xr-x  5 analyst  analyst   1024 Jul 17 16:03 ..
-r--r--r--  1 analyst  analyst  93506 Jul 16 20:38 sample
```

Never analyze original copies of traces unless those traces are stored on read-only media. By changing the permissions of a copied trace to read-only, the analyst guards against simple mistakes that could damage the trace or alter the results of the investigation.

3. Hash the trace and copy the hash elsewhere.

```
$ sha256 sample > sample.sha256
$ cat sample.sha256
SHA256 (sample) =
1a6da6a2a849eb27fb7522939afab63ec59bcdb9412c2460fe611543b573d95f
$ cp sample.sha256 /home/analyst/hashes/
```

Here we use the FreeBSD tool sha256, since problems with MD5 and SHA1 have been reported[2]. By generating a hash and storing it elsewhere, other analysts can be sure they are inspecting the same trace should a second investigation be required. Hashes also ensure the integrity of the data being inspected; they reveal evidence of tampering or corruption when run on modified files.

4. (optional) Run Capinfos on the trace to acquire initial statistical data.

```
$ capinfos sample > sample.capinfos
File name: sample
File type: libpcap (tcpdump, Ethereal, etc.)
Number of packets: 1194
File size: 93506 bytes
Data size: 213308 bytes
Capture duration: 342.141581 seconds
Start time: Thu Jun 23 14:55:18 2005
End time: Thu Jun 23 15:01:01 2005
Data rate: 623.45 bytes/s
Data rate: 4987.60 bits/s
Average packet size: 178.65 bytes
```

This step is optional because much of the same data is obtained using Tcpdstat below. However, Capinfos is packaged with Ethereal, and hence available on Windows. It is good to have Capinfos available as an alternative tool. The most important aspects of Capinfos statistical data include trace start and end times, the number of packets in the trace, and the size of the trace. This data helps an analyst gain an initial sense of what data is available.

5. Run Tcpdstat on the trace to obtain basic statistical data.

```
$ tcpdstat sample > sample.tcpdstat
$ cat sample.tcpdstat
DumpFile:  sample
FileSize: 0.09MB
Id: 200506231455
StartTime: Thu Jun 23 14:55:18 2005
EndTime:   Thu Jun 23 15:01:01 2005
TotalTime: 342.14 seconds
TotalCapSize: 0.07MB  CapLen: 68 bytes
# of packets: 1194 (208.31KB)
AvgRate: 5.08Kbps  stddev:30.22K
### IP flow (unique src/dst pair) Information ###
# of flows: 66  (avg. 18.09 pkts/flow)
Top 10 big flow size (bytes/total in %):
 20.0% 16.3% 15.7% 12.9%  4.8%  4.0%  2.9%  1.3%  1.3%  1.2%
### IP address Information ###
# of IPv4 addresses: 68
Top 10 bandwidth usage (bytes/total in %):
 69.9% 21.5% 18.5% 17.5% 16.9% 13.9%  5.4%  5.2%  4.5%  4.3%
# of IPv6 addresses: 4
Top 10 bandwidth usage (bytes/total in %):
 81.5% 59.2% 40.8% 18.5%
### Packet Size Distribution (including MAC headers) ###
<<<<
 [   32-   63]:        857
 [   64-  127]:        104
 [  128-  255]:         79
 [  256-  511]:         61
 [  512- 1023]:         14
 [ 1024- 2047]:         79
### Protocol Breakdown ###
<<<<
     protocol          packets         bytes           bytes/pkt
----------------------------------------------------------------
[0] total            1194 (100.00%)   213308 (100.00%)    178.65
[1] ip                988 ( 82.75%)   198381 ( 93.00%)    200.79
[2]  tcp              884 ( 74.04%)   180408 ( 84.58%)    204.08
[3]   http(s)         219 ( 18.34%)   124825 ( 58.52%)    569.98
[3]   other           665 ( 55.70%)    55583 ( 26.06%)     83.58
[2]  udp               94 (  7.87%)    17247 (  8.09%)    183.48
[3]   dns               9 (  0.75%)     2752 (  1.29%)    305.78
[3]   other            85 (  7.12%)    14495 (  6.80%)    170.53
[2]  icmp               7 (  0.59%)      546 (  0.26%)     78.00
[2]  igmp               3 (  0.25%)      180 (  0.08%)     60.00
[1] ip6                 5 (  0.42%)      422 (  0.20%)     84.40
[2]  icmp6              5 (  0.42%)      422 (  0.20%)     84.40
>>>>
```

Dave Dittrich's Tcpdstat is helpful because the statistical data it provides gives the first real insights to the nature of the trace[3]. In addition to the data also available in Capinfos, Tcpdstat provides a useful Protocol Breakdown. The number of protocols recognized by Tcpdstat is relatively small, and the breakdown is based on ports and IP protocols. Nevertheless, irregular patterns can often be detected. Notice that the sample trace features almost 56% unrecognized TCP traffic. Details such as these give an analyst a starting point for subsequent inspection.

6. Run Argus on the trace to extract session data.

```
$ argus -r sample -w sample.argus
```

This step provides session data using Carter Bullard's Argus[4]. The analyst does not yet look at the session records. That will happen in step 11, once sessions of interest can be more easily identified. The -w flag writes session records to the specified file in Argus format. Only tools packaged with Argus can read its output.

7. (optional) Run Ragator on the Argus file to collapse redundant session records.

```
$ ragator -r sample.argus -w sample.argus.ragator
```

When a particularly large trace is being analyzed, it may be helpful to collapse redundant session records using the Ragator program packaged with Argus. The Argus server generates multiple entries for longer sessions. Ragator will combine these into a single entry. In the following steps the analyst can run the various "Ra" tools against either the **sample.argus** or **sample.argus.ragator** file. For the sake of consistency, examples that follow inspect the **sample.argus** file.

8. Run Racount on the Argus file to count session records.

```
$ racount -ar sample.argus
racount  recs  tot_pkts  src_pkts  dst_pkts  tot_byt  src_byt  dst_byt
    tcp    50       884       634       250   178162    49818   128344
    udp    46        94        94         0    17237    17237        0
   icmp     6         7         7         0      546      546        0
     ip     3         3         3         0      126      126        0
    arp   159       176       175         1    10560    10500       60
 non-ip    27        30        30         0     4367     4367        0
    sum   293      1194       943       251   210998    82594   128404
```

The Racount utility breaks down the number of session records by protocol. This session metadata helps the analyst get a better sense of the contents of the trace, again without looking at a single packet. The -a flag tells Racount to show all protocols.

9. Run Rahosts on the Argus file to list all IP addresses.

```
$ rahosts -n -r sample.argus > sample.argus.rahosts
$ wc -l sample.argus.rahosts
     129 sample.argus.rahosts
$ cat sample.argus.rahosts
0.0.0.0
0.43.224.0
1.2.170.51
4.0.0.0
4.0.255.255
10.10.3.11
15.0.56.247
15.0.246.214
...edited...
216.74.132.13
255.255.255.255
```

The Rahosts utility lists all of the IP addresses seen in an Argus file. This process helps the analyst get a grip on the scope of the investigation by seeing a summary of all IP addresses present in a trace.

The -n flag disables hostname resolution.

10. Run Ra on the Argus file to enumerate source IP, destination IP, and destination port combinations.

```
$ ra -nn -r sample.argus -s saddr daddr dport proto
 | sort -n -t . -k 1,1 -k 2,2 -k 3,3 -k 4,4 | uniq -c >
 sample.argus.saddr-daddr-dport-proto

$ cat sample.argus.saddr-daddr-dport-proto
   1         0.0.0.0 255.255.255.255.67     udp
   4  0:4:0:92:90:da ff:ff:ff:ff:ff:        0
   1  0:4:0:92:90:da ff:ff:ff:ff:ff:     3307
...edited...
   1   1.2.170.51   16.74.132.12.80     tcp
   1   1.2.170.51   16.74.132.13.80     tcp
   7   1.2.170.51   6.5.161.250.80     tcp
   1   1.2.170.51   216.74.132.13.21     tcp
...truncated...
```

This first invocation of Ra helps the analyst understand what sources are talking to what destinations, using what protocols and services[5]. Source ports are ignored at this stage of the investigation. Often this session data is sufficient to identify suspicious activity[6]. Here the analyst wishes to know why 1.2.170.51 decides to connect to 216.74.132.13 using File Transfer Protocol[7].

Identifying suspicious outbound connections is one way to perform extrusion detection, which is the process of identifying unauthorized activity by inspecting outbound network traffic[8].

11. Run Ra on the Argus file to observe session data directly.

```
$ ra -nn -L0 -A -Z b -r sample.argus  -A -Z b - host 216.74.132.13

    StartTime     Type  SrcAddr  Sport Dir   DstAddr     Dport
 SrcPkt    DstPkt    SAppBytes    DAppBytes    Status
23 Jun 05 15:00:40 tcp  1.2.70.51.49129  ->   216.74.132.13.21
 4        2         0            0           FSA_FSA
23 Jun 05 15:00:41 tcp  1.2.70.51.42939  ->   216.74.132.13.21
 13       10        4207         946         FSRPA_FSPA
```

If the analyst identifies one or more interesting sessions in step 10, he may wish to run Ra with one or more host or network filters. In the example above, the analyst decides to look for traffic involving host 216.74.132.13. The Ra switches have the following meaning:

- -nn disables resolution of IPs to hostnames and port numbers to words.
- -L0 prints column headers.
- -A prints counts of application data passed by each party. The default invocation of Ra counts packet header bytes as data.
- -A Z tells Argus to show TCP flag records.
- – host 216.74.132.12 is a host filter. Alternatively, a network filter like – net 216.74 could have been applied.

12. Run Tcpflow on the trace to rebuild full content data of interest.

```
$ tcpflow -r sample -c port 21

DST: 220 Serv-U FTP Server v5.1 for WinSock ready...
SRC: USER example
DST: 331 User name okay, need password.
SRC: PASS example
DST: 230 User logged in, proceed.
```

```
SRC: SYST
DST: 215 UNIX Type: L8
SRC: FEAT
DST: 211-Extension supported
DST: AUTH TLS
...truncated...
```

Jeremy Elson's Tcpflow rebuilds TCP sessions[9]. When used with the -c switch, it sends the results to standard output. When the -c switch is omitted, Tcpflow rebuilds each side of the session into individual files in the analysis directory. For initial investigation, the author prefers sending results to the screen.-The sample here shows an FTP session. The DST is the FTP server and the SRC is the FTP client.

13. (optional) Run Snort on the trace to generate alert data.

```
$ snort -r sample -c /usr/local/etc/snort.conf -l . -b
```

Marty Roesch's Snort can generate alert data as a fast way to identify low-hanging security fruit. Suspicious traffic identified by Snort can be examined using session data in step 11 or full content data in step 12.

An analyst who must perform zero-knowledge assessments of network traces will be well-equipped to understand network traffic after completing the STA process. The analyst will have gathered statistical, session, full content, and possibly alert data on the trace. Session data often reveals conversations that demand additional attention, and TCP-based flows can be rebuilt using Tcpflow. Should any sessions require direct packet-by-packet analysis, Ethereal is a good choice. The trace or an excerpt can be loaded into Ethereal for examination of the specific packets of interest. Had the analyst simply started with Ethereal, he may have had little or no idea where to begin his investigation.

STA forms the backbone of the investigation process the author uses for network forensics, but that is a topic for another article[10].

Richard Bejtlich (richard@taosecurity.com) is founder of TaoSecurity (www.taosecurity.com) and the TaoSecurity blog (taosecurity.blogspot.com), author of The Tao of Network Security Monitoring and Extrusion Detection, and co-author of Real Digital Forensics.

References

(1) The Tao of Network Security Monitoring: Beyond Intrusion Detection by Richard Bejtlich

(2) www.freshports.org/sysutils/freebsd-sha256/

(3) staff.washington.edu/dittrich/talks/core02/tools/tools.html

(4) www.qosient.com/argus

(5) Thanks to Paul Heinlein for his IP address sorting syntax at http://www.madboa.com/geek/sort-addr/.

(6) In a recent trace the author discovered thousands of sessions to port 3789 TCP. These connections represented traffic to a back door.

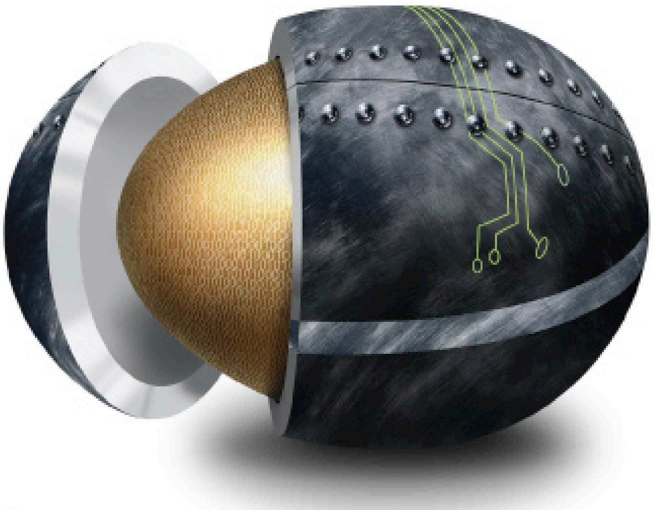
(7) This is only an example for this article. One could imagine any sort of anomalous outbound or inbound session prompting an analyst to desire a closer look.

(8) Extrusion Detection: Security Monitoring for Internal Intrusions by Richard Bejtlich.

(9) www.circlemud.org/~jelson/software/tcpflow/

(10) Real Digital Forensics: Computer Security and Incident Response by Keith Jones, Richard Bejtlich, and Curtis Rose.

FREE Exhibit Hall Admission!
www.infosecurityevent.com/mitigaterisk



Protecting your assets is more challenging than ever.

Attend infosecurity NEW YORK

December 6-8, 2005
Jacob K. Javits
Convention Center

FREE KEYNOTES

Tom Ridge, Opening Keynote
Former Secretary, US Department of Homeland Security
Wednesday
December 7th • 11:30 am

Bruce Schneier, Luncheon Keynote
Founder & CTO, Counterpane
Thursday
December 8th • 11:30 am

Honorary Conference Chairman

Howard A. Schmidt
Chief Security Strategist, US CERT
President & CEO, R & H Security Consulting LLC and Former White House Cyber Security Advisor

Infosecurity delivers information, education, and networking for a more secure and compliant infrastructure

The stakes have been raised. Your job of protecting your business and its information assets continues to increase in difficulty. Protecting against outside threats like hackers, spyware and targeted Trojans is now coupled with the need to be current on the latest regulations and compliance issues, guarding against data leakage, and the ever present insider threat. Infosecurity is geared to assist you in maintaining the confidentiality, availability, integrity compliance and risk mitigation of your organization.

Conference Tracks

Security Leadership Conference Series:
- T1: Security Management
- T2: Technical
- T3: Emerging Threats
- T4: Defense-in-Depth

T5: Wireless & Mobility Security – NYMISSA
T6: Compliance & Governance – Concordant, Inc.
T7: Privacy: Issues for Clients, Customers and Your Organization – IAPP
T8: Best Practices in Achieving Financial Justification – Larstan's Black Book Series

CISSPs/SSCPs Earn Up To 18* CPEs

Only (ISC)²®, the leader dedicated to educating, qualifying and certifying information security professionals and Infosecurity, the global leader in information security events, can offer such a high caliber education program. The CISSP® and SSCP® credentials identify you as an individual capable of developing and implementing solid information security practices.

**Combination of full conference and pre-conference workshops*

Real Risks – Real Solutions

Information security technology is complex and ever-changing to meet continued threats. Visit with over 150 leading vendors at Infosecurity to evaluate competing and complementary products and services such as Access Control/Authorization, Assessment & Audit, Authentication, Content Filtering, Encryption, Network Security, Perimeter Security, Security Management Products, Storage and more.

REGISTER TODAY AND SAVE! For early-bird conference discounts and free exhibition admission, register online before October 18 at **www.infosecurityevent.com/mitigaterisk**

Premier Education Sponsor: (ISC)²
Exclusive Global Media Sponsor: SC Magazine
Premier Media Sponsor: CSO
Silver Sponsors: PatchLink, Citadel Security Software, Websense, AirMagnet, Network Intelligence, nCircle, Secured eMail, TriGeo Network Security, Juniper Networks
Media Sponsors: CIO Insight, eWeek.com, Baseline, Security IT Hub
Bronze Sponsors: nitrosecurity, Sandstorm Enterprises, SecureWave, radware
Education Sponsors: ISSA New York Metro, Concordant, The Black Book Series, iapp
Technical Track Sponsor: ICSA Labs
Produced and Managed by: Reed Exhibitions

Access control lists in Tiger and Tiger Server: true permission management

By Matt Willmore

Are Access Control Lists the Next Big Thing for Apple's new Tiger Server? Matt Willmore explains ACL's and how they can benefit OS X Server admins and regular users alike.

Access Control Lists (ACLs) are an improved way of managing access to objects and directories in a file system. Instead of the traditional UNIX-style approach of read/write/execute, ACL's give administrators an unprecedented amount of control over how file and directory permissions are managed. ACLs have been existent in server Operating Systems such as Windows and OpenVMS for some time; this is one area where Apple is playing catch up. Luckily, Apple has done a thorough job of adding ACL support without any headaches on the part of the user.

Why Should I Care?

True File Access Management

ACLs bring about fine-grained permissions for objects and directories, something that OS X has been sorely lacking up to now. Without ACLs, server administrators are limited to the standard UNIX permissions: one owner, one user and allowance/denial to read, write and execute objects and directories. While these permissions are often adequate for a single user, they rarely fit the needs of an organization. ACLs allow admins to more accurately recreate the structure of an organization when defining permissions.

Stored in Metadata

Coinciding with the release of support for extensible metadata in Tiger, ACEs (access control entries) are stored in that very fashion. Unfortunately, Apple has chosen to not allow tools like xattr, which can view and edit that metadata, to see ACE entries. According to John Siracusa's review of Tiger, they're most likely prefixed with the reserved system namespace and therefore not editable with `xattr`.

Using ACL's with Tiger

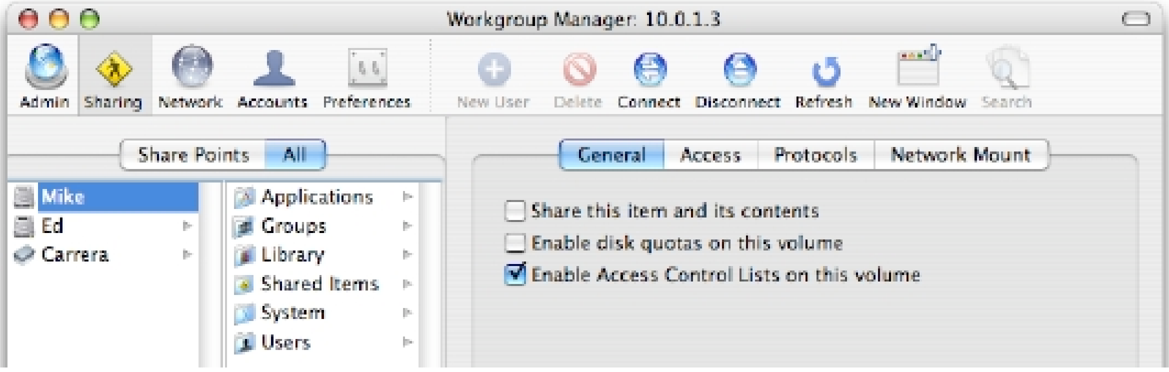
Unlike some file systems that require you to reformat the drive to support ACLs, with Tiger you can turn ACL support on or off with a simple command. Besides the obvious ease of use, one immediate benefit is that if you ever get yourself stuck — say, a file that cannot be deleted by root, which you can certainly do — you can just turn ACLs off, and Tiger will go back to observing the standard UNIX permissions. The only requirement that Apple makes — besides Tiger, of course — is that the volume be formatted as HFS+. Since ACEs are stored in extended attributes (see "Stored in Metadata" above) Apple needs an HFS+-formatted drive to store the entries.

To enable ACLs on the client (non-server) version of Tiger, you'll have to use **fsaclctl**. To enable ACL support on the boot volume, for example, you would use **fsaclctl** as follows:

```
sudo fsaclctl -p / -e
```

Because the command affects a system-level property of the file system sudo will be required. The **-p** flag states the path of the mount point (in this case, **/**) and the **-e** flag tells **fsaclctl** to enable support. Additionally, you can mirror the command on all HFS+ volumes at once by adding the **-a** flag. To disable support for a volume, substitue **-d** for **-e**.

In Tiger Server, you have the added benefit of a GUI front for enabling or disabling ACL support on each volume.



☐ Share this item and its contents
☐ Enable disk quotas on this volume
☑ Enable Access Control Lists on this volume

Apple has also added an API for integrating ACLs into programming for OS X. Details are available on the **acl** manpage. Also, note that API calls ending in **_np** indicate that the routine is non-portable; that is, it differs from the standard POSIX.1e library for ACL's. (Apple's implementation of ACL is based on the POSIX 1003.1e draft and is extensible for future improvements and additions.)

Before You Go Nuts…

Once you have ACL support enabled for a particular volume, we'll use **chmod** to change ACL entries on Tiger. Don't immediately go nuts and start changing everything you can find; rather, let's first look at what attributes can be set with a particular ACE. The **manpage** for **chmod** lists 17 distinct attributes, separated into sections:

File System Objects

- delete: Delete the item. Deletion may be granted by either this permission on an object or the delete_child right on the containing directory.
- readattr: Read an objects basic attributes. This is implicitly granted if the object can be looked up and not explicitly denied.
- writeattr: Write an object's basic attributes.
- readextattr: Read extended attributes.
- writeextattr: Write extended attributes.
- readsecurity: Read an object's extended security information (ACL).
- writesecurity: Write an object's security information (ownership, mode, ACL).
- chown: Change an object's ownership.

Directories

- list: List entries.
- search: Look up files by name.
- add_file: Add a file.
- add_subdirectory: Add a subdirectory.
- delete_child: Delete a contained object.

Files

- read: Open for reading.
- write: Open for writing.
- append: Open for writing, but in a fashion that only allows writes into areas of the file not previously written.
- execute: Execute the file as a script or program.

Interitance (Directories Only)

- file_inherit: Inherit to files.
- directory_inherit: Inherit to directories.
- limit_inherit: This flag is only relevant to entries inherited by subdirectories; it causes the directory_inherit flag to be cleared in the entry that is inherited, preventing further nested subdirectories from also inheriting the entry.
- only_inherit: The entry is inherited by created items but not considered when processing the ACL.

Let’s Try This Out

To start working with ACLs, let’s take a simple approach by working on files used just for practice. Create a new empty file with **touch**, and look at the permissions with **ls -al**. You can see that the user logged in is the owner, (in my case) the user is also the group and the permissions are set to 644 (owner [me] can read/write, group and everyone can both read). Without modifying the standard UNIX permissions, we’ll change it so the group “admin” has access to write to and append the file but not read it. Accomplish the first part with this command:

```
chmod +a "admin allow write,append"
file
```

Use **ls -le** (the **-e** flag instructs ls to print the ACL associated with each file that has one) to view the directory again, and you’ll see that a plus “+” has been added to the listing for that file. Also, below the file the ACL is printed; in this case it will be “**0: group:admin allow write,append**”.

Also note that ACLs do not have the ability to differentiate between a user and group with the same name. While this is common sense for most people, occasionally there will be an instance where a name is used by both. Avoiding this will avert a lot of problems.

Notice how we never modified the UNIX permissions? One thing to note here is that the ACL arguments are evaluated before the UNIX permissions; if there’s no matching ACE, the system then evaluates the request based on the UNIX permission settings.

Now let’s complete the other half: declaring that the admin group does not have the right to read the file. We can express that as so:

```
chmod +a "admin deny read" file
```

By using **ls -le** again, we can see that the new rule has been inserted. But why is it before our first rule? As it turns out, there’s a specific order in which these arguments are evaluated. The correct canonical order is user-deny, user-allow, group-deny and group-allow. ACL entries can be entered at custom points using the **+a#** flag.

This is definitely something to keep in mind when designing your permission structure with ACLs. Now let’s say we want to get rid of the entries we just created and not delete the file, but just remove the ACEs. We accomplish this with **chmod** again:

```
chmod -a "admin deny read" file
```

Sort of a pain to retype the exact argument, isn’t it? We can speed things along by referencing the argument’s number instead! If we look back to our example, we see that there’s only the write/append allowance left. Instead of retyping the argument and using the **-a** flag, let’s reference the number. We can do this by replacing the **-a** flag with **-a#**; the pound sign signifies that a number is being used instead of the string to reference the argument.

```
chmod -a# 0 file
```

Using **ls -le** will now not show any ACEs, but note that there is still an ACL attached to the file; it’s simply empty right now. There are many other cool things that you can do with **chmod**; to learn more, **man chmod** is all you need (skip down to ACL Manipulation Options).

Inheritance

Inheritance is one of the trickier components of ACLs. By default, no inheritance is used when assigning an ACL. You can change that, however, by specifying an inheritence preference in the third part of an ACL entry with the standard permissions.

- file_inherit just says that files created in the directory will inherit the ACL entries of the parent directory.
- directory_inherit is the same as file_inherit, but for directories.
- limit_inherit specifies that only the immediate subdirectories and files of the parent directory will inherit its ACL entries.

only_inherit states that the directory containing the only_inherit ACL entry will not be affected by the entry, but subdirectories and files will.

However, it’s also peritent to remember that the inheritance is “static”, which means that permissions will only be applied the first time that the subdirectory or file is created; subsequent changes to the parent’s ACL listings will not affect existing subdirectories and files (although new ones will certainly inherit the updated listings).

Summary

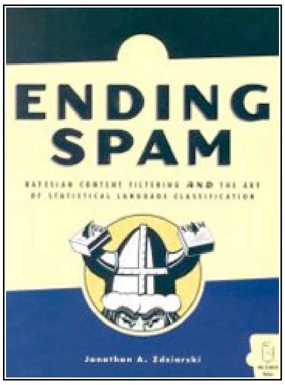
The introduction of Access Control Lists into OS X Tiger is very exciting to say the least. Once a subject of much want by administrators used to the flexibility and power of ACL’s in other (arguably) industrial-strength server OS’s like Windows 2003 Server, ACLs bring a strong, relevant tool to the OS X server toolbox and allow administrators to much more accurately match server permission structures to represent their company's organizational structure, and is scalable enough to do so well for small and enterprise companies alike. Although Apple has failed to give ACLs their due recognition, this is one tool that OS X Server admins should take a very long look at and consider integrating into their OS X Server setup.

Matt Willmore is a graduate student pursuing is Master's degree at Purdue University. When not writing for MacZealots.com, Matt does Macintosh support for Purdue's Engineering Computer Network. He is also a Staff Resident on campus and Vice President of the Purdue Mac Users Group. Matt hails from Columbus, Indiana.

Ending Spam

by Jonathan Zdziarski

No Starch Press, ISBN: 1593270526



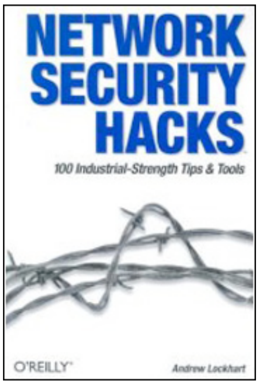
This is yet another spam related book we are taking a look in (IN)SECURE Magazine. It is obvious that spam is one of the major Internet problems, so besides the countless products and services developed around the globe, we also see an increase in books that discuss spam.

Although I believe that many of titles that deal with spam in general are not very good, there are publications like this one that proved to be a very good read. Besides giving a pretty insightful view on spammer techniques, the book's value is the vast coverage of mathematical approaches to spam filtering.

Network Security Hacks

by Andrew Lockhart

O'Reilly, ISBN: 0596006438



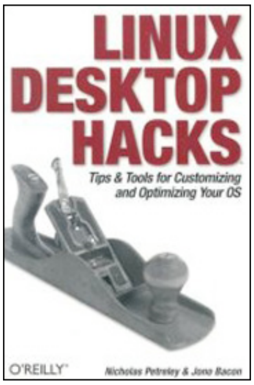
O'Reilly Hacks are one of my favorite IT books. From the first one I came across (Flickenger's Wireless Hacks), to this one, Hacks series provides continuity in quality information. Network Security Hacks features 100 hacks specifically connected to network security.

The majority of info is connected to *NIX based systems, but there is also a good portion for Windows people. Besides the cool tips on general secure networking, the book hosts some good sets of info on intrusion detection and tunneling. This is a book every security professional should have on their bookshelf.

Linux Desktop Hacks

by Nicholas Petreley and Jono Bacon

O'Reilly, ISBN: 0596009119



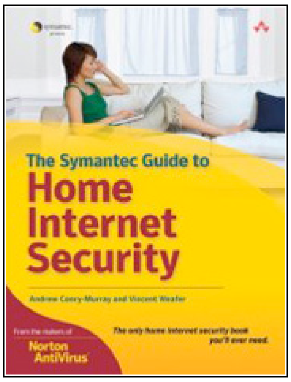
As this is the second "Hacks" book mentioned in this issue of (IN)SECURE, I will limit my scope on mentioning just some of the good tips and tricks you can learn from it.

Here are some interesting tidbits: reduce OpenOffice.org startup time, trigger on-screen events with some applications, accelerating remote X applications, twitching KDE and GNOME windows managers, view PDF and Word files within Mutt, connect to Microsoft PPTP VPN's and my favorite setting up a custom Con Konivas Overloaded kernel.

The Symantec Guide to Home Internet Security

by Andrew Conry-Murray and Vincent Weafer

Addison-Wesley Professional, ISBN: 0321356411



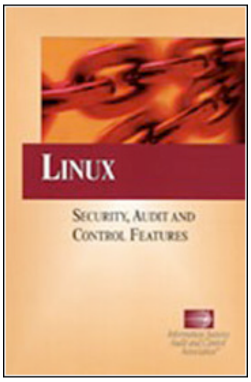
The majority of global internet security problems are related to home users. Viruses, trojans, worms, spyware, evil “hackers”, the list goes on and on.

As the most important thing in this situation is proper education, books like these come really handy. This is a very easy to read manual on basics of home Internet security. Besides extending the reader’s knowledge on subjects like firewalls and identity theft, the authors give tips and suggestions on preventing problems, as well as fixing them.

Linux: Security, Audit & Control Features

by K.K. Mookhe and Nilesh Burghate

ISACA, ISBN: 1893209687



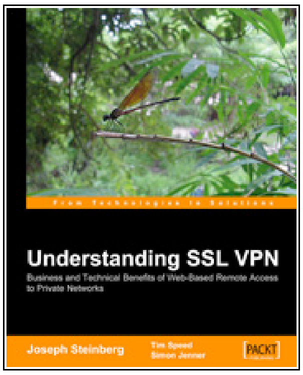
This book is light reading material that doesn’t go too deep into details, but provides a rather good overview of Linux operating system with emphasis on security control and auditing issues.

This isn’t a technical Linux security book, but a very nicely written introduction to the major high-level aspects of Linux security. The book is published by Information Systems Audit and Control Association (ISACA) and it is not widely available. You can get your copy from the ISACA web site – www.isaca.org.

SSL VPN: Understanding, Evaluating And Planning Secure, Web-based Remote Access

by J. Steinberg and T. Speed

Packt Publishing, ISBN: 1904811078



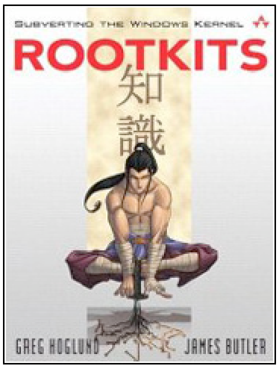
SSL VPNs became quite a hit in the last couple of years. When you take a look at acquisitions and mergers, you will see that majority of early SSL VPN developers were acquired by larger companies.

This book is quite a good read, as it is concise and straight to the point. The author shows clear advantages of SSL VPNs and discusses various facts surrounding this technology. Although the book has just above 150 pages, it contains everything you should need to know on SSL VPNs.

Rootkits: Subverting the Windows Kernel

by Greg Hoglund and Jamie Butler

Addison-Wesley Professional, ISBN: 0321294319



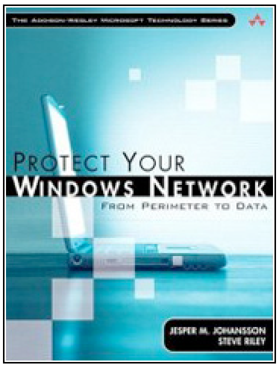
I remember meeting Greg Hoglund at RSA Conference in San Francisco a couple of years ago. He was promoting a book he previously co-wrote "Exploiting Software". In a short interview we've done with Greg, he had a lot of thoughts on rootkits and mentioned that he is going to write THE book about them. Fast forward two years – Hoglund and Butler wrote a perfect read that goes very deep into tactics used by rootkits.

Their main idea is not to cover specific pieces of malware, but go through all the concepts that rootkits normally use. As a sidenote, the book focuses on Windows kernel rootkits.

Protect Your Windows Network: From Perimeter to Data

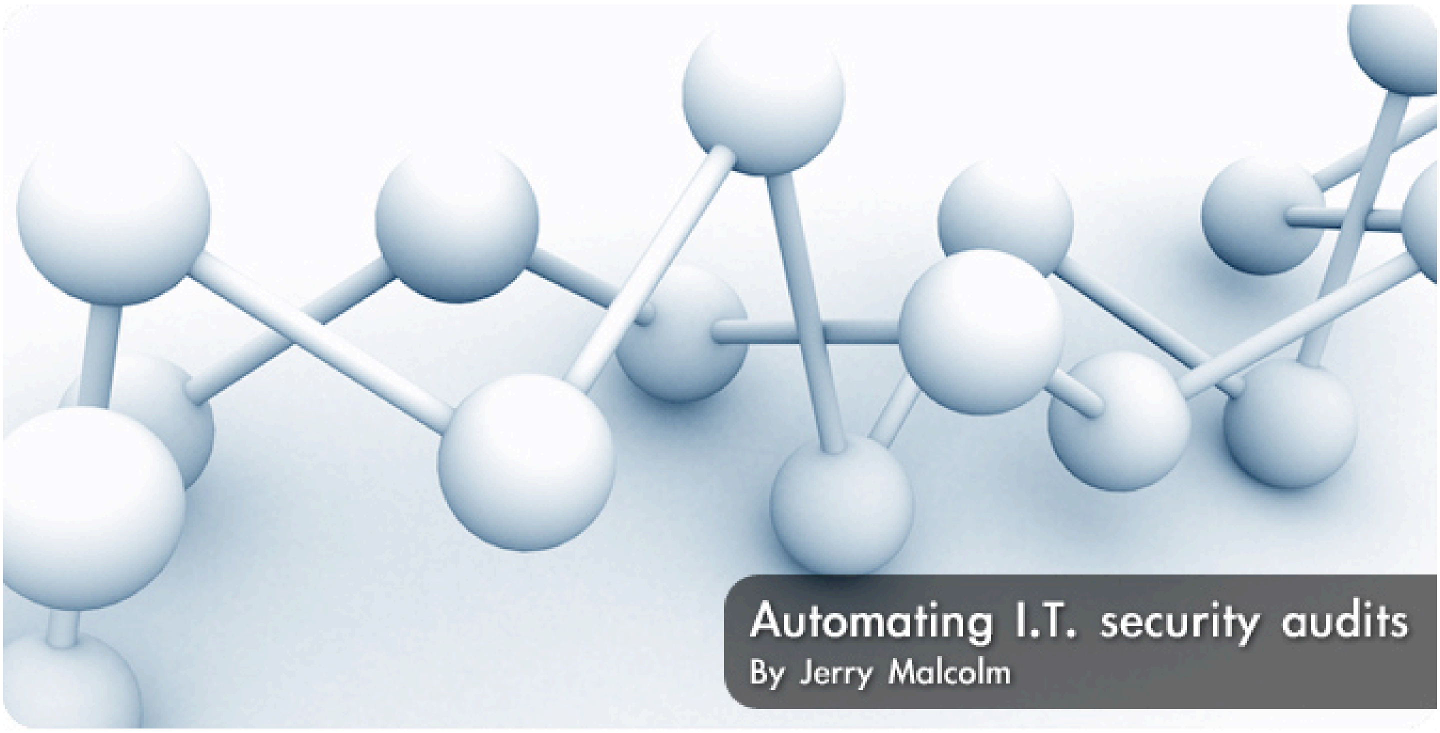
by Jesper M. Johansson and Steve Riley

Addison-Wesley Professional, ISBN: 0321336437



Johansson and Riley are both Senior Managers in security units of the Redmond giant. As they are quite experienced in topics surrounding Windows security, their combined knowledge provides the reader with one of the better and most complete guides on fortifying a Windows network.

While the content is not based on step-by-step instructions, the authors use a number of examples and case studies to prove their points and help administrators to clearly understand all the aspects of their job.



Automating I.T. security audits

By Jerry Malcolm

Through out the computing world, the need for increased wireless security is well known and wide spread. Network security is the focus of the attention because of the acceleration and the publicity related to network and intranet related computer crimes. Well known are the spread of virus and more malicious computer code that can shut down or even destroy a computer data centers ability to support the needs and demands of the organization. The hysteria in the media and in public discussion is compelling but it is not a false hysteria. The crippling affect of virus and other network attacks are quite real and pose a genuine threat to the infrastructure of any organization.

Further, network security is under barrage from hackers and criminals intent upon breaking through the firewalls and other security obstacles to access sensitive data.

Whether for corporate espionage or to access financial information for identity theft or to alter accounts, these attacks are frequent and coming at both a greater degree of frequency and with greater sophistication as well.

In the past, security compromise to the network was the result of hackers who were taking on the challenge of breaking into a network for fun or thrills. While these attacks were serious and deserving of significant response, they masked the real threat to come. In recent years, the attacks have been orchestrated by well organized criminals intent on stealing information, damaging the infrastructure of a corporation or performing other digital violence against the infrastructure of the company. A final level of threat but a very serious one is the classification of attacker who is doing so as an act of violence as a result of international hostilities.

Cyber attacks from terrorist organizations is a very real possibility and just as dangerous and insidious as a physical attack on a building or an act of open warfare in the battlefield.

There is rising evidence that antagonists are targeting vital infrastructure institutions such as telephone companies, nuclear power plants, electrical generation facilities, government offices, water storage and distribution facilities, oil and gas transportation systems and transportation facilities such as ports and airports.

The Challenge for Network Security Professionals

Those charged with security of the organization must be vigilant against such attacks both from without and within the organization as any portal into the network must be viewed as a potential open door for attack. Unlike any profession or application area in computing, network security is extraordinarily dynamic. Each new wave of network attacks comes at greater levels of sophistication and ability to bypass, shut down or in other ways circumvent the most advanced protections network security puts into place. The struggle is a continuous cat and mouse game with network security in a constant war to keep up with the newest methodologies and technological advances of hackers and computer criminals and the adversary continuously maintaining pace with state of the art network protection concepts and tools and finding ways to neutralize those protections.

At the same time, that utmost security must be constantly maintained, enforced, monitored and tested, all of this access control but occur without interfering or even slowing down the genuine value of the network to the business entity. The trade off of performance for security is often one that is not acceptable for the company which complicates the challenge for network security experts. Network security needs to operate in an "invisible" mode, maintaining optimum security for the network, the systems that operate the computing environment and the crucial data structures that support the systems but doing so without hindering the flow of data or the commerce which those systems are designed to protect.

Another tier of accountability about which the network professional must be concerned are layers of regulation that are imposed upon certain industries. Certain critical industries such as the financial sector or any industry related to national defense or government related commerce are held to a strict standard of security that must be maintained to be allowed to continue to function in that environment. The Health Care industry similarly has a detailed regimen of security standards that not only do health care institutions must be aware of but support organizations as well. Further, international standards often are imposed if the organization conducts business with foreign governments or across country boundaries, particularly if that commerce occurs with areas of the world where hostilities have occurred in the past such as the middle east. The result may mean that any given organization may be accountable to multiple tiers of externally imposed security standards.

Certain critical industries such as the financial sector or any industry related to national defense or government related commerce are held to a strict standard of security that must be maintained to be allowed to continue to function in that environment.

It is the responsibility of network security professionals to stay current on these regulations and assure that the network security infrastructure is compliant on a day-to-day basis and stays in sync with mandated security audits to satisfy such external security standards.

Security professionals also must content with the dichotomy between perceived risk and genuine risk. Like many infrastructure functions, to an income generating institution the keen eye on the bottom line means that the security function is often under tight budgetary controls. While being held to a high standard, the organization is reluctant to fund extensive security resources without significant justification.

Nonetheless, because of intense coverage of network security breaches, the hysteria for high levels of security may be intense within the organization and the security department may have a challenge in assuring cost justifiable security without giving the impression of neglecting security for those who feel strongly that their systems are in need of additional protection.

Risk Management

One methodology that is commonly employed to determine real versus perceived risk is to compute the risk using commonly accepted risk management formulas. The formula for determining business impact analysis of risk is:

Risk = Threats x Vulnerabilities X Impact.

Of the three variables, vulnerabilities and impact are the most concrete. The formula can be applied to each subsystem of application area. Impact would be a computed value based on the relative value of the application to the organization and to what extent that application or function is business critical. Hence, the discipline of determining the correct level of security would be to perform an impact analysis of the function. That exercise by itself could do much to calm perceived risk.

Vulnerabilities are more difficult to determine but security standards that are commonly used within the security industry are helpful in gauging what level of vulnerability is permitted within any network security system.

Some vulnerabilities are tolerable and necessary to permit fluidity in the movement of data and communications through the network. Further, if the impact analysis comes back with a low level of impact for the application, vulnerabilities may be permitted because plugging those security gaps is just not justified.

The level of threat is the most difficult to gauge but security mechanisms already in place should be able to provide statistics on the amount, severity and frequency of attempted security breaches. Those statistics can be used to complete the formula and provide each functional area of the enterprise with a fairly reliable risk analysis by function or application area.

I.T. Auditing

Historically the approach to I.T. Auditing has been slow to move toward automated solutions. However, with the advent of network dominated systems, auditing techniques have evolved as well. To date the three most common approaches to auditing have been manual audits and script based auditing and network availability scanners.

The manual audit which can trace its roots back to the heart of auditing in the accounting world is one that gives the auditor a high sense of confidence that the audit is complete.

While automated procedures clearly are the wave of the future in I.T. auditing, some level of manual inspection of the outcome will always be necessary.

The automation process does not remove the judgment and analysis that only a human auditor can bring.

Moreover, it facilitates a greater efficiency when the manual auditor steps in by automating the data gathering and some level of key indicator evaluation so that the most common audit triggers are handled by the automation software thus reserving the high-level skills of the network security auditors for the types of problems that demand their unique contributions.

Scripts have their role in auditing of the network infrastructure because they remove much of the tedium of data gathering.

In script driven data collection is the only practical way to perform a continuous monitoring of such a fast moving and fast changing environment of a network. But keeping in place certain customized software snifters, bots, spies or other types of scripts, security data that flows through the network on a continuous basis can be collected, evaluated and responded to.

Historically the approach to I.T. Auditing has been slow to move toward automated solutions. However, with the advent of network dominated systems, auditing techniques have evolved as well.

Alert and alarm triggers can be very useful within the monitoring scripts but those too are under the constant review and control of network security and auditing expertise. So in that way, scripts simply serve as an extension of the arm of manual auditors.

Requirements for an Automated Solution

Network security automation and management is an area of the enterprise that is under constant review, improvement and scrutiny.

If the reader is researching automating I.T. security audits as a new initiative or as a continuous process of network infrastructure review and upgrade, the requirements for a viable solution remain the same.

The following six criteria then can be used in two ways. First, they can be used to evaluate the current methodology. By applying an independent set of standards to the structures and systems in place today, the areas of potential improvement will surface quickly. In that way, we know where we need to apply automation and our efforts are efficient in terms of time, labor and cost needed to keep the infrastructure secure.

Secondly, these criteria can become the basis for evaluation of potential solutions. The marketplace is rich with products and vendors who are skilled in showing the need and effectiveness of their solution for your network security.

While each solution no doubt has its value, the appropriateness of each solution should be evaluated by the same criteria so as to “level the playing field” so to speak and assure that solutions selected and purchased or leased are the correct fit to your particular network problems.

CRITERIA 1 – Efficiency

By automating security procedures, much of the day in, day out hands on operational and evaluative activities needed by personnel within the security or operations area is minimized. This improves the use of highly trained and expensive human resources.

Not only is this good for the enterprise by reducing the overhead for operations of security, it utilizes staff at their most valued level which is good for the individuals involved in maintaining the security and ongoing health of computing systems for the business entity.

CRITERIA 2 – Compliance

As we mentioned before, security policy and procedures can be quite strict and the security team may have multiple tears of security compliance from the enterprise and from external auditing entities that impose standards upon the organization to be allowed to conduct business in a secured marketplace. Auditing to the level of maximum compliance with these standards using manual audit techniques is inefficient and seldom is able to sustain compliance in an ongoing fashion.

Well designed automation can utilize internal alert and alarm compliance rules that can be fine tuned as regulations change or new threats are identified and protected against. In this way, compliance becomes a daily standard rather than something that is corrected after each manual audit.

CRITERIA 3 – Standardization

Modern computer systems and networks are diverse in nature, design and operation across the network and the enterprise. If each of these subsystems is audited using tools unique to that system, there is no standard methodology to determine if each element of the network is at the same level of compliance. Any automation solution must utilize a single measurement system and provide translation of security data that is peculiar to each subsystem back to a single measurement and reporting logical system. In doing so, the auditing team can evaluate the entire infrastructure against a single logical scheme and not have to learn numerous measurement schemes and attempt that translation manually.

CRITERIA 4 – Accuracy

Manual security audits by necessity cannot sample the data stream as thoroughly as an automated system. Automating the auditing process not only places the whole of the security picture under audit on a continuous basis rather than on an ad hoc basis, it gathers thorough data samples and is must be able to compile, evaluate and model that data in an accurate fashion. When evaluating security systems, whether in place or under consideration, those systems must deliver results that line up with observed reality within the systems themselves.

The ability to gather, store, report and respond to divergent data in a timely and accurate fashion is the backbone of the value of an automated security plan.

CRITERIA 5 – Auditing Frequency

As mentioned above, because security data is collected continuously by automated security data collection agents, the frequency of reporting can be continuous.

Reporting on a daily, weekly and monthly data using standardized and automated web based reporting can transform the auditing process from an infrequent intrusive activity to an ongoing fact of life in the security environment.

Further, management reporting from that same database of security metrics can produce inquiry driven reports for management or auditing agencies that must hold the security infrastructure to high levels of policy compliance.

The reliability and integrity of the automate solution must be beyond question to achieve the high levels of reporting and response that we are calling for in this criteria list.

CRITERIA 6 – Flexibility

The infrastructure of a complex I.T. environment will have a wide array of systems, operating environments and components. The automated solution must be able to adapt to the operating system in which it is performing, function with little impact within that setting and report interpreted data back to the auditing console without difficulty.

Furthermore, as the infrastructure evolves and grows, the automated auditing facility must be able to scale and adapt to pick up the changes in the network and recalibrate its evaluation algorithms so little or no hands on effort must be made to assure that continuous security auditing goes on uninterrupted even in a quickly changing I.T. setting.

The automated solution must be able to adapt to the operating system in which it is performing, function with little impact within that setting and report interpreted data back to the auditing console without difficulty.

The Process of Putting Automation into Place

The rationale for automating the security auditing process is compelling. If the determination has been made to move forward with new or increased levels of automation, the steps for moving forward are often similar from enterprise to enterprise. Before evaluating the solutions, conduct a study of the current security tools and levels of auditing. This study will both identify what is of value in present day systems and where change would benefit the enterprise the most. Those change criteria then become the scope document for a project to improve security auditing for the organization.

Further, as that study is executed and before potential solutions are interviewed, certain standards and security requirements will become clear. You will no doubt have the compliance requirements already in mind particularly if compliance is a weak area of the current solution. Other business and technical requirements for the solution to automating the security function will suggest themselves to be added to the list already provided earlier in this white paper. That list of requirements has three applications. First, it becomes the standard against which all possible solutions will be evaluated. Second, it will provide a starting place for a list of policies to become the automation checklist when the pilot of an automated solution is set in motion. Finally, it will provide the organization system for a future reporting strategy and as such give insight into the data collection and retention policies that will be important in the early design phase of the solution.

An additional parameter to be an important part of the early study which can be viewed as a needs analysis or requirements definition project phase is that a model for the criticality of systems to be included will emerge. Using similar definition making criteria that was part of the Impact Analysis we discussed under Risk Management earlier, systems, applications and functions can be tagged as high, medium or low candidates for implementation of audited security based on the impact and potential risk to that system. Because the evaluation process is independent and devoid of bias, the subsystems are judged entirely on value rather than on importance to individuals within the organization. This is far more healthy for the enterprise overall than similar evaluations done manually that might be clouded by personal preference or opinion. Upon conclusion of the study, the organization will have a detailed snap shot not only of how the infrastructure looks but also its current state of security readiness and what will be the high priority issues in developing a solution for automating the security infrastructure.

From that study, management can be kept abreast of the process and be given break points to steer the process, abort it or sanction what is going on and provide funding and management support to push the process through to completion. A systematic approach to the problem will increase management confidence that the proposed solution, even if it is an expensive one, is the right solution for the organization. From that point forward, the process of entertaining solutions, selecting a software package or designing a custom developed solution and preparing a development, testing and implementation strategy can go forward along familiar project management methods. But because the security team has taken the initiative to bring to management solutions to the security dilemma, the likelihood of success is increased tremendously and the potential that the implemented solution will provide a highly secured infrastructure is excellent. In that way, the process of putting in an automated solution improves the health of the organization as well as the health and operating ability of the security department itself.

Jerry Malcolm is an I.T. professional with 30 years of experience at all levels of IT project development, design, management and documentation. Since 2003 Mr. Malcolm has been the owner/principle of Malcolm Systems Services, an IT services consulting firm.

CNIS 2005: IASTED International Conference on Communication, Network and Information Security
14 November-16 November 2005 – Phoenix, USA
http://www.iasted.org/conferences/2005/phoenix/cnis.htm

IBM European SecureWorld 2005
21 November-24 October 2005 – Prague, Czech Republic
http://www.ibm.com/training/conf/europe/secureworld

Asiacrypt 2005
1 December-4 December 2005 – Chenna, Madras
http://www.iacr.org/conferences/asiacrypt2005

Infosecurity New York 2005
6 December-8 December 2005 – Jacob K. Javits Convention Center, New York, NY
http://www.infosecurityevent.com/digitalsecure

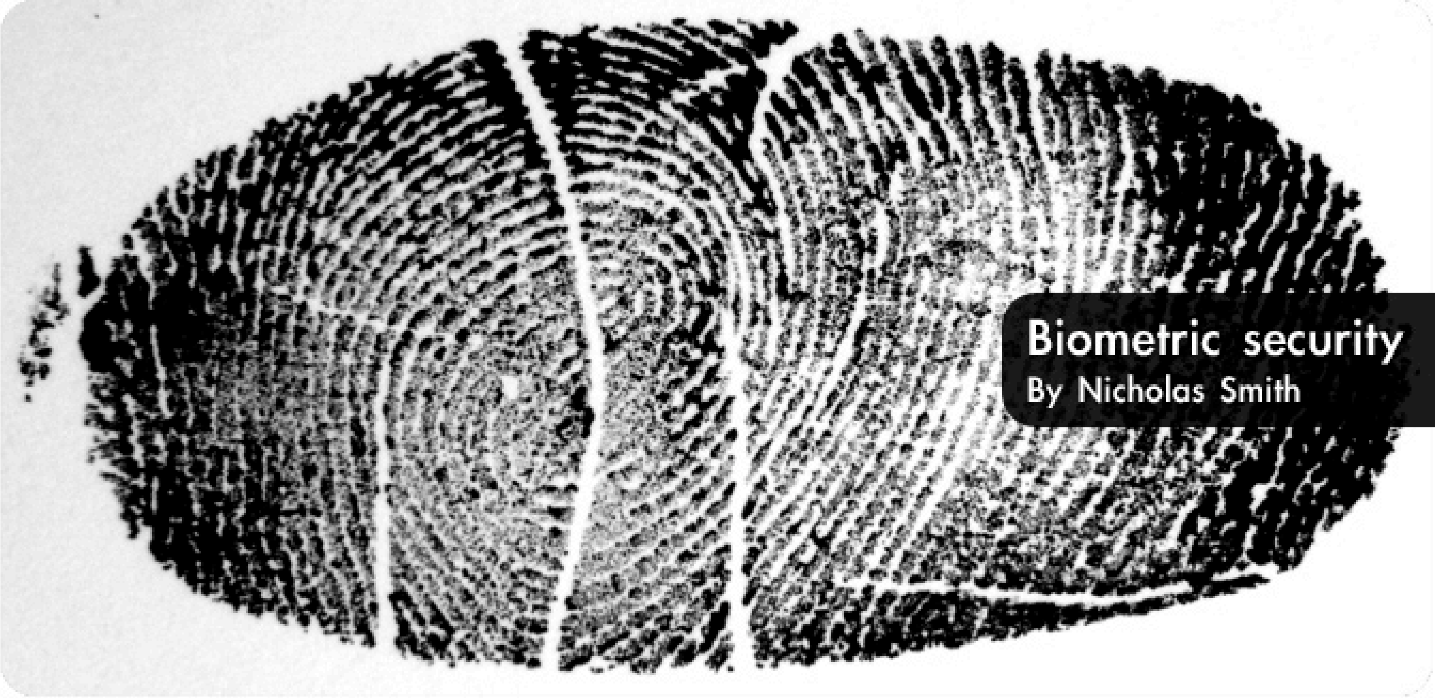
3rd International IEEE Security in Storage Workshop
13 December – Golden Gate Holiday Inn, San Francisco, California, USA
http://www.ieeeia.org/sisw/2005

Black Hat Federal 2006 Briefings and Training
23 January-26 January 2006 – Sheraton Crystal City, Washington DC, USA
http://www.blackhat.com

RSA Conference 2006
13 February-17 February 2006 – McEnery Convention Center, San Jose, CA, USA
http://www.rsaconference.com

Black Hat Europe 2006 Briefings and Training
28 February-3 March 2006 – Grand Hotel Krasnapolsky, Amsterdam, Netherlands
http://www.blackhat.com

iTrust 2006
16 May-19 May 2006 – Piza, Italy
http://www.iit.cnr.it/iTrust2006/index.htm



Biometric security

By Nicholas Smith

Biometric security is a growing form of security used to allow physical identification, but such security measures bring up ethical issues concerning things such as privacy, identity theft, and integrity over a remote transfer. Although the security of a person's biometric is normally greater than that of a variable data means of authentication, such as a username/password pair, credit card number, PIN number, etc., it is less flexible and more damaging in the event of compromise.

Introduction

As businesses, government, and other organizations look for a means to increase security to access data such as bank and credit card accounts, website membership, personal information records, etc., biometrics are becoming a popular new means to identify people [1]. Some biometrics are more suitable for various applications, and so many different ones are used. However, each has its own different, but similar, advantages and disadvantages. The disadvantages will be the largest topic of this article, as they concern various and debatable ethical issues. The most popular biometrics and their properties will be discussed.

Fingerprints

Fingerprints are known to be unique for each person, and therefore seem to be a good means of authentication to determine that a person is who they say they are. The government has been using fingerprints for a long time to identify people for various reasons, but most popularly for identifying criminals. Fingerprints have been used in criminal investigations for more than the last 100 years, and have been dated back as far as the 1500s where they were used for identification [2]. The first issue with taking a person's fingerprint is that the person has a right to privacy of their personal information. By requiring a person to submit their fingerprint to the government, or any organization for that matter, some people may consider it a violation of privacy. Regardless of whether or not a person consents to submitting a fingerprint, or fingerprints, the result is a copy of a unique entity, which will be kept in some sort of data store. This data can then be used against the person in a prosecution because a fingerprint is a valid and concrete means of identification in the courts. So, consider the possibility that a crime was committed, the police dusted for fingerprints, and the fingerprints incorrectly matched that of an innocent person.

How could this incorrect match happen? The first possibility is that the software used to match the crime scene fingerprints to a data store of fingerprints made a false conclusion of the match. It is not impossible for the software to make a mistake, and more likely that it will at some point in time. Practically no software ever written has been 100% perfect, so it is possible and likely that software used to match fingerprints will give a false positive at some point. The second possibility is that the criminal planted forged, or even fake, fingerprints.

A fake fingerprint is most likely to come up with no match, but possibly could. A forged fingerprint, however, is much worse (and more difficult to do) and would falsely identify the criminal. If the criminal went out of his way enough to copy someone's fingerprint and plant it at the crime scene, it is very probable that he/she used a fingerprint of someone they know and/or someone who would make a prime suspect for the crime. In any case, an innocent person could be sentenced to months, years, or a lifetime in prison. Had the person's fingerprints never been required to be submitted, the situation would have never happened.

The more severe case, in which a person's fingerprint is forged/stolen, is actually the easiest and most probable means of false fingerprint identification. It is not difficult at all to steal a person's fingerprint. It can be done while the person is sleeping, after they touch something and walk away, or can be forcefully copied in an extreme situation.

The issue of false identification by fingerprint does not apply only to crime, but anywhere that a person may suffer consequences that they do not deserve. Just because the chance is normally extremely small, should such an issue not be important? Does it not matter if even one in a billion people are falsely identified by their fingerprint, and consequently suffer damages for it? If fingerprints were to become a widespread means of authentication, obviously something would have to be done to increase security. Maybe a password or second biometric would be required in conjunction with the fingerprint.

IT IS POSSIBLE AND LIKELY THAT SOFTWARE USED TO MATCH FINGERPRINTS WILL GIVE A FALSE POSITIVE AT SOME POINT.

Retina Scans

A retina scan looks at the formation of blood vessels in the back of a person's eye because the patterns are different for each person. Such a scan requires a person to be very close to the scanning device and focus on a given point [1]. Unlike a fingerprint, physically replicating this data is nearly (if not) impossible. Because physical compromise is not a threat, it would seem that this method of identification would work great. However, it is not convenient for the person being identified, and so has not become very popular.

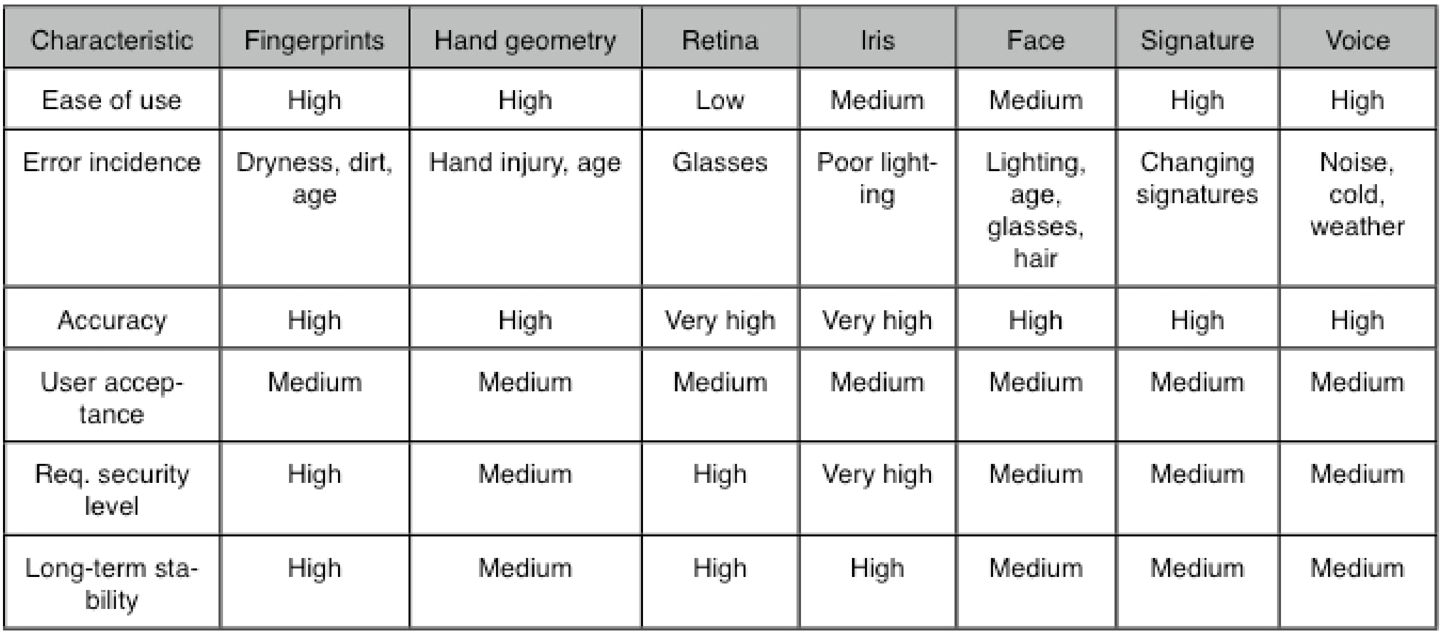
Iris Scans

An iris scan analyzes features on the colored ring of a person's eye that surrounds their pupil. This data can be collected much easier than a retina scan and usually gives more accurate results [1]. Like a person's retina, the iris is nearly (if not) impossible to duplicate. The fallback about this method of identification is that the technology has not advanced to the point of being as usable and easy to implement into a system as other methods. Currently, some U.S. state law enforcements have put iris-scanning technology to use in select jails and prisons. The first correctional facility to use such technology was Lancaster County Prison in Pennsylvania, which began using it in 1994 for prisoner identification. Since then, multiple other jails and prisons have begun using it. The Charlotte/Douglas Airport in North Carolina and Flughafen Frankfort Airport in Germany allow frequent fliers to register their iris data to allow faster boarding because they can be identified quickly with an iris scan [2].

Face Recognition

Currently, facial recognition is used mostly in areas where a small database of faces is being used for matching (usually criminals). The facial recognition camera looks for distinguishing facial features such as the distance between features (eyes, nose, mouth) and shape [1]. In an environment where a person was willingly being authenticated and would stand directly in front of the camera, this authentication would work relatively well, and would obviously be easy for the person to use. However, if the user is willing to be authenticated, other methods are easier and faster to match, and more reliable. Because of this, it is used for "lookout" rather than authentication in areas where people walk through, and does not require them to look at the camera or even know that it's there. These cameras are used commonly in casinos to detect scam artists, or in places where there are large gatherings of people to detect wanted felons and/or terrorists,

Each biometric has characteristics that vary in degree of quality. Table 1[1] on the following page gives an idea of the relative quality for common biometrics.



Characteristic	Fingerprints	Hand geometry	Retina	Iris	Face	Signature	Voice
Ease of use	High	High	Low	Medium	Medium	High	High
Error incidence	Dryness, dirt, age	Hand injury, age	Glasses	Poor lighting	Lighting, age, glasses, hair	Changing signatures	Noise, cold, weather
Accuracy	High	High	Very high	Very high	High	High	High
User acceptance	Medium	Medium	Medium	Medium	Medium	Medium	Medium
Req. security level	High	Medium	High	Very high	Medium	Medium	Medium
Long-term stability	High	Medium	High	High	Medium	Medium	Medium

Issues Regarding All Biometrics

All biometrics suffer from a common security flaw by the definition that they are a biometric. The flaw is that they are physical aspects of a person, and so cannot be easily changed[1]. This is good in the regard that it is difficult to forge/compromise, and so a criminal cannot easily take the identity of another person. The ease of doing this varies in degree in different biometrics, but is especially vulnerable with fingerprints. The drawback is very severe, however, in the event that a criminal manages to successfully forge/compromise a person's biometric. The innocent person is then left with no option but to change his/her own biometric. This is a huge problem, and in the event that it occurs, who is responsible for fixing it? And, how can it be fixed? Unfortunately, the person responsible for fixing it would be the victim in most cases.

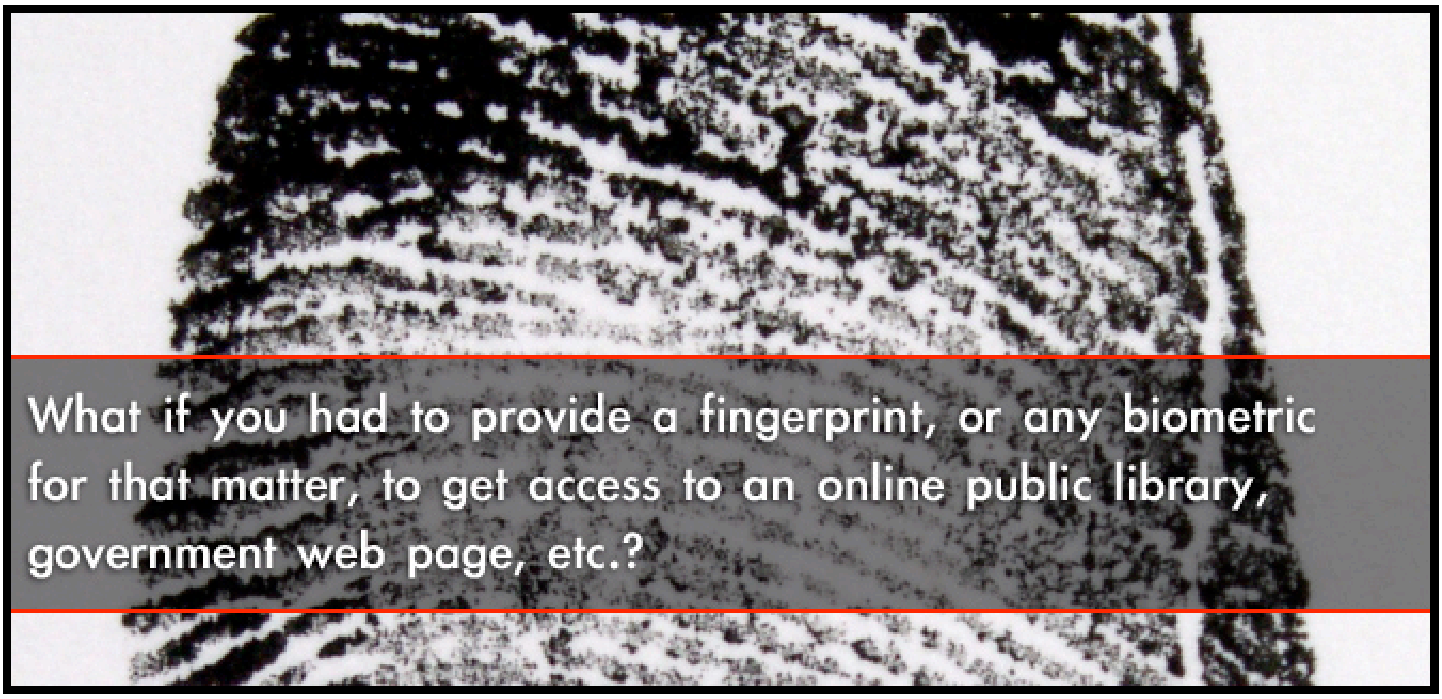
This may be different in cases where the criminal is the victim's employer, or anyone who is legally responsible for the integrity of the victim's biometric information. For example, if a business were to authenticate its employees access to a building by using a biometric, and the business then misused the employees information (i.e. to commit identity theft), the business may then be held responsible for any costs associated with changing the victim's biometric, and possibly for the damages/losses accrued by the criminal. Regardless of who is responsible, though, it still requires that the victim change a physical aspect of his/her body. This is practically intolerable.

Also, where will biometric information data stores be? It depends on what it is being used for, and who (or what company) is in charge of the authorization. In any case, how are these data stores expected to be kept secure? A small business might hire just about anyone with knowledge of how to operate and maintain the data stores. Who is to say that he/she is trustworthy? A large business or government would most likely hire experienced, qualified professionals to do the same job. Their experience and qualifications don't necessarily make them any more trustworthy, although people generally tend to think so, and they may be right or wrong. Regardless, someone at some point (probably more often than you'd like to think) will leak or misuse the information. In addition to people with physical access to the data, there will be people who manage to bypass security measures and gain access to the data remotely. The database obviously cannot be stored in a biometric scanning peripheral, so it must communicate remotely to a data store. This line of communication is where another problem lies...

Issues Regarding Biometric Information Transfer

The need for remote user authentication is obvious and has been discussed. However, the transfer of biometric information, regardless of the communication method, will always be vulnerable in some way. The Internet is a very popular method for information transfer, and the information is available to the hands of whomever has access to the hardware used to do the transfer. Of course, secure communication can be set up by encrypting the data. But, methods used to do this have vulnerabilities, and the encrypted data can be, has been, and will be compromised.

The focus of this issue is not on the details of how data can be compromised, but the fact that it can and will be. In the event that this occurs, the victim is put in a similar situation as if their biometric information had been physically copied.



What if you had to provide a fingerprint, or any biometric for that matter, to get access to an online public library, government web page, etc.?

The victim cannot easily, or even possibly, depending on the biometric, change their physical aspect(s) used for authentication. Once again, this is an ethical issue because the victim may or may not have ever wanted to submit his/her personal biometric information to use for authentication. Of course, not in all cases are people required to submit such information. But, consider an example where biometric authentication becomes widespread for government restricted or monitored access. What if you had to provide a fingerprint, or any biometric for that matter, to get access to an online public library, government web page, etc.?

Real Examples of Privacy Issues

To follow up on the previous section about the example of requiring biometric authentication for remote government restricted or monitored access, this could just as easily, and probably more likely, happen to authenticate physical access to government property. This is already beginning to happen in Tampa Bay, FL, where the school system (Pinellas) is considering the implementation of a system that would keep track of children on school buses [3]. Each child has to scan his/her fingerprint when entering and exiting the school bus to ensure they each get on the right bus and get off at the right bus stop. It also ensures that only the allowed students for each bus will be let on, so it serves as a means of authorization. Regardless of its purpose, the scan is required, which means that the children do not have a choice as to whether or not they want their personal physical information given to the government. A parent can opt out from having their child's fingerprint scanned, but this is not the decision of the child whose fingerprint will be taken. According to the article from which this information came from, parents are expressing concerns about their children's privacy, and how such a proposed system is similar to a "big brother" scheme. The school system assures them that the fingerprints will be held in a secure database, which will only be accessible by "school district employees" and will be password protected. The point that they argue is weak not only because they don't define who or what a "school district employee" is, but also because they say the database is password protected. They also didn’t mention what means of encryption will be used for the wireless communication between buses and the database.

Wireless communications are notorious for being weakly secured, so what's to stop a hacker from intercepting all the student records? With all of these weak points combined, the children's private data could be lost from being leaked by a "school district employee", having the database password compromised, or even having the data be intercepted in wireless transit. Is it worth the chance of losing your children's personal, unchangeable data just so they get on and off the right bus?

Some car manufacturers, such as Audi, have been developing technology that will allow keyless entry and ignition of a car by the use of the owners fingerprint [4]. When the owner puts his/her hand on the door handle, their fingerprints are scanned and checked for authentication. Their prints are scanned again when they attempt to turn the ignition. This all seems like a good idea because you can't lose your fingerprint like you can lose keys. However, as discussed earlier, your fingerprint can be compromised, which gives your impersonator unquestionable access to your car! Currently, unless your keys are stolen, which is rare, a thief must somehow break into a car in order to steal it, which in most cases involves finding a way around the alarm (if any) and hot-wiring it to start it up. Even then, a car's steering column is normally locked to make sure that it cannot be driven without keys in the ignition. Imagine that a criminal is following you, and dusts for fingerprints on things that you touch. You would most likely never notice, and, in the meantime, the criminal is making molds of some sort with your fingerprints. Then, he/she waits for you to go into work, the grocery store, or wherever, and you come back to find your car missing. This seems like a lot of trouble for the criminal to go through, but when it's worth $20k+, a few days, or even weeks, of work seems to be worth it.

Conclusion

Stepping back to look at biometrics as a whole, it is evident that the opportunity exists to use biometrics as means of enhanced security over the currently used methods. However, a person's biometric is private information, and protection of privacy is supposed to be supported by law. There seems to be a limited extent to which biometric authorization should expand before it becomes widespread to the point where it is required by all citizens. Who will decide where to stop, if it does? The government already takes people's fingerprints, so what will keep them from taking other biometric information? Assuming that the government does take such information at some point, will they be responsible for the reparation of damages such as those caused by identity theft? Given that the government would house all of the data by some means, how are the people supposed to put full trust into the belief that their private information is secure? The questions can only be answered with time.

Nicholas Smith is a senior college student at Georgia Southern University in the field of Computer Science. He is interested in network applications/protocol design/security, and synchronization systems modeling. You may contact him at nsmith22@georgiasouthern.edu.

References

(1) Liu, Simon, and Silverman, Mark. "A Practical Guide to Biometric Security Technology".

(2) "Iris Scan" – tinyurl.com/cw376

(3) Koch, Nora. "Have your thumb ready to ride the bus" – tinyurl.com/3cpdd

(4) "Audi at Forefront of Automotive Electronics Development" – tinyurl.com/ap5xk.

PDA attacks, part 2: airborne viruses evolution of the latest threats

By Cyrus Peikari, M.D.

This (IN)SECURE Magazine article is the second of a three-part series on PDA and Smartphone security. This second part is loosely based on the talk I gave at this year's RSA Security Conference 2005 in San Francisco, as well as various review articles I have published elsewhere. However, I have updated it to reflect many of the latest threats in the rapidly evolving mobile malware field.

Contributions to this article came from MARA members Seth Fogie, David Hettel, Petr Matousek, and Jonathan Read, and some excerpts appear with kind permission from InformIT.com

Overview

At the time of this writing (September 2005), airborne viruses are increasing in sophistication at a spectacular pace. For example, the very first Pocket PC virus to appear (Dust) was incredibly complex. It achieved a technological breakthrough equivalent to the Win32 Chernobyl virus, which was the first PC based virus to break into the protected "Ring 0" of the Windows operating system.

Moreover, less than a year after Dust, we have already seen numerous "blended" threats. For example, virus writers have developed anti-antivirus trojans, and have even combined these with the Bluetooth spreading capability of the Carib (Cabir) virus. So in the space of one year, we have seen a viral evolution equivalent to what took 20 years on desktop PCs.

One problem with this rapid evolution in threats is the fact that mobile devices can't support sophisticated antivirus software on current platforms. For example, embedded operating systems don't use "interrupts" (system calls to the kernel), so a heuristic virus scanner on the PDA or Smartphone can't hook a specific interrupt that it might otherwise suspect is a virus.

Another problem is the seeming lassitude of a crumbling and outdated antivirus industry. The old guard antivirus industry often operates under the antiquated principle of "security through obscurity." Some try to keep knowledge of vulnerabilities secret within a closed priesthood of a few selected men. Meanwhile, hackers, identity thieves and virus writers may have access to the same information, while the larger security community is kept in the dark. This can become problematic in the new world of rapidly evolving field mobile threats, where millions of unprotected wireless devices now share data promiscuously "any time, anywhere."

Mobile Antivirus Researcher's Association

For those who have an interest in the field of mobile malware, please consider applying to the Mobile Antivirus Researchers Association (www.mobileav.org).

It is open to security industry members who have a legitimate need to obtain source code and binaries of the latest mobile viruses and security vulnerabilities. MARA is a free, vendor neutral organization whose members come from the global community of professional, embedded malware and wireless security researchers. Our research and discoveries are published as full disclosure to the security community as soon as threats can be verified and tested.

The Threats

The following will describe the evolution of mobile malware from its beginning until the present time. We will progress more or less in chronological order, as well as showing the evolution from simple, nuisance viruses into more sophisticated ones.

Phage

Phage was the first PDA Virus and was discovered on the Palm OS in Sept. 2000. When the virus is executed, infected PDA files display a grey box that covers the screen, whereupon the application terminates. In addition, the virus infects all other applications on the device.

When a “carrier” device is synchronized with a clean PDA, the clean PDA receives the Phage virus in any infected file. This virus will in turn copy itself to all other applications on the clean device.

Liberty Crack

This virus acts as a Trojan horse because it comes in a disguise (although it does not open a backdoor). “Liberty” is a program that allows you to run Nintendo GameBoy games on the Palm OS. Liberty is shareware, but like all useful shareware it has a corresponding crack that converts it to the full registered version. The authors of Liberty decided to pay back the pirates by releasing a “counter-crack” for Liberty that was actually a virus. The developers distributed it on IRC. Unfortunately for the pirate, when executed, the Liberty crack virus deletes all applications from the PDA.

This virus can spread both through the desktop and through wireless email. In fact, it may be the first known PDA virus to spread wirelessly in the wild.

Vapor

The “Vapor” virus does just what it sounds like it should; when infected with Vapor, all the files on the PDA “disappear.” When the infected file is executed, all application icons will vanish as if deleted. This is a trick, since the files still exist. In reality, the virus simply removed their icons from the display. This is similar to setting all file attributes to ‘hidden.’

911 Virus

Older handsets were relatively immune from airborne viruses because they lacked functionality. However, Internet-enabled Smartphones are facile hosts for infection and attack. For example, the “911 virus” flooded Tokyo's emergency response phone system using an SMS (short message system) message. The message, which hit over 100,000 mobile phones, invited recipients to visit a Web page. Unfortunately, when the users attempted to visit the website, they activated a script that caused the phones to call 110, which Tokyo's equivalent of the 911 emergency number in the United States. Thus, the virus could have indirectly resulted in deaths by denying emergency services.

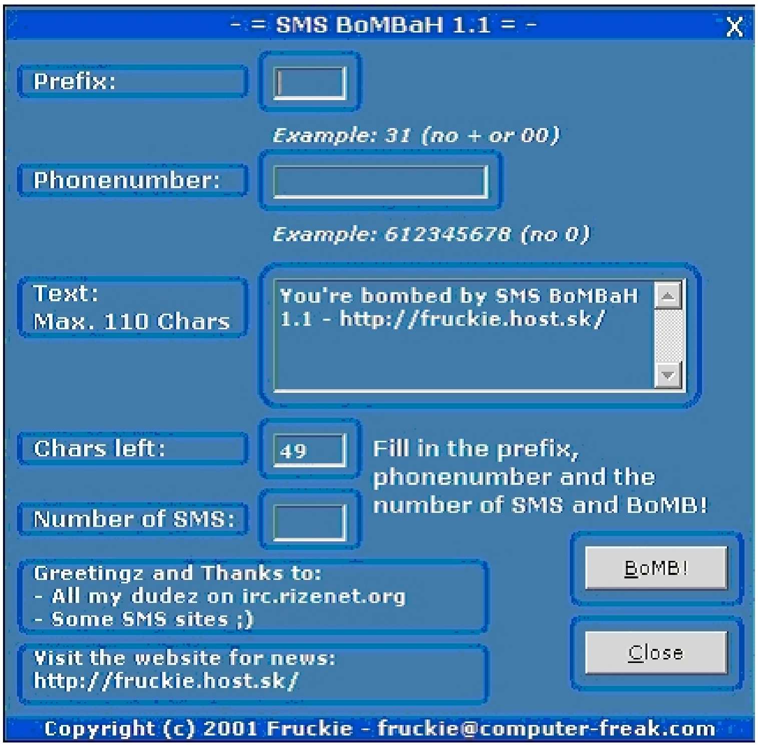
SMS attacks

A potential vulnerability of SMS is that it allows a handset to receive or submit a short message at any time, independent of whether a voice or data call is in progress. In addition, if the handset is unavailable, the message will be stored on the central server. The server will then retry the handset until it can deliver the message.

In fact, there are desktop tools that script kiddies use for SMS Denial of Service bombing, such as Fruckie’s SMS “Bombah” (Fig. 1 on the following page). The same principle of this tool, when coupled to the power of a replicating virus, can potentially result in wide scale Denial of Service attacks.

Another example of such an SMS flooding virus occurred in Scandinavia. When a user received the short message, the virus locked out the handset buttons. This effectively became a denial of service attack against the entire system.

Similarly, a Norwegian company found another example of malicious code. In this case, a Norway-based WAP service developer known as Web2WAP was testing its software on Nokia phones. During the testing, they found that a certain SMS was freezing phones that received it. The code knocked out the keypad for up to a minute after the SMS was received. This is similar to format attacks that cause crashes or denial of service attacks against Internet servers.



- = SMS BoMBaH 1.1 = -

Prefix:

Example: 31 (no + or 00)

Phonenumber:

Example: 612345678 (no 0)

Text:
Max. 110 Chars

You're bombed by SMS BoMBaH 1.1 - http://fruckie.host.sk/

Chars left: 49

Fill in the prefix, phonenumber and the number of SMS and BoMB!

Number of SMS:

Greetingz and Thanks to:
- All my dudez on irc.rizenet.org
- Some SMS sites ;)

Visit the website for news:
http://fruckie.host.sk/

Copyright (c) 2001 Fruckie - fruckie@computer-freak.com

Fig. 1: Fruckie's SMS "Bombah"

Carib (Caribe, Cabir)

When translated into Spanish, 'The Caribbean' is Carib, which is also the name for the first known Bluetooth virus that targets mobile phones. While the virus itself is relatively harmless, it represents a new era for virus writers and a new set of worries for mobile phone owners.

While a virus can be written for a single, unconnected device, such a program will have a very short life. To keep the virus 'alive', it must be able to spread from one device to another. This part of the virus process is called the 'infector' or 'infection vector'.

While infectors have changed and evolved over the years, the most common infector is via email. Once infected, a virus either uses its own email engine, or borrows the victim's email program to send out copies of itself to email addresses it finds on the victim's system. Other infectors include floppy disks and pirated software. New infectors are very rare simply because there are a very limited number of new methods of communication.

In 1999 a new protocol was designed, developed, and implemented into numerous mobile devices. Known as Bluetooth, the wireless communications protocol was built to facilitate the transmission of information within a relatively small range (about 10 meters). Operating in the 2.45 GHz range, Bluetooth uses spread spectrum frequency hopping (1600 hops/s) to reduce the risk of interference between one Bluetooth device and another. By changing the frequency 1600 times a second, it is very unlikely two devices in the same area will conflict with each other.

What makes Bluetooth really useful is that it automatically handles the connection process between two devices. By scanning the surround area for other Bluetooth devices, and by setting up a connection between discovered devices, a piconet is created. This piconet spreads from device to device, thus creating a self-maintaining web of connected devices.

While this concept is great for those who are honest, in the last couple years several vulnerabilities have been found within the protocol that make it possible to extract information from users, send users unsolicited messages, and perform other miscellaneous attacks. However, up until now these 'attacks' have been used only in labs or to send unsolicited messages to other Bluetooth devices (toothing).

Understanding Carib

Carib is the first known virus/worm that uses Bluetooth as an infector. While the worm itself is basically harmless, it has recently been coupled with other malware to create "blended" threats.

In fact, Carib is already a worm/Trojan by definition. Once a device is infected, it scans the airwaves for other Bluetooth enabled devices until it finds one (and only one). It then tries to send the new target device a copy of itself.

At this point the victim is prompted with a dialogue box as to whether or not they would like to accept the incoming file. If the victim accepts, the file will transfer and another prompt will be presented warning the victim that the no supplier could be verified. If the victim hits the ‘Yes’ key, the device will prompt the victim one final time if it should install the program. At this point the device is infected.

Once infected, Bluetooth is enabled on the device, in case the file is downloaded or transferred via an alternate route. Then a splash screen is presented on the victim’s device with the message “Caribe-VZ/29a”. Finally, the infected devices starts scanning for any new Bluetooth enabled devices that it can infect.

Of note, Carib will only transfer the file once. In addition, only Nokia Series 60 phones appear to be vulnerable, at least according to an internal memo written by Symbian. Finally, this worm requires user interaction three separate times, which means it would take a naïve user with a complete lack of viruses knowledge for Carib to spread. That is, until it was coupled with other malware.

Mosquito Trojan

The Symbian operating system powers many cellular phones, and also supports a wide range of third party applications – including games. Unfortunately, one popular game turned out to have a “cracked” version that was secretly infected with a Trojan horse. The Mosquito Dialer Trojan infects the popular game “Mosquito” with code that secretly messages pay-per-call numbers.

What is a dialer Trojan?

A dialer trojan is malware coded to secretly dial phone numbers, thus leaving the infected victim with a large phone bill. There are two reasons why someone might code and spread a dialer trojan. The first reason is destructive; perhaps as tool of revenge. The second reason is for financial gain. Simply set up a premium 900 number and charge $5.99 a minute. Then, all the malicious coder needs is a few hundred infected victims to make a decent amount of money.

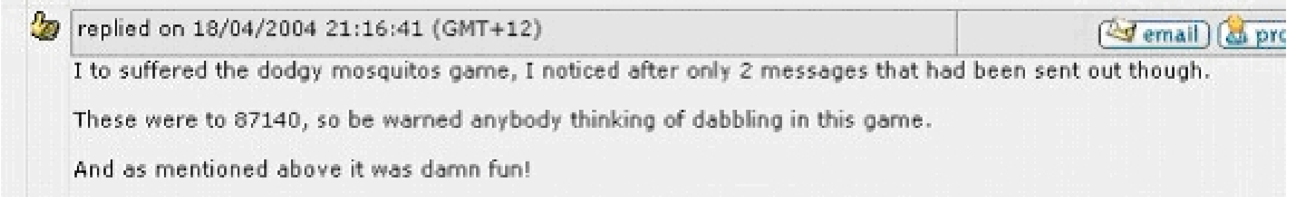
As a desktop analogy, many free porn websites use browser-based exploits to infect PC users with dialer trojans. This is a classic example of dialer trojans being used for financial gain. Dialer trojans have been around on PCs for many years. Traditional PC dialer trojans rely on the infected computer having a working modem; and the modem needs to be connected to a wall socket. It was only a matter of time before someone realized that coding dialer malware for computers that mostly rely on broadband was a waste of time. Such malware coders have now moved to cellular phones.

Symbian-based cellular phones offer the ability to run far more code than earlier cellphones. Cellular phones can now be used to play games, surf the Internet and perform many other activities traditionally done from a desktop computer. While these features are useful for consumers, it also means that malware coders have an increasing amount of scope in which to apply malicious code. Code that once only worked on desktop computers can now be easily ported to work on handheld devices running the Symbian OS.

Mosquito, the game that plays the player

We first heard of the Mosquito dialer trojan, while researching online. Various web forum users where complaining that they had installed a game and now their phones where sending text messages to the number 87140 (Fig. 2 below). But some users had not noticed this problem, so it was evident that there were at least two discrete versions of this game in circulation — and that at least one of these versions was malware.

The game that the infected users had installed was called Mosquito v2.0. The game is unique in that it makes use of the phone’s built-in camera. The user walks around shooting mosquitoes in a virtual reality-like atmosphere. This game appeared to be a “cracked” version that appeared on the many cell phone warez and p2p networks that plague the Internet underground. It appeared that 87140 was a UK number costing a hefty £1.50 per text message.



replied on 18/04/2004 21:16:41 (GMT+12)

I to suffered the dodgy mosquitos game, I noticed after only 2 messages that had been sent out though.

These were to 87140, so be warned anybody thinking of dabbling in this game.

And as mentioned above it was damn fun!

Fig. 2: Screenshot of the first known mention of an infected Mosquito game (source – tinyurl.com/4ght6)

How to detect the malicious version of Mosquitoes v2

Cellular phone malware is a relatively new phenomenon. There were no clear instructions that we knew of for dissecting Symbian malware, and we had no prior experience with this platform. But we have developed a successful antivirus scanner for a similar platform (Windows Mobile/Pocket PC), and we have written some papers on ARM-based reverse engineering.

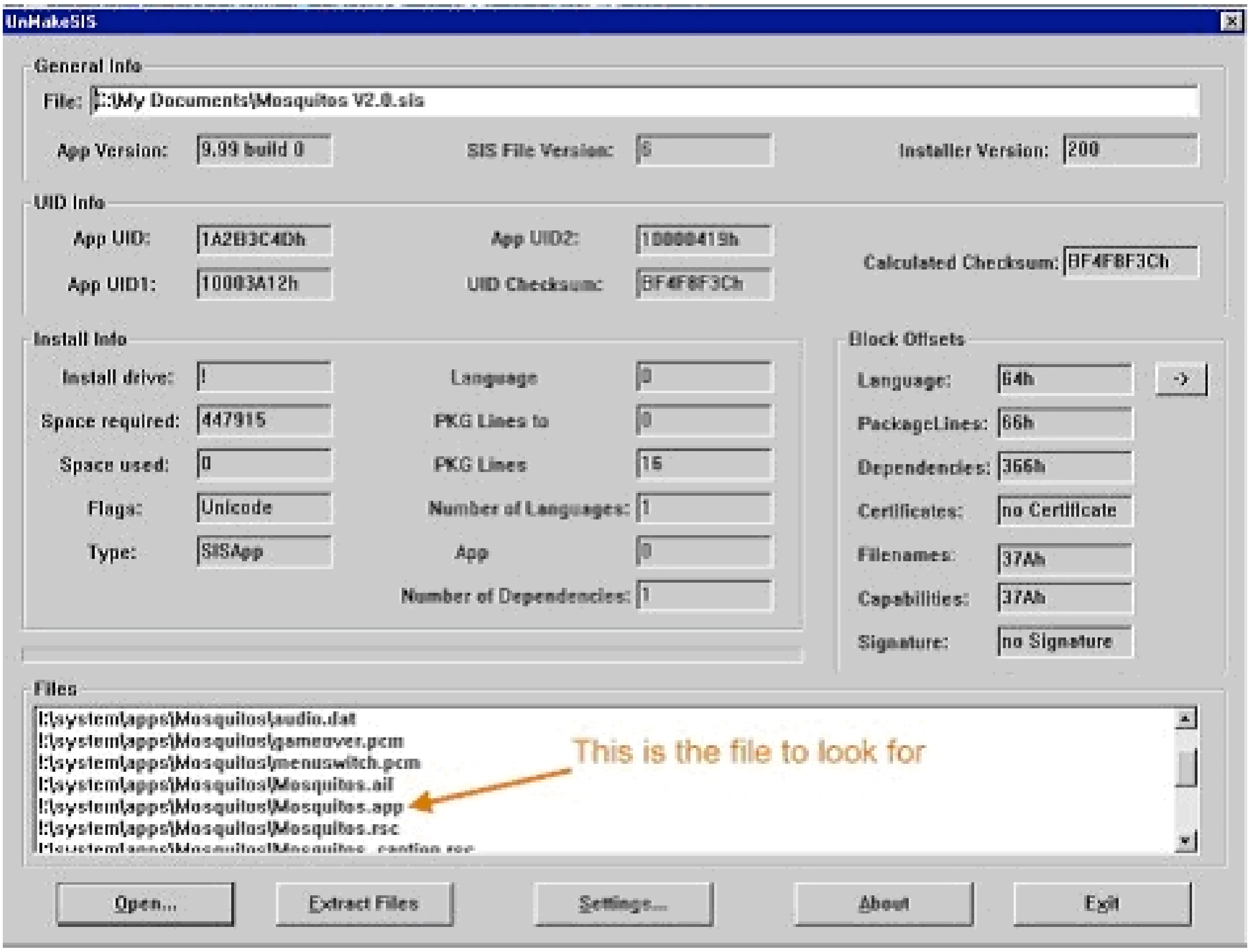
So out of curiosity, we decided to download the infected warez and see if we could take a look under the hood. Hopefully, by reporting our findings here, we will inspire others to take the analysis further.

For every instance of this trojan that we have encountered, the file is packed as a .sis file type. Specific tools are needed to view the contents of a .sis file on a PC. Most of the tools are freeware and are easily available.

SisView (www.dalibor.cz/epoc/sisview.htm) is a freeware plug-in that has been created for the shareware program Total Commander (www.ghisler.com/download.htm). This tool allows you to view the contents of any .sis file. System admins can view .sis files that are stored on their servers to see if they have the .nfo files often associated with cracker group releases (pirated software). The malware version of the mosquito game is cracked, so using this method could help in initial detection.

Unmakesis (mitglied.lycos.de/atzplzw/) is a freeware tool for unpacking a Symbian .sis file. With Unmakesis, analyzing and extracting Symbian sis files is relatively easy. When using Unmakesis it is important to set your screen resolution to 1024x768; the program does not dynamically adjust to your screen size and you may find that you cannot use the functions needed to extract the sis file.

Using Unmakesis on the mosquito file we can see that this file contains Mosquitos.app (Fig. 3).



File: C:\My Documents\Mosquitos V2.0.sis
App Version: 9.99 build 0
SIS File Version: 6
Installer Version: 200
App UID: 1A2B3C4Dh
App UID1: 10003A12h
App UID2: 10000419h
UID Checksum: BF4F8F3Ch
Calculated Checksum: BF4F8F3Ch
Install drive: !
Space required: 447915
Space used: 0
Flags: Unicode
Type: SISApp
Language 0
PKG Lines to 0
PKG Lines 16
Number of Languages: 1
App 0
Number of Dependencies: 1
Language: 64h
PackageLines: 66h
Dependencies: 366h
Certificates: no Certificate
Filenames: 37Ah
Capabilities: 37Ah
Signature: no Signature

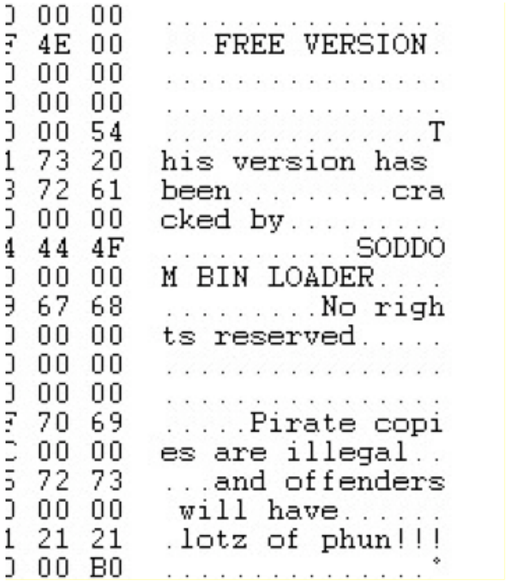
```
!:\system\apps\Mosquitos\audio.dat
!:\system\apps\Mosquitos\gameover.pcm
!:\system\apps\Mosquitos\menuswitch.pcm
!:\system\apps\Mosquitos\Mosquitos.aif
!:\system\apps\Mosquitos\Mosquitos.app
!:\system\apps\Mosquitos\Mosquitos.rsc
!:\system\apps\Mosquitos\Mosquitos_caption.rsc
```

This is the file to look for

Fig. 3: Using Unmakesis to extract the Mosquitos.app file

Extracting the Mosquitos.app file onto your computer allows you to view the code of this file using the following two basic techniques. The first technique used to view the file is hex editing. Using a simple hex editor (Fig. 4) allows you to access written comments and hex code that the file contains. Looking at the code closely we find the following line:

Free Version cracked by Soddom bin Loader

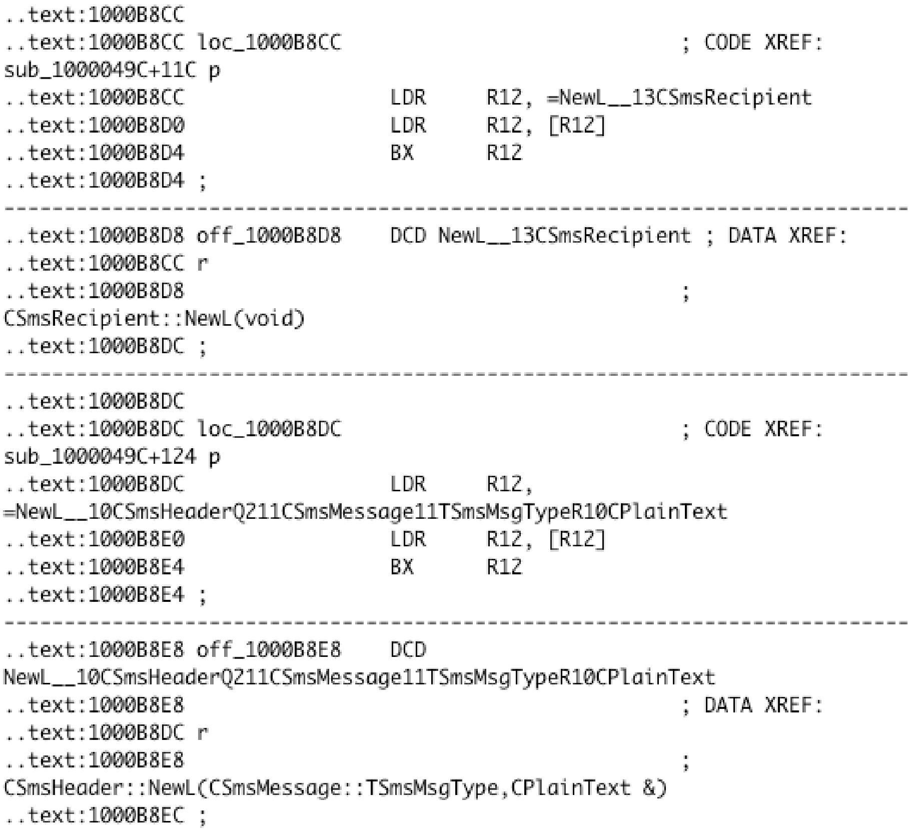


```
0 00 00   ................
F 4E 00   ...FREE VERSION.
0 00 00   ................
0 00 00   ................
0 00 54   ...............T
1 73 20  his version has 
3 72 61  been.........cra
0 00 00  cked by.........
4 44 4F   ...........SODDO
0 00 00  M BIN LOADER....
9 67 68   .........No righ
0 00 00  ts reserved.....
0 00 00   ................
0 00 00   ................
F 70 69   .....Pirate copi
C 00 00  es are illegal..
5 72 73   ...and offenders
0 00 00    will have......
1 21 21   .lotz of phun!!!
0 00 B0   ...............°
```

Fig. 4: Selected hex dump of Mosqitos.app showing what appears to be the malware author

Finding the SMS call routine in the Mosquitos.app file

Using a hex editor gives us some preliminary information on the file, but it does not give us enough information to prove that this file is a malicious dialer. What we need is a more complex debugging tool. The tool we highly recommend is IDA Pro (www.datarescue.com/idabase/). IDA comes up with the following SMS call routines; these are just a few selected examples; the SMS routines make up 5 pages of data, which are too large to post here:



```
..text:1000B8CC
..text:1000B8CC loc_1000B8CC                            ; CODE XREF:
sub_1000049C+11C p
..text:1000B8CC                 LDR     R12, =NewL__13CSmsRecipient
..text:1000B8D0                 LDR     R12, [R12]
..text:1000B8D4                 BX      R12
..text:1000B8D4 ;
---------------------------------------------------------------------------
..text:1000B8D8 off_1000B8D8    DCD NewL__13CSmsRecipient ; DATA XREF:
..text:1000B8CC r
..text:1000B8D8                                         ;
CSmsRecipient::NewL(void)
..text:1000B8DC ;
---------------------------------------------------------------------------
..text:1000B8DC
..text:1000B8DC loc_1000B8DC                            ; CODE XREF:
sub_1000049C+124 p
..text:1000B8DC                 LDR     R12,
=NewL__10CSmsHeaderQ211CSmsMessage11TSmsMsgTypeR10CPlainText
..text:1000B8E0                 LDR     R12, [R12]
..text:1000B8E4                 BX      R12
..text:1000B8E4 ;
---------------------------------------------------------------------------
..text:1000B8E8 off_1000B8E8    DCD
NewL__10CSmsHeaderQ211CSmsMessage11TSmsMsgTypeR10CPlainText
..text:1000B8E8                                         ; DATA XREF:
..text:1000B8DC r
..text:1000B8E8                                         ;
CSmsHeader::NewL(CSmsMessage::TSmsMsgType,CPlainText &)
..text:1000B8EC ;
```

As you can see, the malicious game uses SMS routines. That makes it one of the first documented trojans written specifically for cellular phones. At the least, it is the first Symbian-based cellular phone dialer Trojan of which we know.

Looking further at the following code snippet, the phone number 87140 is clearly visible, along with other numbers that may also be SMS text targets.

a9222	1000BA84
a4636	1000BA90
a87140	1000BA9C
a33333	1000BAA8

There is no need for this game to use SMS routines. When combined with multiple user reports of surreptitious, paid text messages, it appears to be a legitimate threat.

The Brador Trojan

Brador was the first backdoor Trojan for Pocket PC. It gives full, wireless remote control of your Pocket PC to a remote hacker who might even be on the other side of the world. This is problematic if, for example, you are a physician with patient medical records on your Pocket PC. Or, you may be a corporate executive or network admin who, like many, use your Pocket PC over a VPN to control your wired infractructure – thus potentially giving a remote hacker total control of your enterprise network.

Brador is successful in part because the Pocket PC operating system doesn't come with a native process monitor. Without a process monitor (such as the Win32 Task Manager), it can be difficult to detect and remove this Trojan and any future Trojans. Pocket PC lacks this feature; when a user attempts to delete the malicious file, the system presents an error message saying that the program is in use. In this case, it might seem that the only way to remove the Trojan is a hard factory reset (similar to formatting the hard drive on a desktop PC). Fortunately, there are now third party tools that can do this without needing a hard reset.

Dust Virus

The following is the virus writer’s description of the significant technical obstacles that prevented anyone from successfully infecting Windows CE for nearly 4 years.

“While programming WinCE4.Dust, I used time-tested techniques from the Win32 world. When infecting, the PE file is altered in the following way: The last section size is increased by the virus code size, and the virus body is copied at the end of the last section. Then the new EntryPoint is set; in other words, the pointer is set to the first instruction to execute when the program is loaded. This way, it's guaranteed that the virus will be run.

Because Dust doesn't use the host's import section, it has to somehow obtain the needed API function addresses. This was the biggest problem to overcome, and finding the solution took some time. As soon as we have the function addresses, we use them to alter the victim's files found on the memory medium. Finding files to infect provides the standard function pair FindFirstFile and FindNextFile. Together with CreateFile, they differ from their Win32 counterparts, which appeared to be another minor problem.

Every file gets mapped into memory, where later needed modifications are made. Windows CE introduced a new function, CreateFileForMapping, that has no equivalent on Win32. Without calling this function, there's no way to get the file handle that could be used to create the mapping object. On the other hand, the advantage of the ARM ISA appeared – the automatic generation of position-independent code. On Win32 x86, you had to determine its actual memory position and use this value later to modify absolute variable addresses (if the host's relocations were not altered, of course).

The virus source code includes deeper comments of given problems and techniques. Please take the time to carefully read through the comments of this source code, in which I explain the Windows CE .NET security weakness that allowed me to create the first successful virus for this platform.”

For those who are interested in further details, the full, annotated source code has been published by Pearson Education, Inc. InformIT. The URL is tinyurl.com/8zmqn

Evolution of “blended” mobile threats

The last few months have seen a rapid evolution of “blended” mobile malware. Much of this activity has been seen on the Symbian Smartphone platform. For example, “Skulls” was the second trojan to infect Symbian Series 60 smart phones (the first was Mosquito, discussed above). When launched, the application claims to be an “Extended Theme Manager by Tee-222.” However, it then disables all other applications on the phone and replaces their icons with a skull and crossbones. Worse, it was more recently merged with Caribe to form the first “crossover” malware for smartphones.

Skulls and Caribe also merged to form Metal Gear, a trojan that masqerades as the game with the same name. Metal Gear uses Skulls to deactivate the devices antivirus. Thus, it was the first anti-AV malware for Symbian phones. The malware also drops SEXXXY.sis to the device, an installer that adds code to disable the handset menu button. The Trojan then uses Caribe to transmit itself to new devices.

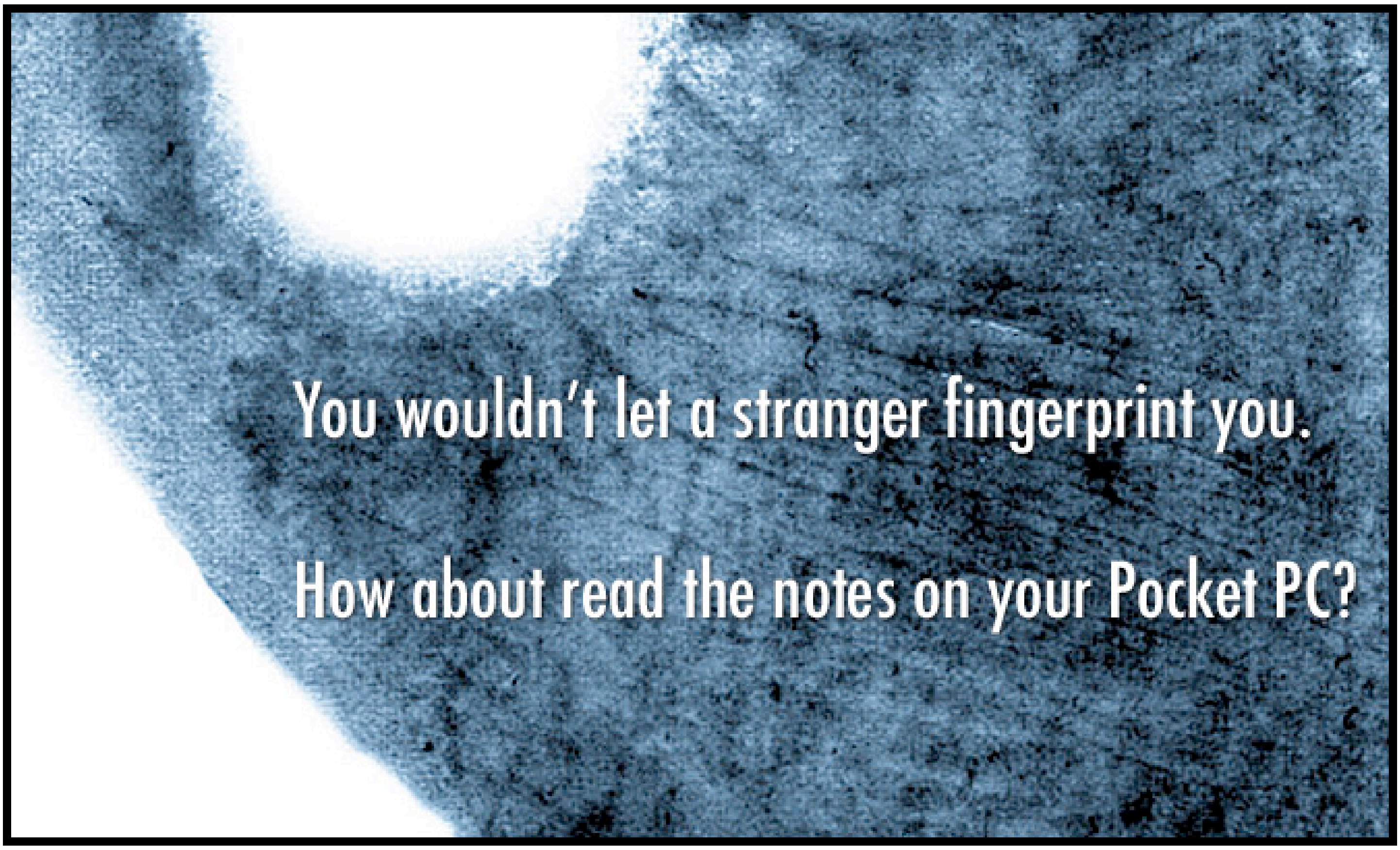
Another example of blending is the Gavno.a Trojan, which is spread via a file called patch.sis (it masquerades as phone patch). Gavno uses a malformed file to crash an internal Symbian process, thus disabling the phone.

The effect is to disable all handset buttons and to completely prevent the user from making calls. It may also cause a continual rebooting loop. It is only 2kb in size, and it has already seen variants merged with Caribe to spread to other phones. Other examples of viral evolution include the following:

- Dampig trojan: Notable in that it corrupts the system uninstallation settings, making it more difficult to remove.
- Mabir virus: Similar to Cabir, but instead of Bluetooth it uses SMS to spread.
- Commwarrior: also tries to disable the onboard antivirus software.
- Frontal virus: causes a total system crash of the phone until it is removed

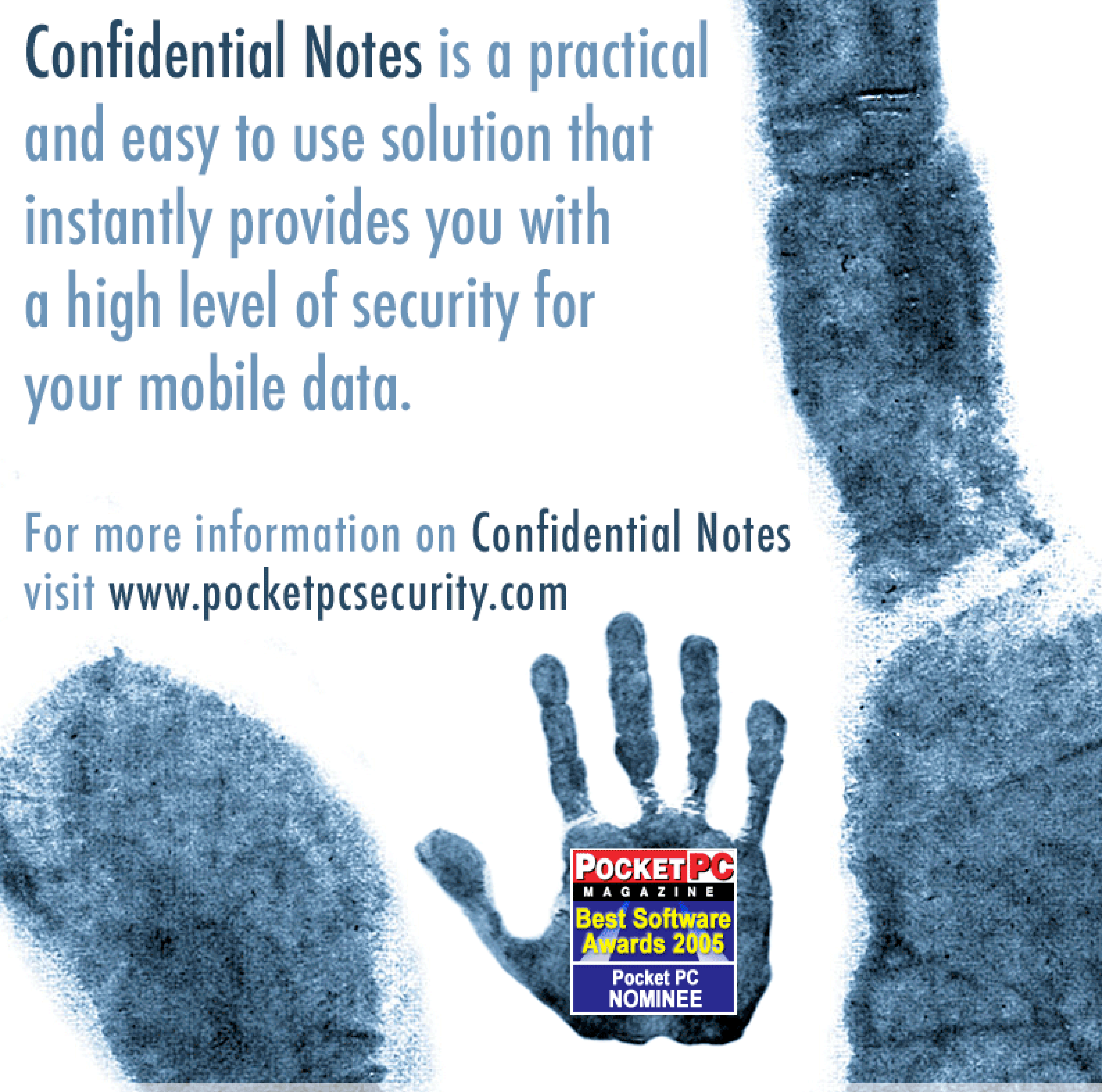
Lastly, a new Symbian Trojan called Doomboot-A that now loads a Commwarrior variant when it infects Smartphones. Doomboot-A destroys the boot process so that the phone is not useable.

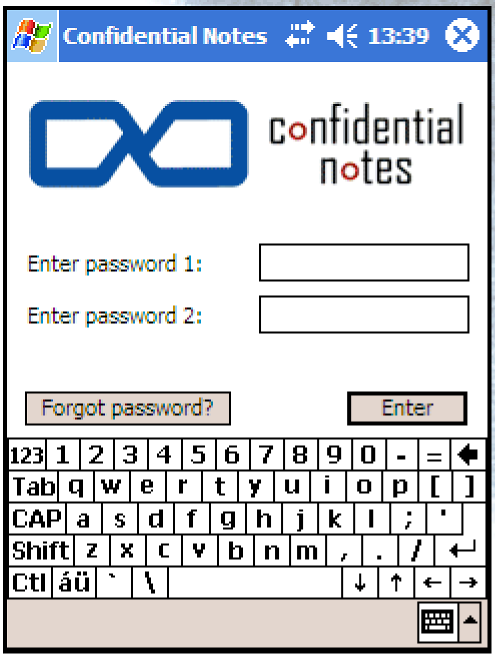
Dr. Cyrus Peikari is the founder of Dallas-based Airscanner Corporation, which produces technologically advanced security software for PDAs and Smartphones. Dr. Peikari finished his undergraduate training with honors in electrical engineering from Southern Methodist University in 1991. He also worked as a telecommunications software engineer for Alcatel before receiving his Doctor of Medicine degree from Southwestern in 1995. Dr. Peikari has co-authored five technical books on information security, including "Maximum Wireless Security" from SAMS and "Security Warrior" from O'Reilly. Dr. Peikari is also a frequent speaker at major information security conferences, including RSA, Defcon, NetSec, CISSPcon and CSI. He holds several patents and patents pending in the mobile security field. Dr. Peikari has helped several universities start brand new infosec degree programs. He is also the Site Host for Security at Pearson Education's InformIT.com division, where his infosec articles are read by millions of network administrators and programmers per year.


You wouldn’t let a stranger fingerprint you.
How about read the notes on your Pocket PC?

Confidential Notes is a practical and easy to use solution that instantly provides you with a high level of security for your mobile data.

For more information on Confidential Notes visit www.pocketpcsecurity.com

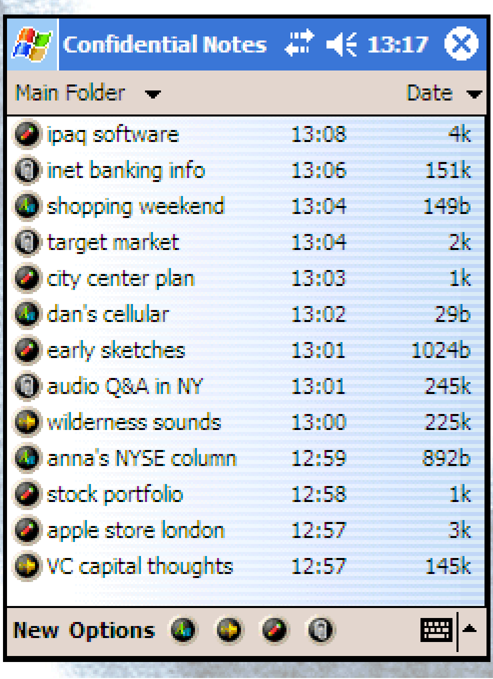




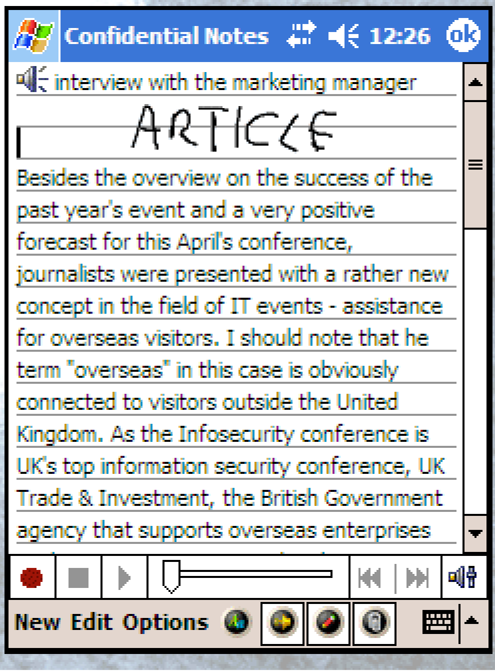
Enter password 1:

Enter password 2:

Forgot password?
Enter
123 1 2 3 4 5 6 7 8 9 0 - =
Tab q w e r t y u i o p [ ]
CAP a s d f g h j k l ; '
Shift z x c v b n m , . /
Ctl áü ` \ ↓ ↑ ← →



ipaq software	13:08	4k
inet banking info	13:06	151k
shopping weekend	13:04	149b
target market	13:04	2k
city center plan	13:03	1k
dan's cellular	13:02	29b
early sketches	13:01	1024b
audio Q&A in NY	13:01	245k
wilderness sounds	13:00	225k
anna's NYSE column	12:59	892b
stock portfolio	12:58	1k
apple store london	12:57	3k
VC capital thoughts	12:57	145k



interview with the marketing manager

ARTICLE

Besides the overview on the success of the past year's event and a very positive forecast for this April's conference, journalists were presented with a rather new concept in the field of IT events - assistance for overseas visitors. I should note that he term "overseas" in this case is obviously connected to visitors outside the United Kingdom. As the Infosecurity conference is UK's top information security conference, UK Trade & Investment, the British Government agency that supports overseas enterprises

Security resources

F-Secure Antivirus Research Weblog

http://www.f-secure.com/weblog/

This blog is run by the F-Secure Antivirus Research Team whose most prominent member is certainly Mikko Hypponen, well known in the security community. This is a great place to look for up-to-date information on new viruses and Trojans as well as interesting facts as they are discovered. The team provides a fresh look at the antivirus world free of marketing hype and filled with texts, screenshots and sometimes videos. Other antivirus vendors could learn a lot from the F-Secure team.

A Day in the Life of an Information Security Investigator

http://blogs.ittoolbox.com/security/investigator/

Interesting stories of apparently real life events that sometimes make you think. The description of the author is: “Follow an Information Security Investigator as he recounts his unique experiences working with federal, corporate, and military institutions and provides his perspective on the security issues impacting the IT industry today.”

SpywareInfo

http://www.spywareinfo.com/

This website is the perfect resource for information on spyware. In the words of the authors: “Has some sleazy web site taken over your browser? Are you getting pop up ads even when your browser has been closed for some time? Are you infected with a spyware program that refuses to go away? If so, our message board has dozens of dedicated volunteers ready to give you step-by-step assistance to remove the malicious software and regain control of your PC.”

Financial Cryptography

https://www.financialcryptography.com/

This is a blog that covers a topic of interest to a specific audience and it does it well. Updated regularly it provides links to stories and commentaries that anyone interested in cryptography should find interesting. The authors publish also original content on the blog. To give you an idea of what to expect here are two titles of the papers linked there: “On Secure Knowledge-Based Authentication”, “An Introduction to Petname Systems”.

Send your favorite security websites and blogs to editor@insecuremag.com for consideration.

Build a custom firewall computer

By Nicholas Petreley and Jono Bacon

As more and more computers are getting plugged into the Internet, the risk factor associated with an online presence has also risen. The increase in hours online combined with the propagation of always-on broadband and high-speed cable/DSL Internet access has resulted in the need to secure even simple, one-computer home networks. As a result, the humble firewall has become a must-have item as opposed to a could-have item in a network.

The basic aim of a firewall is to keep unwanted people off of your network. The virtual wall of fire is essential in keeping out crackers who want to invade your security, as well as blocking the growing armies of worms, viruses, and other Internet nasties that crawl the Web looking for computers to exploit. The situation is very bad; an unprotected Windows machine can become infected in as little as four minutes after it is put on the Internet. If you are considering a firewall but are uncertain you want to put the effort into it, ask a friend who has one for a list of attempted intrusions on his network. You will probably be surprised by the frequency of attacks. My own firewall logged more than 100 attempted intrusions in the first few hours after I put it up.

Both software and hardware firewalls are available. Software firewalls are installed on each desktop on the network, and they protect that single machine. The hardware approach is to use a dedicated machine to protect the entire network from malicious traffic. This hack explores a dedicated firewall Linux distribution called SmoothWall, which you can install on an aging computer to provide a dedicated firewall appliance to protect your entire network. After the initial setup, you will find your SmoothWall box to be invaluable.

Gather the Ingredients

To create a SmoothWall firewall appliance, you need a computer to use. Anything from a ‘486 with 16MB of RAM on up is fine, but if you want to keep several days’ worth of log files, I recommend you use at least a 4GB disk. You also need at least two Linux-supported network cards in the computer.

Here is how you will use your network cards:

- If you have a cable/DSL modem that plugs into a network card, you need a card for this. This card is referred to as the RED interface.

- You need a network card to connect to the internal network. If you have more than one computer on your internal network, this interface is usually plugged into a hub, switch, or wireless access point. This card is referred to as the GREEN interface.

- If you have any computers that need to be accessed publicly, you need another network card for these. This card is referred to as the ORANGE interface and also is known as the snazzily titled De-Militarized Zone (DMZ), because it exists in a sort of no man's land between the public Internet and your private network.

You should install the cards you need in the computer, download the SmoothWall ISO from www.smoothwall.org, and then burn the ISO to CD.

The next step is to boot from the CD and install the SmoothWall software. If you cannot boot from the CD, try using the Smart Boot Manager. If this does not work, you can create a series of boot floppies from the files found in the images directory on the SmoothWall CD. There you'll find two boot floppy images called bootdiskone-x.x.img and bootdisktwo-x.x.img. Use **dd** to create the floppies (unmounting and changing the floppy between images, of course):

```
foo@bar:~$ dd if=bootdiskone-x.x.img
of=/dev/fd0 bs=1024 conv=sync ; sync
foo@bar:~$ dd if=bootdisktwo-x.x.img
of=/dev/fd0 bs=1024 conv=sync ; sync
```

If you need to create the floppies on a Windows system, you can use the rawrite program (uranus.it.swin.edu.au/~jn/linux/rawwrite.htm) to create the disks. Installing SmoothWall is a fairly simple process, but you need to know how you want your network to be set up in terms of IP addresses. Within the setup routine are a Networking section and an Addresses subsection. You set the IP addresses for each interface here. For example, a common setting for the GREEN interface is the IP address 192.168.0.1 and the network mask 255.255.255.0. The RED interface is typically set to DHCP to grab your Internet IP address from the cable modem, but you should check with your ISP to see how the cable modem gets its IP address. The other setting to configure is in the "DNS and Gateway settings" section. Set this to 192.168.0.1. Now you have your firewall set up as your Internet gateway that other machines can refer to when requiring Internet access.

Configure the Firewall

Once the SmoothWall firewall is installed, you can access it in two main ways. The most common and popular way is to access its special web-based interface, which is available on port 81. So, if your firewall's IP address is 192.168.0.1, you can access the web interface at 192.168.0.1:81. SmoothWall's default configuration does not allow access from outside your internal network, so you cannot make changes to it from work or while traveling.

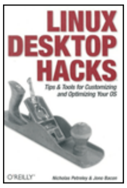
When you access the web interface, you are asked for the administrator password for the machine (which you created when you installed SmoothWall) and then you can configure it. Within the web panel is a huge range of options and features that you can configure. These options are grouped into categories which are visible at the top of the page.

If you need to do something that is not accessible in the web interface, you can use the included Java SSH applet to log in to the machine and type in commands to an SSH shell.

Enable Port Forwarding

A common requirement when running a network of machines is the need to have a connection from outside the firewall serviced by a machine inside the firewall (usually in the DMZ). This is the scenario for those who run a web or email server and need to have the relevant ports accessible to the outside world. When a computer connects to your IP address/domain, the first computer that receives the connection is the firewall. Because it is unlikely you are running a web or email server on the firewall itself (if you are, you really shouldn't be because bugs in these programs can compromise the security of the firewall) you need a method to get that request to the computer that can handle it. This is where port forwarding comes in. Its purpose is to take the request for a service and forward it to the specific machine on the network that can service the request.

To do this with SmoothWall, access the web interface and select Networking → Port Forwarding. You can leave the external source IP box blank if you want to accept all connection requests for the port in question (this is commonly the case for a public service such as web serving). In the Source Port box, specify the port you want to forward (such as port 80 for a web server). Finally, you can enter the destination computer IP address and its port number in the other two boxes. This is quite useful if you want to forward a normal port 80 connection to a machine with a different port number, such as port 8080; a common request with Apache virtual hosts. Once you have forwarded your ports, you need to select the External Services Access page and add the ports you have forwarded to that page. This enables access to the ports from outside the network. SmoothWall is proven to be an incredibly capable and flexible firewall. Because of this a lot of organizations and homes use it to protect their networks. Although the GPL version of the firewall is very capable, the commercial version and its included support can be really useful for commercial organizations. Both versions give you the flexibility of a powerful and supported firewall that can protect a network of Linux, Windows, or Mac OS X machines.


LINUX DESKTOP HACKS
Tips & Tools for Customizing and Optimizing Your OS
O'REILLY
Nicholas Petreley & Jono Bacon

Excerpted from “Linux Desktop Hacks” by Nicholas Petreley, Jono Bacon (ISBN: 0596009119). Copyright 2005, O'Reilly Media, Inc. www.oreilly.com All rights reserved.

Software spotlight

WINDOWS – Acunetix Web Vulnerability Scanner
http://www.net-security.org/software.php?id=633

Acunetix WVS first crawls the whole website, analyzes in-depth each file it finds, and displays the entire website structure. After this discovery stage, it performs an automatic audit for common security vulnerabilities by launching a series of web attacks.

LINUX – WifiScanner 1.0.0
http://www.net-security.org/software.php?id=381

WifiScanner is an analyzer and detector of 802.11b stations and access points. It can listen alternatively on all the 14 channels, write packet information in real time, can search access points and associated client stations, and can generate a graphic of the architecture using GraphViz.

MAC OS X – iStumbler
http://www.net-security.org/software.php?id=620

iStumbler is a free, open source tool for finding AirPort networks, Bluetooth devices and now mDNS services with your Mac. iStumbler combines a compact user interface with a real time display of signal strength and complete debugging information.

POCKET PC – Crippin
http://www.net-security.org/software.php?id=544

Crippin was designed to protect confidential files in case a Pocket PC is lost or stolen. It's been designed to be:

* small (minimal executable size).
* miserly (with respect to storage required both during and after encryption).
* secure (using RSA encryption).

If you want your software title included in the HNS Software Database e-mail us at software@net-security.org

Lock down your kernel with grsecurity

By Andrew Lockhart

Hardening a Unix system can be a difficult process. It typically involves setting up all the services that the system will run in the most secure fashion possible, as well as locking down the system to prevent local compromises.

However, putting effort into securing the services that you're running does little for the rest of the system and for unknown vulnerabilities. Luckily, even though the standard Linux kernel provides few features for proactively securing a system, there are patches available that can help the enterprising system administrator do so. One such patch is grsecurity (www.grsecurity.net).

grsecurity started out as a port of the OpenWall patch (www.openwall.com) to the 2.4.x series of Linux kernels. This patch added features such as nonexecutable stacks, some filesystem security enhancements, restrictions on access to **`/proc`**, as well as some enhanced resource limits. These features helped to protect the system against stack-based buffer overflow attacks, prevented filesystem attacks involving race conditions on files created in **`/tmp`**, limited a user to only seeing his own processes, and even enhanced Linux's resource limits to perform more checks. Since its inception, grsecurity has grown to include many features beyond those provided by the OpenWall patch. grsecurity now includes many additional memory address space protections to prevent buffer overflow exploits from succeeding, as well as enhanced **`chroot()`** jail restrictions, increased randomization of process and IP IDs, and increased auditing features that enable you to track every process executed on a system. grsecurity adds a sophisticated access control list (ACL) system that makes use of Linux's capabilities system.

This ACL system can be used to limit the privileged operations that individual processes are able to perform on a case-by-case basis.

To compile a kernel with grsecurity, you will need to download the patch that corresponds to your kernel version and apply it to your kernel using the patch utility.

For example, if you are running Linux 2.4.24:

```
# cd /usr/src/linux-2.4.24
# patch -p1 < ~andrew/grsecurity-1.9.13-2.4.24.patch
```

While the command is running, you should see a line for each kernel source file that is being patched. After the command has finished, you can make sure that the patch applied cleanly by looking for any files that end in .rej.

The patch program creates these when it cannot apply the patch cleanly to a file. A quick way to see if there are any .rej files is to use the find command:

```
# find ./ -name \*.rej
```

If there are any rejected files, they will be listed on the screen. If the patch applied cleanly, you should be returned back to the shell prompt without any additional output.

After the patch has been applied, you can configure the kernel to enable grsecurity's features by running **`make config`** to use text prompts, **`make menuconfig`** for a curses-based interface, or **`make xconfig`** to use a Tk-based GUI. If you went the graphical route and used make **`xconfig`**, you should then see a dialog similar to Figure 1–1. If you ran **`make menuconfig`** or **`make config`**, the relevant kernel options have the same name as the menu options described in this example.

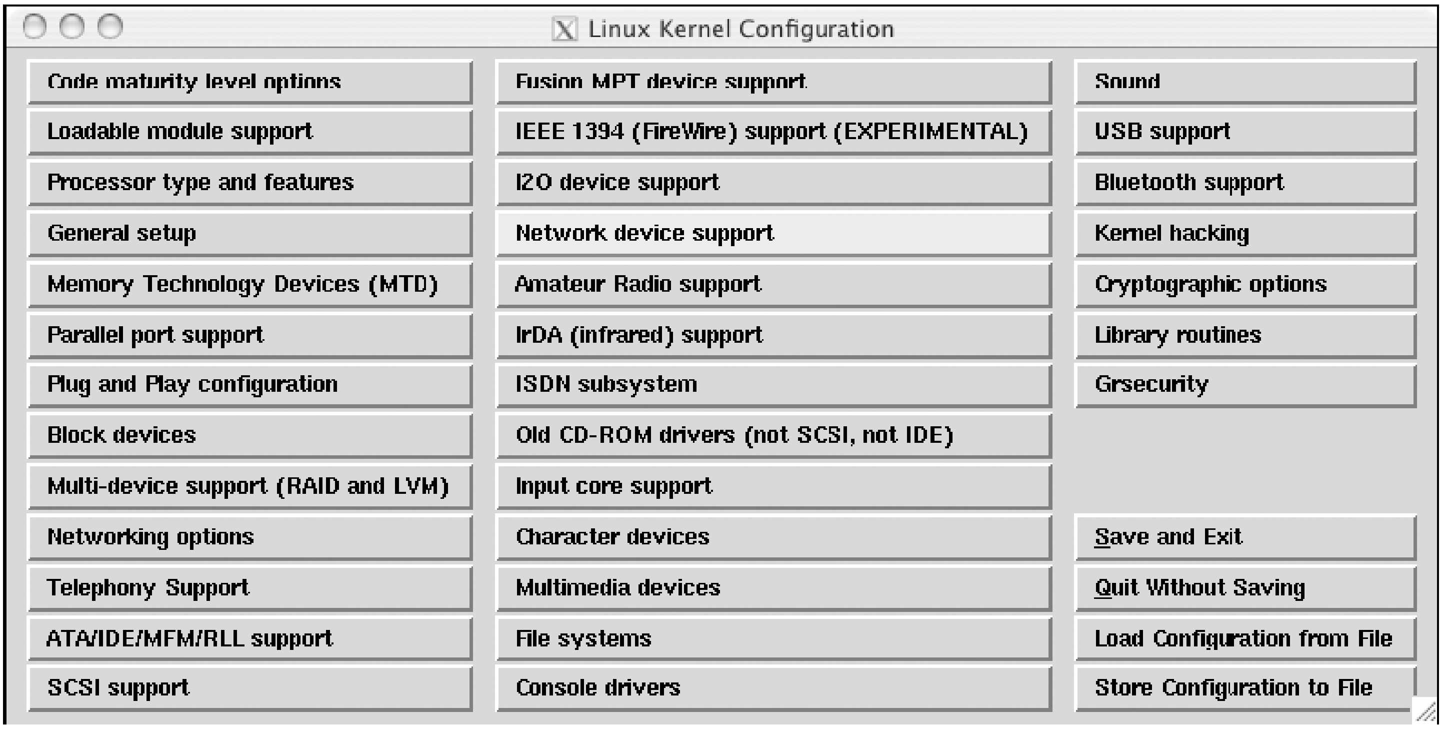


Figure 1–1. Linux kernel configuration after the grsecurity patch has been applied.

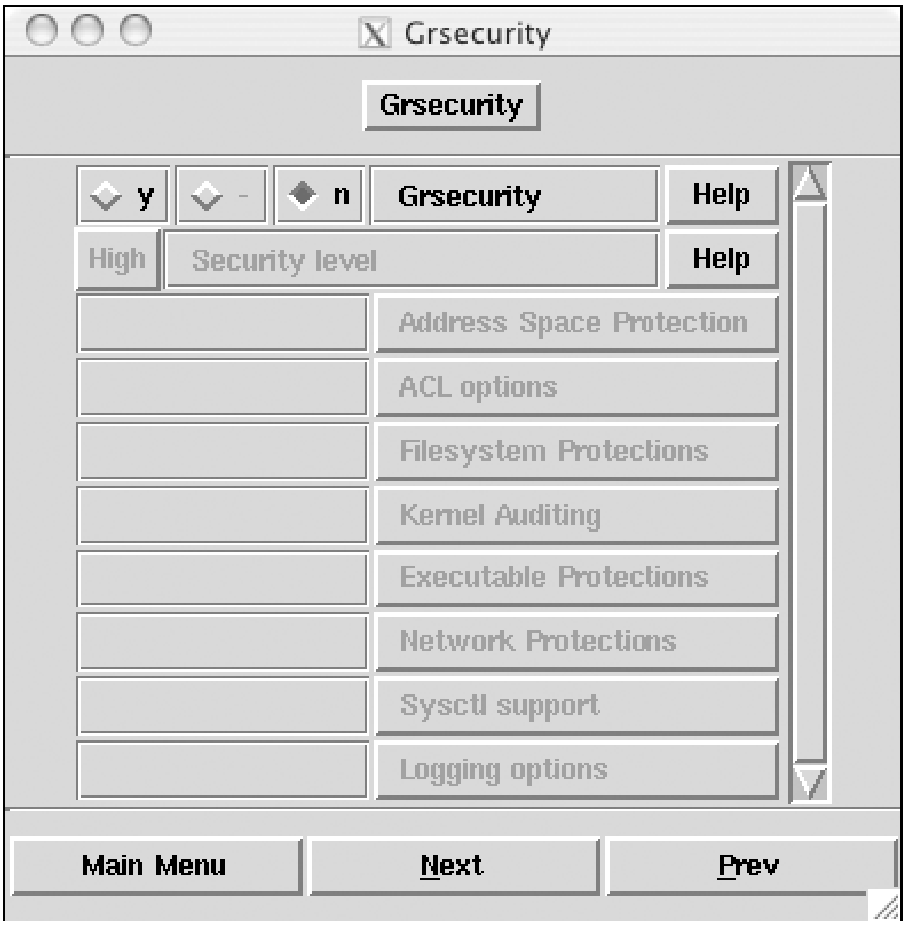
To configure which grsecurity features will be enabled in the kernel, click the button labeled Grsecurity. After doing that, you should see a dialog similar to Figure 1–2.

To enable grsecurity, click the y radio button. After you've done that, you can enable predefined sets of features with the Security Level drop-down list, or set it to Custom and go through the menus to pick and choose which features to enable.

Choosing Low is safe for any system and should not affect any software's normal operation. Using this setting will enable linking restrictions in directories with mode 1777. This prevents race conditions in **`/tmp`** from being exploited, by only following symlinks to files that are owned by the process following the link. Similarly, users won't be able to write to FIFOs that they do not own if they are within a directory with permissions of 1777.

In addition to the tighter symlink and FIFO restrictions, the Low setting increases the randomness of process and IP IDs. This helps to prevent attackers from using remote detection techniques to correctly guess the operating system your machine is running, and it also makes it difficult to guess the process ID of a given program.

The Low security level also forces programs that use **`chroot()`** to change their current working directory to **`/`** after the **`chroot()`** call. Otherwise, if a program left its working directory outside of the **`chroot`** environment, it could be used to break out of the sandbox.



◇ y ◇ - ◆ n Grsecurity

High Security level

Figure 1–2. The grsecurity configuration dialog

Choosing the Low security level also prevents nonroot users from using **dmesg**, a utility that can be used to view recent kernel messages.

Choosing Medium enables all of the same features as the Low security level, but this level also includes features that make **chroot()**-based sandboxed environments more secure. The ability to mount filesystems, call **chroot()**, write to **sysctl** variables, or create device nodes within a **chrooted** environment are all restricted, thus eliminating much of the risk involved in running a service in a sandboxed environment under Linux. In addition, TCP source ports will be randomized, and failed **fork()** calls, changes to the system time, and segmentation faults will all be logged. Enabling the Medium security level will also restrict total access to **/proc** to those who are in the wheel group. This hides each user's processes from other users and denies writing to **/dev/kmem**, **/dev/mem**, and **/dev/port**.

This makes it more difficult to patch kernel-based root kits into the running kernel. Also, process memory address space layouts are randomized, making it harder for an attacker to successfully exploit buffer overrun attacks. Because of this, information on process address space layouts is removed from **/proc** as well. Because of these **/proc** restrictions, you will need to run your **identd** daemon (if you are running one) as an account that belongs to the wheel group. According to the grsecurity documentation, none of these features should affect the operation of your software, unless it is very old or poorly written.

To enable nearly all of grsecurity's features, you can choose the High security level. In addition to the features provided by the lower security levels, this level implements additional **/proc** restrictions by limiting access to device and CPU information to users who are in the wheel group. Sandboxed environments are also further restricted by disallowing **chmod** to set the SUID or SGID bit when operating within such an environment. Additionally, applications that are running within such an environment will not be allowed to insert loadable modules, perform raw I/O, configure network devices, reboot the system, modify immutable files, or change the system's time.

Choosing this security level will also cause the kernel's stack to be laid out randomly, to prevent kernel-based buffer overrun exploits from succeeding. In addition, the kernel's symbols will be hidden – making it even more difficult for an intruder to install Trojan code into the running kernel - and filesystem mounting, remounting, and unmounting will be logged.

The High security level also enables grsecurity's PaX code, which enables nonexecutable memory pages. Enabling this will cause many buffer overrun exploits to fail, since any code injected into the stack through an overrun will be unable to execute. However, it is still possible to exploit a program with buffer overrun vulnerabilities, although this is made much more difficult by grsecurity's address space layout randomization features. PaX can also carry with it some performance penalties on the x86 architecture, although they are said to be minimal. In addition, some programs – such as XFree86, wine, and Java virtual machines – will expect that the memory addresses returned by **`malloc()`** will be executable. Unfortunately, PaX breaks this behavior, so enabling it will cause those programs and others that depend on it to fail. Luckily, PaX can be disabled on a per-program basis with the **`chpax`** utility (chpax.grsecurity.net).

To disable PaX for a program, you can run a command similar to this one:

```
# chpax -ps /usr/bin/java
```

There are also other programs that make use of special GCC features, such as trampoline functions. This allows a programmer to define a small function within a function, so that the defined function is only in the scope of the function in which it is defined. Unfortunately, GCC puts the trampoline function's code on the stack, so PaX will break any programs that rely on this. However, PaX can provide emulation for trampoline functions, which can be enabled on a per-program basis with chpax, as well by using the **`-E`** switch.

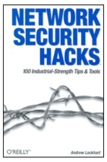
If you do not like the sets of features that are enabled with any of the predefined security levels, you can just set the kernel option to "custom" and enable only the features you need.

After you've set a security level or enabled the specific options you want to use, just recompile your kernel and modules as you normally would. You can do that with commands similar to these:

```
# make dep clean && make bzImage
# make modules && make
modules_install
```

Then reboot with your new kernel. In addition to the kernel restrictions already in effect, you can now use **`gradm`** to set up ACLs for your system.

As you can see, grsecurity is a complex but tremendously useful modification of the Linux kernel. For more detailed information on installing and configuring the patches, consult the extensive documentation at www.grsecurity.net/papers.php.



Excerpted from "Network Security Hacks" by Andrew Lockhart, (ISBN: 0–596–00643–8). Copyright 2005, O'Reilly Media, Inc. www.oreilly.com All rights reserved.

Interview with Sergey Ryzhikov

By Berislav Kucan

Please introduce yourself, your company and your line of work.

My name is Sergey Ryzhikov and I am the Director of Bitrix (www.bitrixsoft.com). Today's base line of our company is the development of the Bitrix Site Manager software. The Bitrix Site Manager is a content management solution (CMS) that enables users to effectively manage a corporate web site. We are technology partners with more than 400 Web-development companies in more than 30 countries.

Almost on a daily basis we are hit by information on new security vulnerabilities affecting popular web applications. In your opinion, what are the current trends related to security issues in web based applications and what can be done to minimize the threat?

Today, attacks on Web-based businesses make up more than 50 percent of all IT crimes. It's obvious that company web sites will continue to be major targets for Internet criminals.

First of all, there are non-profit hacks targeting web sites of large corporations and state institutions. Such hacks usually aim to produce self-advertisement and hit the company image. In addition to this, there are also a large number of commercial hacks. Cases like this are rarely discovered right away, since the purpose of these hacks is information theft.

Undoubtedly, hacks of corporate web sites affect both the image and reputation of these companies. When a hack becomes public knowledge, these incidents become extremely troublesome. The greatest concern of these companies, however, is the loss of sensitive information about the organization and its clients, and the potential for blackmailing of proprietors, all of which can lead to direct material losses.

What differentiates Bitrix Site Manager from other similar content management systems?

It is a common practice to develop a stand-alone CMS or an e-store without content management capabilities. Different Web-development companies produce stand-alone Web statistics analyzers, forums, and technical support systems. As a result, these components offer different user profiles, different security policies, and different user interfaces.

Not only is this inconvenient, but it prohibits growth and development of Internet projects and increases the cost of implementation and ownership. The result is obvious: companies cannot run their e-businesses effectively.

We create a package which includes everything that a modern web site might require, and then some. Our software covers virtually every business objective related to e-business. Today we offer 19 modules that share a unified style of architecture and a common user interface. This set of modules satisfies the business-requirement needs of 95% of all web sites.

There is one serious advantage of our product that I can't pass over in silence. Any edition of the Bitrix Site Manager allows for multiple sites to be built using a single product copy. These sites can be managed, maintained, and analyzed in a centralized fashion, using a common interface. As companies master the Internet, their sites grow and become more developed; this further implies that it's only sensible to use a single software package to manage all of the sites.

Your company launched a security section of your web site, showing your current and potential users that security is one of your top priorities. What security mechanisms and options does Bitrix Site Manager offer?

Indeed, security issues are of utmost importance to us and our clients. As you mentioned, we have launched on our site a special section solely devoted to security issues. In this section, we have spotlighted common categories of vulnerabilities, and explained how we have implemented systems within our product to protect the client's security.

While designing the Bitrix Site Manager architecture, we have placed special emphasis on security-related problems. Allow me to list the major features which will enable Web developers to increase the security of their Web projects:

- unified authorization system – all permissions can be assigned to user groups only;
- user profile is common for all modules;
- two-level system of access-level delimitation;
- an access-control system which is independent of the business logic;
- information can be encrypted during transmission;
- the SiteUpdate system;
- independent journaling of executed pages in the Statistics module;
- special policy of handling the volatile and external data;
- double-checking of unsafe code portions.

The methods by which we have implemented these features in our product, as well as the features themselves, should indicate that our system is secure. Nevertheless, we have engaged an independent auditing company to verify this. Presently, all the updates that we publish are subjected to an independent audit.

Your software recently undergone an independent security audit. If any, what kind of potential security risks the auditors came across?

We have called for engineers of the Positive Technologies company to conduct an independent security audit of the Bitrix Site Manager. Their task was to check the system architecture and how well the security precautions are implemented.

The Positive Technologies company carries out commercial "hacks" of Internet projects at the resource owner's request. They also produce a range of software products for monitoring the information environment security.

The vulnerability analysis was conducted using both the system source codes and a real web project built upon the system. This has provided for thorough "internal" and "external" testing, which resulted in an exhaustive and trustworthy analysis.

The audit results induced a range of updates. Right after the press conference, we have issued an update which helped to strengthen all modules. This significantly improved the administrative sections' security against specific attacks – which I must say, were very exotic attacks. Furthermore, none of the other existing Web solutions provide any protection against such attacks. Only a handful of developers are informed of these, though their risk and threat is very high. We have quickly responded to the recommendations of the Positive Technologies company and implemented the necessary means for protection.

As a result of the final audit stage, the Positive Technologies company has issued a certificate entitled "The Secure Web Application". As a matter of fact, this is the first certificate of this kind issued by the Positive Technologies company. I must add that we have signed a treaty to constantly monitor the security of updates. Finally, the Positive Technologies company has made a decision to move their site to the Bitrix Site Manager. We consider this decision to be an honor, coming from such a professional and experienced team as the Positive Technologies company.

Bitrix Site Manager uses SiteUpdate technology which allows seamless upgrading possibilities. Automatic updates can be a security risk, how is this taken care of within your software?

The SiteUpdate security methods are paramount. This technology provides for downloading and installing the latest product updates without having to enlist the services of specially educated engineers. This significantly improves the security of the site. By a simple button click, the client's site engine connects to our server via the license key and provides the system updates that our client needs.

To prevent downloading illegal updates, our clients may use the SSL certificate. This ensures that updates will be downloaded encrypted from our server only. But, as HTTPS requires special libraries to be installed on the client's side, we have set the default transport mode for updates to HTTP.

Some open source content management systems are (in)famous for their security problems. Are commercial content management systems more secure because there is money involved in the development and therefore this phase is done in a more professional way, or their main advantage is that everyone can't poke around their source code and hunt for vulnerabilities?

Many open source content management systems pay little attention to security. This can be the result of decentralized management and development of such systems, frequent change of developers, and the lack of unified security environment architecture.

We consider security as one of the key factors in product development. We invest much effort and money into security analysis and constant audit.

If you would run a webserver for your company, would you use Linux or a Windows based system?

Our product is developed with PHP using MySQL as a free database or Oracle as a commercial database.

The Bitrix Site Manager runs on any software platform, such as UNIX, Linux, Sun, FreeBSD, HP-UX and Windows. The Apache web server is mostly used, even on the Windows platform. Most of our clients install the system on Linux or FreeBSD.

At the same time, many projects having a high attendance are implemented on the Windows platform. In my opinion, this is due to the fact that the UNIX system is capable of serving heavier load than the Windows systems, given similar hardware resources. Traditionally, experienced administrators install FreeBSD or Linux because they consider these systems more secure and efficient.

In general, Oracle is used for large-scale projects with heavy traffic (more than 50.000 unique visitor per day or 500.000 pages), or if the web resource hosts large information content (for example, over 100.000 articles in on-line magazines).

Generally, Oracle based projects are more secure and reliable, and can be easily scaled to serve unlimited loads. Some of the clients simply make Oracle their corporate standard and choose only the Oracle editions of the product.

By the way, our clients can migrate from MySQL to Oracle in few hours by simply converting their databases while leaving the function and design of the pages and applications untouched.

What are the plans for the future product development?

Recently we have introduced a new module: "Active Directory/LDAP Integrator", which provides more simple and comfortable integration of the client's sites in their intranet and extranet networks built upon Windows or UNIX/Novell. We have launched a new redesigned beta version of the Search module with support for English morphology. Search has become faster and more precise.

This year we plan to release a new version – 4.1. It will implement a new visual editor featuring Firefox support and a full-featured visual editor for design templates. This will help to make the first steps in mastering the Bitrix Site Manager easier.

The new version will include a rich set of functions that provide support for selling digital content, manager of the client's virtual accounts, and a new version of the e-store. This version will include updates of the Statistics, Newsletter, Helpdesk, and other modules. Without a doubt, this new version will come under the banner of security! Version 4.1 is not a full stop, but just a stage. We keep extending the task list according to the dialogue with our partners and clients.

This year we plan to release the first .NET edition of the product in the "Standard" and "Web Analytics" editions. The new edition will be based on conceptually different architecture, though. This will allow our partners to offer their clients a complete range of solutions for all popular platforms and databases.

Best practices for database encryption solutions

By Ulf Mattsson

Encryption can provide strong security for data at rest, but developing a database encryption strategy must take many factors into consideration. Encryption at the database level, versus application level and file level has proved to be the ideal method to protect sensitive data and deliver performance.

Organizations must balance between the requirement for security and the desire for excellent performance. Building and maintaining a secure and efficient cryptographic engine is not the easiest task. This is a specialized and complex solution area and if internal resources don't have the cryptography expertise in relation to IT environment, outside expertise should be used to ensure superior performance.

The best practice, in nearly all cases, is to use an engine that's already available and tested. Packaged database encryption solutions have proven to be the best alternative to protect sensitive data. There is a multitude of techniques and alternative topologies for encryption at the database level. In real-world scenarios, are complex issues and experts should be used who understand all available options and the impact for each particular customer environment. Encryption engines and services come in three flavors: central, local and dedicated. In a straight comparison of costs, Local Encryption Services are generally cheaper but not secure. Dedicated Encryption Services provides high availability with key caching and real cpu offloading. Benchmarks in customer environments demonstrated the criticality of making the right selection between the different topologies for database encryption implementations.

This article reviews the performance aspects of three dominant topologies for database encryption and offers detailed guidance on scalable implementations of data at rest encryption in an enterprise environment, including encryption, key management, backup, auditing and logging should be deployed to optimize security, performance, scalability, and administration.

Introduction

Building and maintaining a secure and efficient cryptographic engine is not the easiest task. The best practice, in nearly all cases, is to use an engine that's already available and tested. In a straight comparison of costs, Local Encryption Services are generally cheaper but not secure. Dedicated Encryption Services provides high availability with key caching and real cpu offloading.

Benchmarks in customer environments demonstrated the criticality of making the right selection between the different topologies for database encryption implementations. A central topology benchmarked decryption of only a few hundred database rows per second and a more distributed hybrid topology benchmarked in the range of million rows per second, typically needed in environments with a high volume OLTP or parallel systems for decision support. Be aware that exposing encryption services as a network resource will introduce an additional point of attack, and very limited scalability in a database environment. Private keys should be stored encrypted with several AES encryption keys that are nested within a hierarchy in which each key is protected by a parent key. This multi-layer hierarchy of keys ensures the highest level of protection against attack. Engines come in three flavors: central, local and dedicated. Not protected properly, stored unprotected in a software environment, and unprotected in server memory, keys are vulnerable to discovery. What ís needed? The best protection against private key compromise is a combination of physical security and key management technology, including stringent security standards throughout the private key lifecycle.

Data at rest encryption – different approaches have its advantages and disadvantages

There are many architectures, techniques, and tools available to Security and IT organizations to ensure security and performance are balanced and optimized. Each of these approaches has its advantages and disadvantages. Database security is a wide research area and includes topics such as statistical database security, intrusion detection, and most recently privacy preserving data mining. In this work we have addressed and evaluated the most critical issue for the success of encryption in databases, performance. To achieve that, we have analysed different solution alternatives. Each topology effects security and performance differently and has advantages and disadvantages.

THE BEST PROTECTION AGAINST PRIVATE KEY COMPROMISE IS A COMBINATION OF PHYSICAL SECURITY AND KEY MANAGEMENT TECHNOLOGY, INCLUDING STRINGENT SECURITY STANDARDS THROUGHOUT THE PRIVATE KEY LIFECYCLE

Issues with application level encryption and storage level encryption

Application-layer encryption requires rewrite of existing applications that is impractical due to limited IT resources, lack of access to source code, or a lack familiarity with old code. Rewriting applications is also very costly, risky and introduces an implementation time delay factor. Lastly, all applications that access the encrypted data must also be changed to support the encryption/decryption model. Storage-layer encryption alone can only protect against a narrow range of threats, namely media theft and storage system attacks.

Database-layer encryption

Database-layer encryption protects the data within the DBMS and also protects against a wide range of threats, including storage media theft, well known storage attacks, database-layer attacks, and malicious DBAs. Deployment at the column level within a database table, coupled with access controls will prevent theft of critical data.

Different dimensions of database encryption support

There are three main dimensions to encryption support in databases:

- One is the granularity of data to be encrypted or decrypted. The field, the row and the page, typically 4KB, are the alternatives. The field is the best choice, because it would minimize the number of bytes encrypted. However, as we have discovered, this will require methods of embedding encryption within relational databases or database servers.

- The second dimension is software versus hardware level implementation of encryption algorithms. Our results show that the choice makes significant impact on the performance. We have discovered encryption within relational databases based on hardware level implementation of encryption algorithms entail a significant start up cost for an encryption operation. Each model also offers different operational performance, tuning possibilities, and encryption offloading capabilities. The loss of granular protection will impact the security level.

- The third dimension is the location of the encryption service – local service, remote procedure service, or network attached service. Choosing the point of implementation not only dictates the work that needs to be done from an integration perspective but also significantly affects the overall security model.

Not always practical to encrypt data as soon as it enters the network

The sooner the encryption of data occurs, the more secure the environment however, due to distributed business logic in application and database environments, it is not always practical to encrypt data as soon as it enters the network. Encryption performed by the DBMS can protect data at rest, but you must decide if you also require protection for data while it's moving between the applications and the database. We considered several possible combinations of different encryption approaches, namely; software and hardware level encryption, and different data granularity. We started with software encryption at field level. We then developed search acceleration support to index encrypted fields, and experienced a low performance overhead when searching on encrypted fields, including primary index fields. We finally directed our research and experiments to hardware level encryption use only for master key encryption.

Algorithm performance and security

Initially we considered several encryption algorithms AES, RSA and Blowfish for the implementation. We conducted experiments using these algorithms and found that the performance and security of the AES algorithm is better than the RSA implementation and the Blowfish algorithm implementation. AES is fast, compared to other well-known encryption algorithms such as DES. DES is a 64-bit block cipher, which means that data is encrypted and decrypted in 64-bit chunks. This has implication on short data. Even 8-bit data, when encrypted by the algorithm will result in 64 bits.

Optimizing database-level encryption implementations

Database-level encryption allows enterprises to secure data as it is written to and read from a database. This type of deployment is typically done at the column level within a database table and, if coupled with database security and access controls, can prevent theft of critical data. Database-level encryption protects the data within the DBMS and also protects against a wide range of threats, including storage media theft, well known storage attacks, database-level attacks, and malicious DBAs. Database-level encryption eliminates all application changes required in the application-level model, and also addresses a growing trend towards embedding business logic within a DBMS through the use of stored procedures and triggers. Since the encryption/decryption only occurs within the database, this solution does not require an enterprise to understand or discover the access characteristics of applications to the data that is encrypted. While this solution can certainly secure data, it does require some integration work at the database level, including modifications of existing database schemas and the use of triggers and stored procedures to undertake encrypt and decrypt functions.

Performance aspects of database-level encryption

Additionally, careful consideration has to be given to the performance impact of implementing a database encryption solution, particularly if support for accelerated index-search on encrypted data is not used. First, enterprises must adopt an approach to encrypting only sensitive fields. Second, this level of encryption must consider leveraging hardware to increase the level of security and potentially to offload the cryptographic process in order to minimize any performance impact. The primary vulnerability of this type of encryption is that it does not protect against application-level attacks as the encryption function is strictly implemented within the DBMS. If we compare the response time for a query on unencrypted data with the response time for the same query over the same data, but with some or all of it encrypted, the response time over encrypted data will increase due to both the cost of decryption as well as routine and/or hardware invocations. This increase is referred to as the encryption penalty. An observation according to recent studies is that, different fields have different sensitivity. It is possible for Hybrid to support encryption only on selected fields of selected tables.

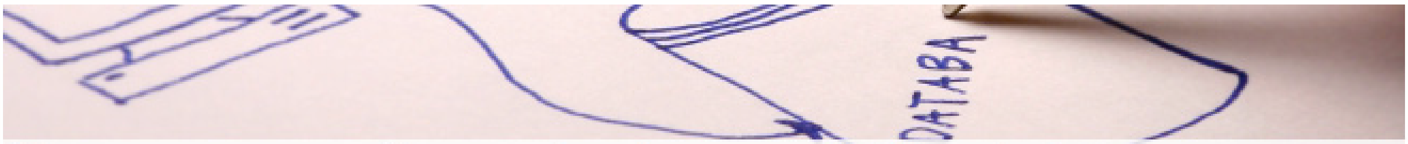
Encryption, by its nature, will slow down most SQL statements. If some care and discretion are used, the amount of extra overhead should be minimal. Also, encrypted data will have a significant impact on your database design. In general, you want to encrypt a few very sensitive data elements in a schema, like Social security numbers, credit card numbers, patient names, etc. Some data values are not very good candidates for encryption — for example booleans (true and false), or other small sets like the integers 1 through 10. These values along with a column name may be easy to guess, so you want to decide whether encryption is really useful. Creating indexes on encrypted data is a good idea in some cases. Exact matches and joins of encrypted data will use the indexes you create. Since encrypted data is essentially binary data, range checking of encrypted data would require table scans. Range checking will require decrypting all the row values for a column, so it should be avoided if not tuned appropriately with an accelerated search index.

Central, local or dedicated encryption services

The cryptographic engine lies at the core of the infrastructure. It implements algorithms, such as AES, DES, RSA, SHA, upon which the rest of the system depends. Any request to encrypt, decrypt, hash, or sign data ultimately passes through the engine. The management of encryption keys are the foundation of all encryption-based security solutions. Each of these approaches has its advantages and disadvantages. Database security is a wide research area and includes topics such as statistical database security, intrusion detection, and most recently privacy preserving data mining. Users wishing to access data will now securely access it using the privacy management infra structure instead of developing multiple customized solutions for each application and data storage system. Applications and databases would not be impacted by an application specific implementation. This would alleviate the problem of maintaining the software and administrating privacy for each specific application and database.

Local encryption services

A Local Encryption Service is one in which the cryptography occurs on the same cpu as the rest of the application ís processing. Typically for a Local Encryption Service, the algorithms are included as part of the application or as a library to which the application is linked. Examples of Local Encryption Services include RSA's Crypto-J, Cryptix, Mcrypt, and encryption libraries and toolkits included in products from some database vendors.



The management of encryption keys are the foundation of all encryption-based security solutions.

Local encryption services are generally cheaper but not secure

A Local Encryption Service is often easier to implement as there is no need to configure a separate hardware interface, and since there is no special hardware or software to purchase, the Local Encryption Service can be significantly less expensive, even free. Of course, the keys used by the engine are also significantly less secure. The application's performance will also suffer to a great extent by the inclusion of cryptographic processing on the application's cpu, and potential requirements in searching of encrypted database columns. Compared to a Dedicated Encryption Service, a Local Encryption Service offers a reduction in complexity in some areas. A Dedicated Encryption Service requires the installation and configuration of a separate layer of hardware. Local Encryption Services avoid this. However, Dedicated Encryption Services typically store the key away from the data and application and then wrap the key in layers of encryption and optionally use tamper-resistant hardware.

Approximating this level of security using a Local Encryption Service requires a comprehensive and mature implementation. In a straight comparison of costs, Local Encryption Services are generally cheaper but not secure. The cost savings need to be balanced against security and performance issues.

Dedicated encryption services

Dedicated Encryption Services contain separate processes dedicated just to cryptography. Dedicated Encryption Services could also contain a set of separated cpu's dedicated just to cryptography. A typical example of a Dedicated Encryption Service is Protegrity Secure.Data Server in which the cryptographic processes are mounted. Another example of a Dedicated Encryption Service is a Hardware Security Module (hsm) in which the cryptographic cpu is mounted within a standalone, typically tamper-resistant, enclosure. An hsm might communicate with the application through a pci card, scsi, or over ipc (inter process communication). The Central Encryption Services may contain a number of separate cpu's dedicated just to cryptography. To maintain a high level of security the server platform should only contain securely encrypted lower level data encryption keys. Key encryption master keys should always be stored separately outside the server platform. Private keys should be stored encrypted with several AES encryption keys that are nested within a hierarchy in which each key is protected by a parent key. This multi-layer hierarchy of keys ensures the highest level of protection against attack.

Central encryption services

Central Encryption Services can be implemented as a remote server with an optional HSM, or a network appliance (NAED). A typical example of a Central Encryption Service is a Server with a Hardware Security Module (hsm) in which the cryptographic cpu is mounted within a standalone, typically tamper-resistant, enclosure. An hsm might communicate with the application over the local area network using ssl. The goal of an hsm is to provide a secure environment for keys thus most hsms combine the functionality of a key vault and engine.

Relying only on remote central encryption support is not satisfactory, since it would always penalize network and total system performance, and more importantly, it is likely to open new security holes in the form of attacks on network exposed encryption services.

The Network Attached Encryption Device (NAED)

The Network Attached Encryption (NAED) is implemented as a Network Attached Encryption Appliance that scales with the number of Network Attached Encryption Appliances available. A NAED is a hardware device that resides on the network, houses the encryption keys and executes all crypto operations. This topology has the added security of physically separating the keys from the data. However, this added security comes with a heavy price; performance can be 10–100 times worse than alternative methods. The benchmarks showed a throughput of between 440 and 1,100 row-decryptions per second. In prior work with IBM Research we addressed some critical performance issues when using HSM support. A coming paper will address the security exposure with API level attacks when using HSM support, including Network Attached Encryption Appliances.

A NAED IS A HARDWARE DEVICE THAT RESIDES ON THE NETWORK, HOUSES THE ENCRYPTION KEYS AND EXECUTES ALL CRYPTO OPERATIONS

Encryption as a network resource – a new point of attack

Be aware that exposing encryption services as a network resource will introduce an additional point of attack. An integrated central and distributed solution can protect from this vulnerability. Also, look for industry standard api support. Adopting a standard such as pkcs#11, will help ease the transition from one vendor's engine to another, and in some cases between different engines from the same vendor.

Denial of Service attacks

A network attached engine, on the other hand, does not provide high availability, unless multiple engines are configured into a high availability cluster. Denial of Service attacks are another related concern with network attached engines. Since the engine is available over tcp/ip, an attacker could flood the engine with traffic and block legitimate cryptographic requests. If required information can't be decrypted, then a customer may not be able to place an order or access account information. If the database stored encrypted records that are critical for the business operation, then a successful denial of service attack could be severe. None of the above are reasons not to use Dedicated Encryption Services, but rather factors to keep in mind when selecting a Dedicated Encryption Service.

Interface to encryption hardware add overhead

Central/Dedicated Encryption Services may be used in environments where performance and scalability are not critical requirements. A Dedicated Encryption Service is typically a specially constructed device which is connected via a cable to the computer needing cryptographic services, including pci or scsi, for directly connected engines, or ethernet for network connected encryption services. When selecting a Dedicated Encryption Service, consider performance, scalability, and availability. Most Dedicated Encryption Services will perform cryptographic operations faster than Local Encryption Services for larger blocks of data, however the interface to the hardware can add considerable overhead for shorter blocks of data. This overhead may be noticeable on directly connected engines at higher transaction volumes. Network based engines, though, might carry a performance penalty from the need to negotiate a secure tcp connection. If the connection remains open between requests, then the overhead may be lower but it certainly should be tested.

The myth that NAEDS off-load work from the database

This example debunks a well-publicized myth, that NAEDs off-load work from the database. There isn't an off-load of work since this solution must perform one encryption operation in the database, which is the same for other topologies, in addition to the encryption functions at the NAED.

There are three points of overhead with this topology

Let's explore a simple example to demonstrate the overhead; a user requests 500,000 rows of encrypted data.

When a user requests secured data, the security system manages the process of retrieving encrypted data from the database, ensuring that the request is from an authorized user, and performing the decryption process.

In this topology, the encryption agent handles the request and retrieves the encrypted data from the database. It sends the encrypted data over the network to be decrypted by the NAED.

Inside the NAED are the keys and the algorithms to decrypt the data. However once decrypted, we have clear-text information that needs to be sent back over the wire to the database server. This requires that we re-secure the information for transit, typically through a secure communication process such as SSL. When the data arrives at the agent on the database server, it has to be returned to clear-text, and then it is served up to the calling application.

1. A NAED topology has three points of encryption versus one for other methods. In the example above, the 500,000 rows of data are sent over the wire to be decrypted at the NAED. The clear text is then encrypted using SSL to send back over the network and decrypted at the database to be served in clear text to the application.

2. Network overhead is caused by sending all 500,000 rows over the network to be decrypted by the NAED and then must return over the network to the database.

3. The NAED is a stateless device and needs to be initialised/set-up before each row is decrypted. In this simple example, the NAED is set-up 500,000 times. The set-up has a large overhead.

The Network Attached Encryption Device (NAED) topology has proven in tests, due to the three points of overhead, to perform by an order of magnitude, worse than alternative structures. Each round trip over the network is roughly 1 millisecond per row. In the example above this would be 500,000 x 1ms = 500 seconds compared to 1-25 seconds with alternative topologies.

A mature hybrid solution

A powerful Hybrid solution combines the benefits of a Central Encryption Service with a Dedicated Encryption Service, on a general purpose computers running a standard operating system, but stripped of all but the most essential services. Amongst those services would be a cryptographic server and a key storage module. To maintain a high level of security the server platform should only contain securely encrypted lower level data encryption keys. Key encryption master keys should always be stored separately outside the server platform on the central encryption services platform. Private keys should be stored encrypted with several AES encryption keys that are nested.

The Hybrid solution Central Encryption Services provides secure and flexible key management and key backup. Dedicated Encryption Services provides high availability with key caching and real cpu offloading. Packaged and Integrated Local Encryption Services operations provides the highest operational performance and the highest availability for encryption services.

All data items and keys are not equal

Some data requires a higher level of protection. Data classification can be used to determine if a specific data item should be processed locally, in a dedicated service, central service, or on a hsm. Risk management can help in defining the balance between these requirement for security, cost, and acceptable performance and scalability.

Some encryption keys requires a higher level of protection. Master keys and some data encryption keys requires a higher level of protection. Data classification can be used to determine if a specific data item encryption key should be processed locally, in a dedicated service, central service, or on a hsm. Risk management can help in defining the balance between these requirements for security, cost, and acceptable performance and scalability.

A high level of security

To maintain a high level of security the server platform should only contain securely encrypted lower level data encryption keys. Key encryption master keys should always be stored separately outside the server platform on the central encryption services platform. While most Dedicated Encryption Services are devices specifically constructed for cryptography, some Dedicated Encryption Services might be general purpose computers running standard OSes, but stripped of all but the most essential services. Amongst those services would be a cryptographic server and a key storage module. At the heart of the server is a library such as the ones used for a Local Encryption Service. For that reason, these types of Dedicated Encryption Services. Private keys should be stored encrypted with several AES encryption keys that are nested within a hierarchy in which each key is protected by a parent key. This multi-layer hierarchy of keys ensures the highest level of protection against attack.

Effective key protection in memory

Memory attacks may be theoretical, but cryptographic keys, unlike most other data in a computer memory, are random. Looking through memory structures for random data is very likely to reveal key material. Well made libraries for use as Local Encryption Services go to great efforts to protect keys even in memory. Key-encryption keys are used to encrypt the key while it is in memory and then the encrypted key is split into several parts and spread throughout the memory space. Decoy structures might be created that look like valid key material. Memory holding the key is quickly zeroed as soon as the cryptographic operation is finished. These techniques reduce the risk of memory attacks. Separate encryption can also be used for different data. These encryption keys can be automatically rotated based on the sensitivity of the protected data. Dedicated Encryption Services are also vulnerable to memory attacks. However, a well made Dedicated Encryption Service runs only the minimal number of services. Since web servers, application servers, and databases have no place on a dedicated cryptographic engine, these common attack points aren't a threat. This severely constrained attack surface makes it much more difficult to gain the access needed to launch a memory attack. To maintain a high level of security backups contain the encrypted data and only securely encrypted lower level keys. Master keys should be backed up separately.

Secure key back up

A weak link in the security of many networks is the backup process. Often, private keys and certificates are archived unprotected along with configuration data from the backend servers. The backup key file may be stored in clear text or protected only by an administrative password. This password is often chosen poorly and/or shared between operators. To take advantage of this weak protection mechanism, hackers can simply launch a dictionary attack (a series of educated guesses based on dictionary words) to obtain private keys and associated certificates.

A WEAK LINK IN THE SECURITY OF MANY NETWORKS IS THE BACKUP PROCESS

High performance and scalability

This topology combines the enhanced performance of the Software structure with the added security of a hardware device. A HSM, in some situations, is an ideal way to add additional protection for the most important element of any encryption solution ñ the encryption keys. HSM devices are fast and tamper proof, so they make an excellent vault to store the crown jewels ñ the encryption keys. The performance in this topology is essentially identical to the earlier pure software structure, with an occasional transit to the HSM to refresh and retrieve the master encryption keys. During the majority of processing time, performance is identical to the software solution.

In our 500,000-row example, in contrast to the NAED structure – where all 500,000 rows flowed over the wire to the NAED – the encryption service in the database server accesses the key from the HSM one time and thereafter all crypto operations are completed in the database by the software encryption service. The Hybrid system is implemented as distributed processes that scales with the number of processors and database server available. In the Software topology the database server becomes the platform for encryption services, removing the network and a remote device from the equation.

Benchmarks from real world customer environments

We studied the industry standard SQL benchmark as a model for workloads. Some simple sample tests on Oracle and DB2. The first benchmark was focus on a particular customer scenario. Subsequent benchmarks used a workload combined from multiple customer case studies. The technological aspects of developing database privacy as an enterprise IT infrastructure component lead to new research challenges. First and fore-most is the issue of encryption key management. Most corporations view their data as a very valuable asset. The key management system would need to provide sufficient security measures to guard the distributed use of encryption keys. We propose a combined hardware and software based data encryption system as the solution to this problem.

A distributed policy and audit capability is proposed for the control the use of different encryption keys. Detailed investigation of this solution is presented below. Since the interaction between the database and the enterprise IT infrastructure component there are potential over-heads introduced by encryption. Therefore the sources of performance degradation and its significance should be determined.

How to off-load encryption and scale with additional servers

As mentioned before, cryptography consumes a fair amount of cpu cycles; with a Local Encryption Service application performance could be spread over a number of local processors. Dedicated Encryption Services off-load the encryption to a pool of separate processors.

Some Dedicated Encryption Services also use special proprietary processors. Similar performance out of general purpose hardware could be easy and less expensive when using state of the art general servers.

Obtaining a relevant degree of key security with a Local Encryption Service will require key storage separated from the local server. A central server with optional hardware to store the key can provide a cost effective solution in some environments.

A final security consideration with Local Encryption Services is due to fact that the same physical memory is shared between the Local Encryption Service and the application. Many modern application architectures scale by adding additional servers. In the case of directly connected engines, each new server requires a new engine. Directly connected engines in a highly available cluster of servers should provide cryptographic availability. Should one engine fail, then processing could shift to the other servers where the engines were still operational.

Conclusion

We addressed performance as a particularly vital problem and evaluated different solutions for database encryption.

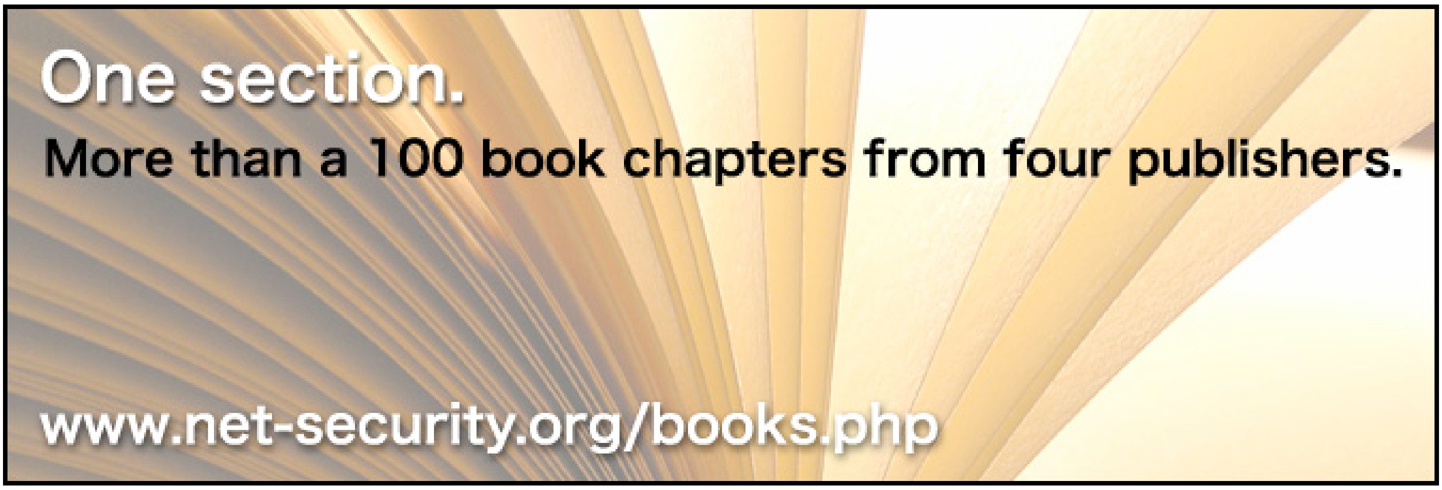
In this article, we discussed the Hybrid, a database privacy solution built on top of all major relational databases.

The Hybrid model introduces many significant challenges primary of which are the additional overhead of searching on encrypted data an infrastructure to guarantee data privacy, and management of such an enterprise IT infrastructure component. We have addressed these issues.

Our experiments using several benchmarks showed that the overhead is tolerable when using suitable encryption architecture. The Hybrid model implements a scalable approach for data privacy and security in which a security administrator protecting privacy at the level of individual fields and records, and providing seamless mechanisms to create, store, and securely access databases. Such a model alleviates the need for organizations to purchase expensive hardware, deal with software modifications, and hire professionals for encryption key management development tasks.

We proposed, implemented, and evaluated different encryption schemes. We showed the drastic decrease in query execution times from distributed software level encryption. We showed that the Hybrid database encryption solution is the most successful offering for most application environments.

Ulf T. Mattsson is the CTO of Protegrity. His extensive IT and security industry experience includes 20 years with IBM as a manager of software development and a consulting resource to IBM's Research and Development organization, in the areas of IT Architecture and IT Security.


One section.
More than a 100 book chapters from four publishers.
www.net-security.org/books.php

We have attended this year's RSA Conference Europe, held near Prater entertainment center in Austrian capital Vienna. If you are not familiar with the concept of this conference, it focuses on providing quality class sessions on all the major aspects of information security. Our coverage is available at:
www.net-security.org/article.php?id=849

Besides these sessions, the most visited presentations included Keynotes by Scott Charney, VP Trustworthy Computing at Microsoft, Art Coviello, RSA Security CEO and ex Director General of MI5 Stella Rimington. The exibition area of this years' event was the biggest one we have seen since our first RSA Conference Europe back in 2002. It was announced that the 2006 conference will be held in Nice, France.