The Cyber Security Mess

Garfinkel, Simson L.
Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/44325

Downloaded from NPS Archive: Calhoun
The Cyber Security Mess

Simson L. Garfinkel
Associate Professor, Naval Postgraduate School
March 26, 2014

DISCLAIMER:
“The views expressed in this document are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.”
NPS is the Navy’s Research University.

Monterey, CA — 1500 students
  • US Military & Civilian (Scholarship for Service & SMART)
  • Foreign Military (30 countries)

Graduate Schools of
Operational & Information Sciences (GSOIS)
  • Computer Science
  • Defense Analysis
  • Information Sciences
  • Operations Research
  • Cyber Academic Group

National Capital Region (NCR) Office
  • 900 N Glebe (Ballston)/Virginia Tech building
The risk of being “hacked”—whatever that expression actually means—is at the heart of our civilization’s chronic cybersecurity problem. Despite decades of computer security research, billions spent on secure operations, and growing training requirements, we seem incapable of operating computers securely.

There are weekly reports of penetrations and data thefts at some of the world’s most sensitive, important, and heavily guarded computer systems. There is good evidence that global interconnectedness combined with the proliferation of hacker tools means that today’s computer systems are actually less secure than equivalent systems a decade ago. Numerous breakthroughs in cryptography, secure coding, and formal methods notwithstanding, cybersecurity is getting worse as we watch.

So why the downward spiral? One reason is that cybersecurity’s goal of reducing successful hacks creates a large target to defend. Attackers have the luxury of choice. They can focus their efforts on the way our computers represent data, the applications that process the data, the operating systems on which those applications run, the networks by which those applications communicate, or any other area that is possibly subverted. And faced with a system that is beyond one’s technical hacking skills, an attacker can go around the security perimeter and use a range of other techniques, including social engineering, supply-chain insertion, or even kidnapping and extortion.

It may be that cybersecurity appears to be getting worse simply because society as a whole is becoming much more dependent upon computers. Even if the vulnerability were not increasing, the successful hacks can have significantly more reach today than a decade ago.
An Introduction to Computer Security

Part 1

Simson L. Garfinkel

"Spies," "vandals," and "crackers" are out there, waiting to get into—or destroy—your databases.

Lawyers must understand issues of computer security, both for the protection of their own interests and the interests of their clients.

Lawyers today must automatically recognize insecure computer systems and lax operating procedures in the same way as lawyers now recognize

Sept. 1987

1991

2000

2006

2006
The lack of security is **inherent** in modern information systems.

- Attack is **easier and cheaper** than defense.
- Cyber “defense in depth” does not work — a single vulnerability compromises.

Today’s systems are less secure than those of the 1970s.

Cyber can directly target inner defenses

It’s easier to break things than to fix them.
How hackers allegedly stole “unlimited” amounts of cash from banks in just hours

Federal authorities have accused eight men of participating in 21st-Century Bank heists that netted a whopping $45 million by hacking into payment systems and eliminating withdrawal limits placed on prepaid debit cards.

The eight men formed the New York-based cell of an international crime ring that organized and executed the hacks and then used fraudulent payment cards in dozens of countries to withdraw the loot from automated teller machines, federal prosecutors alleged in court papers unsealed Thursday. In a matter of hours on two separate occasions, the eight defendants and their confederates withdrew about $2.8 million from New York City ATMs alone. At the same time, “cashing crews” in cities in at least 26 countries withdrew more than $40 million in a similar fashion.
Hacked AP Twitter feed reporting fake White House attack rocks markets

Account compromise comes after AP targeted by malware and phishing e-mails.

by Dan Goodin - Apr 23 2013, 3:44pm EDT

The seven-minute drop in the Dow Jones Industrial Average touched off by a single tweet falsely claiming the White House had been bombed. It temporarily wiped out about 1 percent of the average, which can translate into millions or billions of dollars in market capitalization.

Stock prices plunged and then quickly recovered after a Twitter account belonging to the Associated Press was hacked and used to send a bogus report falsely claiming that the White House had been bombed and President Obama was injured.

March 2014:
IRS Employee Took Home Data on 20,000 Workers

A U.S. Internal Revenue Service employee took home a computer thumb drive containing unencrypted data on 20,000 fellow workers, the agency said in a statement today.

The tax agency's systems that hold personal data on hundreds of millions of Americans weren't breached, the statement said.

"This incident is a powerful reminder to all of us that we must do everything we can to protect sensitive data — whether it involves our fellow employees or taxpayers," IRS Commissioner John Koskinen said in a message to employees. "This was not a problem with our network or systems, but rather an isolated incident."

The IRS is contacting the current and former employees involved, almost all of whom worked in Pennsylvania, Delaware and New Jersey. The information dates to 2007, before the IRS started using automatic encryption.

IRS officials were told of the breach "a few days ago," Koskinen's message said.

The Social Security numbers, names and addresses of employees and contract workers were potentially accessible online because the thumb drive was plugged into the employee's "unsecure home network," Koskinen's message said.

The IRS said it had no knowledge of the information being used to commit identity theft.

March 2014: Stolen F-35 secrets showing up in China’s stealth Fighter
March 2014: Target ignored alarms before hack.

The biggest retail hack in U.S. history wasn’t particularly inventive, nor did it appear destined for success. In the days prior to Thanksgiving 2013, someone installed malware in Target’s (TGT) security and payments system designed to steal every credit card used at the company’s 1,797 U.S. stores. At the critical moment—when
The cyber security mess: it’s technical *and* social.

Most attention is focused on technical issues:

- Malware and anti-viruses
- Access controls, authentication & cryptography
- Supply chain issues
- Cyberspace as a globally connected “domain”

Non-technical issues are at the heart of the cyber security mess.

- Education & career paths
- Immigration
- Manufacturing policy

We will do better when we *want* to do better.
What do we know about cyber security today?
Cyber Security... is undefined.

There is no good definition for “cyber”
- Something having to do with cybernetics
- Computers?
- Computer networks?
- Hacking?
- Using “network security” to secure desktops & servers?

There is no way to measure the security of a “cyber” system
- Which OS is more secure?
- Which computer is more secure?
- Is “open source” more secure?

—A system that seems “more secure”
can suffer a total compromise from a single unknown attack.
We can measure expenditures. Cyber Security is expensive.

Global cyber security spending: $60 billion in 2011
- *Cyber Security M&A, pwc, 2011*

172 Fortune 500 companies surveyed:
- Spending $5.3 billion per year on cyber security.
- Stopping 69% of attacks.

If they raise spending...
- $10.2 billion stops 84%
- $46.67 billion stops 95%
- “highest attainable level”

95% is not good enough.
Spending more money does not make a computer more secure.
Paradox: Cyber security research makes computers less secure!

- Data
- Encoding
- Apps
- OS (programs & patches)
- Network & VPNs
- DNS, DNSSEC
- IPv4 / IPv6
- Embedded Systems
- Human operators
- Hiring process
- Supply chain
- Family members

The more we learn about securing computers, the better we get at attacking them.
Cyber Security is an “insider problem.”

bad actors
good people with bad instructions
remote access
malware

If we can stop insiders, we might be able to secure cyberspace….

—... but we can’t stop insiders.

Ames  Hanssen  Manning  Snowden
Cyber Security is a “network security” problem.

We can’t secure the hosts, so secure the network!

- Isolated networks for critical functions.
- Stand-alone hosts for most important functions.

But strong crypto limits visibility into network traffic, and...
... stuxnet shows that there are no isolated hosts.

Iranian President Mahmoud Ahmadinejad inspects nuclear centrifuges
“to a first approximation, every computer in the world is connected to every other computer.”

—Robert Morris (1932-2001), to the National Research Council’s Computer Science and Technology Board, Sept. 19, 1988

“Action is needed on many fronts to protect computer systems and communications from unauthorized use and manipulation.”

The wonders of the Internet and the promise of the worldwide information infrastructure have recently reached headline status. Connectedness has become the Holy Grail of the 1990s. But expansion of the electronic network brings with it increased potential for harm as well as good. With a broader cross-section of people logging on to the electronic superhighway and with the enhanced interconnectedness of all computer systems, the likelihood of mischievous or even criminal behavior grows, as does the potential extent of the damage that can be done.

But in spite of the higher risks and higher stakes, little attention has been paid to the need for enhanced security. The stories that appear in the press from time to time about prankster hackers breaking into a computer network or computer viruses infecting government computers focus more on the skill of the culprit than the harm done. The popular assumption is that break-ins are relatively harmless. Most computer users complacently believe that if there was real cause for alarm, government or corporate computer experts would recognize the problem and take appropriate action.

Unfortunately, experts and neophytes alike have their heads in the sand on this issue. In spite of repeated examples of the vulnerability of almost all computer systems to invasion and manipulation, very few people recognize the magnitude of the damage that can be done and even fewer have taken adequate steps to fix the problem.

Peter G. Neumann is a principal scientist in the Computer Science Laboratory at SRI International in Menlo Park, California. His new book, Computer-Related Risks (ACM Press/Addison-Wesley, 1994), discusses reliability and safety problems as well as security.
Cyber Security is a “process” problem.

Security encompasses all aspects of an organization’s IT and HR operations.

Microsoft Security Development Lifecycle

- Few organizations can afford SDL.
- Windows 7, Windows 8 is still hackable...
January 2013: Windows RT jailbroken

Microsoft controlled the hardware and the software.

Windows RT — still hacked
Cyber Security is a money problem.

Security is a cost.....Not an “enabler”
- No ROI

Chief Security Officers are in a no-win situation:
- Security = passwords = frustration
- No reward for spending money to secure the infrastructure
- Money spent on security is “wasted” if there is no attack

—“If you have responsibility for security but have no authority to set rules or punish violators, your own role in the organization is to take the blame when something big goes wrong.”
- Spaf’s first principle of security administration
  *Practical Unix Security*, 1991
NEW YORK —

(AP) — Target Corp. (NYSE:TGT) Chief Information Officer Beth Jacob is resigning effective Wednesday as the retailer overhauls its information security and compliance division in the wake of a massive pre-Christmas data breach.

Target Chairman, President and CEO Gregg Steinhafel said in a statement released to The Associated Press that the company will search for an interim chief information officer who can help guide the company through the transformation.

Jacob had been in her current role since 2008 and oversaw teams in the U.S. and India.

Target disclosed on Dec. 19 that the data breach compromised 40 million credit and debit card accounts between Nov. 27 and Dec. 15. Then on Jan. 10 it said hackers also stole personal information — including names, phone numbers as well as email and mailing addresses — from as many as 70 million customers.
Cyber Security is a “wicked problem”

No clear definition of the wicked problem
   — *You don’t understand the problem until you have a solution.*

No “stopping rule”
   — *The problem can never be solved.*

Solutions not right or wrong
   — *Benefits to one player hurt another — Information security vs. Free speech*

Solutions are “one-shot” — no learning by trial and error
   — *No two systems are the same. The game keeps changing.*

Every wicked problem is a symptom of another problem

- Dave Clement, “Cyber Security as a Wicked Problem,” Chatham House, October 2011
Is it the technology?

Why is cyber security so hard?
Cyber Security has an active, malicious adversary.

The adversary...

— *Turns your bugs into exploits*
— *Adapts to your defenses*
— *Waits until you make a mistake*
— *Attacks your employees when your systems are secure*
For example...

Compiler bugs are security vulnerabilities!

Compilers are core technology used in software development.

We have seen:

• Optimizations to make programs run faster can become security vulnerabilities
• The same errors are repeatedly made by different programmers

What’s difference between a bug and an attack?

—The programmer’s intent.
Bugs in CPU silicon are remotely exploitable!

This means:

- Programs that are “secure” on one CPU may be vulnerable on another.
- Auditing the code & the compiler isn’t enough.

Kaspersky:

- “Fact: malware that uses CPU bugs really does exist;”
- “not apocalypse, just a new threat,”
The supply chain creates numerous security vulnerabilities

- App Developers
- 3rd Party Kits
- Open Source
- Apple Developers
- iOS
- CPU
- Carrier
- Wireless
- Apps
The attacker is smarter than you are... and has more time to find a good attack.

Smartphone designers were sure that there was no privacy leakage in accelerometers. We now know they can:

- Reveal your position
- Reveal your PIN

ACCoplice: Location Inference using Accelerometers on Smartphones

Jun Han, Emmanuel Owsusu, Le T. Nguyen, Adrian Perrig, Joy Zhang
[junhan, owusu, le.nguyen, adrian, joy]@cmu.edu
Carnegie Mellon University

Abstract

The security and privacy risks posed by smartphone accelerometers have largely been ignored. We show that accelerometers can be used to locate a device owner to within a 200 meter radius of the true location. Our results are comparable to those of prior work using GPS, even when access to GPS is disabled. However, the importance of accelerometers have been largely ignored. We demonstrate that accelerometers can be used to infer highly sensitive information such as personal location and position without requiring special privileges. We demonstrate that accelerometer readings can be used to fingerprint personal accelerometers, determine their location and infer the identity of the owner. This enables adversaries to infer the location of individuals or organizations that use mobile phone accelerometers. We define a framework for studying the privacy risks of accelerometers and show how existing techniques can be used to infer the location of nearby devices.

I. INTRODUCTION

Location privacy has been a hot topic in recent years after reports that Apple, Google, and Microsoft collect records of the location of millions of users of their mobile operating systems. These records are used for certain tasks against the operator [1]. For example, in the case of mobile phone users, it is often difficult to assess the degree of personal information that can be collected about a user's location. The goal of this paper is to study the security and privacy risks posed by smartphone accelerometers. Accelerometers are particularly interesting because they are available on virtually all smartphones and tablets, and can be used to track a user's location long after location services have been disabled [6]. But as we show, the accelerometer can be used to infer a location with an initial location information. There is a very personal side-channel that can be exploited even when location information is not shared.

A) Problem Definition: We assume that the adversary can monitor operations on the mobile device, without any special privileges except the capability to send information over the network. The application will tell our framework where to obtain access to network communications. This is easily achieved by monitoring a popular application that many users access on a daily basis.

1) We define a framework to study the security and privacy risks posed by smartphone accelerometers. Accelerometers can be used to fingerprint devices and infer the location of nearby devices.

2) We show that accelerometers can be used to infer highly sensitive information such as personal location and position without requiring special privileges. We demonstrate that accelerometer readings can be used to fingerprint personal accelerometers, determine their location and infer the identity of the owner. This enables adversaries to infer the location of individuals or organizations that use mobile phone accelerometers.

3) We define a framework for studying the privacy risks of accelerometers and show how existing techniques can be used to infer the location of nearby devices.

II. SYSTEM OVERVIEW

A) Contributions:

1) We develop a new framework to study the security and privacy risks posed by smartphone accelerometers. Accelerometers can be used to fingerprint devices and infer the location of nearby devices.

2) We show that accelerometers can be used to infer highly sensitive information such as personal location and position without requiring special privileges. We demonstrate that accelerometer readings can be used to fingerprint personal accelerometers, determine their location and infer the identity of the owner. This enables adversaries to infer the location of individuals or organizations that use mobile phone accelerometers.

3) We define a framework for studying the privacy risks of accelerometers and show how existing techniques can be used to infer the location of nearby devices.

III. METHODOLOGY

A) Dataset:

1) We collect data from a variety of mobile devices, including smartphones and tablets, and track their accelerometer readings over time.

2) We then analyze the data to determine the location of the nearby devices.

IV. RESULTS

A) Discussion:

1) We show that accelerometers can be used to fingerprint devices and infer the location of nearby devices.

2) We demonstrate that accelerometer readings can be used to fingerprint personal accelerometers, determine their location and infer the identity of the owner. This enables adversaries to infer the location of individuals or organizations that use mobile phone accelerometers.

3) We define a framework for studying the privacy risks of accelerometers and show how existing techniques can be used to infer the location of nearby devices.

V. CONCLUSION

A) Conclusion:

1) We develop a new framework to study the security and privacy risks posed by smartphone accelerometers.

2) We show that accelerometers can be used to fingerprint devices and infer the location of nearby devices.

3) We demonstrate that accelerometer readings can be used to fingerprint personal accelerometers, determine their location and infer the identity of the owner. This enables adversaries to infer the location of individuals or organizations that use mobile phone accelerometers.

4) We define a framework for studying the privacy risks of accelerometers and show how existing techniques can be used to infer the location of nearby devices.

6 accelerometers
no privacy

https://sparrow.ece.cmu.edu/group/pub/han_ACCoplice_comsnets12.pdf

Jun Han, Emmanuel Owsusu, Thanh-Le Nguyen, Adrian Perrig, and Joy Zhang
Many people liken cyber security to the flu.

DHS calls for “cyber hygiene”

- install anti-virus
- update your OS
- back up key files

—“STOP, THINK, CONNECT”
Another model might be *obesity*....

Making people fat is good business:

- Farm subsidies
- Restaurants
- Healthcare and medical utilization
- Weight loss plans
  
  —*Few make money when Americans stay trim and healthy.*

Lax security is also good business:

- Cheaper cost of deploying software
- Private information for marketing
- Selling anti-virus & security products
- Cleaning up incidents
  
  —*Few benefit from secure computers*
Some people say that cyber war is like nuclear war.

http://www.acus.org/new_atlanticist/mind-cyber-gap-deterrence-cyberspace

http://www.beyondnuclear.org/security/
Biowar may be a better model for cyberwar.

—Cheap to produce
—Easy to attack
—Hard to control
—Hard to defend
—No clear end
Security problems are bad for society as a whole... 

... because [wireless] computers are everywhere.

50 microprocessors per average car

http://www.autosec.org/

– Comprehensive Experimental Analysis of Automotive Attack Surfaces (2011)
– Experimental Security Analysis of a Modern Automobile (2010)

Remote take-over of EVERY safety-critical system from ANY wired or wireless interface

2008: demonstrated wireless attack on implantable pacemakers

2012: demonstrated wireless attack on insulin pump

DDoS the endocrine system!
It's still premature to say you need firewall or antivirus protection for your television set, but a duo of recently diagnosed firmware vulnerabilities in widely used TV models made by two leading manufacturers suggests the notion isn't as far-fetched as many may think.

... While poking around a Samsung D6000 model belonging to his brother, he inadvertently discovered a way to remotely send the TV into an endless restart mode that persists even after unplugging the device and turning it back on.

"It wasn't even planned," Auriemma told Ars, referring to the most damaging of his two attacks, which rendered the device useless for three days...
Lax security makes non-banking sites prime targets for skimming attacks...
Cell phones cannot be secured.

Cell phones have:
• Wireless networks, microphone, camera, & batteries
• Downloaded apps
• Bad crypto

Cell phones can be used for:
• Tracking individuals
• Wiretapping rooms
• Personal data

How do we address the cybersecurity challenge?

1. Deploy technology that works.

2. Address the non-technical issues.
We have made major advances in cyber security.

Major security breakthroughs since 1980:

- Public key cryptography (RSA with certificates to distribute public keys)
- Fast symmetric cryptography (AES)
- Fast public key cryptography (elliptic curves)
- Easy-to-use cryptography (SSL/TLS)
- Sandboxing (Java, C# and virtualization)
- Firewalls
- BAN logic
- Fuzzing.

None of these breakthroughs has been a “silver bullet,” but they have all helped.

We must continue to deploy technology that works, because adversaries are not all powerful.

Adversaries are impacted by:

— *Economic factors*
— *Attention span*
— *Other opportunities*

You don’t have to run faster than the bear….
There are solutions to many cyber security problems... We should use them!

30% of the computers on the Internet run Windows XP
- Windows 7 & 8 have vulnerabilities, but they’re better.

Apple users don’t use anti-virus.
- Yes, Apple tries to fix bugs, but

Most “SSL” websites only use it for logging in.

DNSSEC

Smart Cards
Example: Google Authenticator’s 2-factor authentication protections against password stealing.
Example: Apply “Recovery Oriented Computing” to security.

Recovery Oriented Computing [Stanford & Berkeley]:

- Systems that are isolated and redundant
- System-wide undo support
- Integrated Diagnostic support
- Online verification and recovery mechanisms
- Modularity, measurability and restorability

Applying to security:

- Pervasive use of digital signatures.
- Disconnected, offline storage.
We must address the non-technical factors that impact cyber security.

These factors reflect deep divisions within our society.

- **Shortened** development cycles

- **Education**: Not enough CS graduates; not enough security in CS.

- **Labor**:
  - *Immigration Policy*: Foreign students; H1B Visa
  - *HR*: Inability to attract and retain the best workers

- **Manufacturing Policy**: Where we are building our computers.

Solving the cyber security mess requires addressing these issues.
Short development cycles

Insufficient planning:
• Security not “baked in” to most products.
• Few or no security reviews
• Little Usable Security

Insufficient testing:
• Testing does not uncover security flaws
• No time to retest after fixing

Poor deployment:
• Little monitoring for security problems
• Difficult to fix current system when new system is under development
Education is not supplying enough security engineers. Software engineers don’t learn enough about security.

Security HR Pipeline

• High School → College → Graduate School → Career

It takes years to master security...

• Many professional programmers learn their craft in college.
• College English graduates: 16 years’ instruction in writing
• College CS graduates: 4 years’ instruction in programming

—Is it any wonder their code has security vulnerabilities?

• 64% of the 2,500 vulnerabilities in the National Vulnerability Database in 2004 were caused by programming errors.

Snapchats Don't Disappear: Forensics Firm Has Pulled Dozens of Supposedly-Deleted Photos From Android Phones

http://www.forbes.com/sites/kashmirhill/2013/05/09/snapchats-dont-disappear/
73% of states require computer “skills” for graduation. Only 37% require CS “concepts”

CS teachers are paid far less than CS engineers.
—Salaries for beginning & average teachers lag CS engineers by 30%
—Adjusting for cost-of-living and shorter work week.

- Linda Darling-Hammond, Stanford University, 2004
  http://www.srnleads.org/data/pdfs/ldh_achievement_gap_summit/inequality_TCR.pdf
High school students are not taking AP computer science!
Computer Science undergraduate enrollment has gone up for 4 years, but still not reached 2001-2004 levels

CRA is psyched at double-digit gains in recent years, but…

Source: Table 3: Bachelor’s Degrees Awarded by Department Type

2013 CRA Taulbee Survey

http://cra.org/govaffairs/blog/2013/03/taulbeereport/
7% of Bachelor’s degrees awarded to “nonresident alien” (12,596 to US citizens)

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>CS</th>
<th>CE</th>
<th>I</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonresident Alien</td>
<td>619</td>
<td>216</td>
<td>98</td>
<td>933</td>
</tr>
<tr>
<td>Amer Indian or Alaska Native</td>
<td>39</td>
<td>6</td>
<td>12</td>
<td>57</td>
</tr>
<tr>
<td>Asian</td>
<td>1,477</td>
<td>447</td>
<td>341</td>
<td>2,265</td>
</tr>
<tr>
<td>Black or African-American</td>
<td>407</td>
<td>107</td>
<td>203</td>
<td>717</td>
</tr>
<tr>
<td>Native Hawaiian/Pac Islander</td>
<td>18</td>
<td>4</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>White</td>
<td>5,793</td>
<td>1,154</td>
<td>1,522</td>
<td>8,469</td>
</tr>
<tr>
<td>Multiracial, not Hispanic</td>
<td>130</td>
<td>27</td>
<td>26</td>
<td>183</td>
</tr>
<tr>
<td>Hispanic, any race</td>
<td>575</td>
<td>102</td>
<td>203</td>
<td>880</td>
</tr>
<tr>
<td>Total Residency &amp; Ethnicity Known</td>
<td>9,058</td>
<td>2,063</td>
<td>2,408</td>
<td>13,529</td>
</tr>
<tr>
<td>Resident, ethnicity unknown</td>
<td>732</td>
<td>117</td>
<td>89</td>
<td>938</td>
</tr>
<tr>
<td>Residency unknown</td>
<td>1,259</td>
<td>176</td>
<td>73</td>
<td>1,508</td>
</tr>
<tr>
<td>Grand Total</td>
<td>11,049</td>
<td>2,356</td>
<td>2,570</td>
<td>15,975</td>
</tr>
</tbody>
</table>

—Most do not go on to advanced degrees.
54% of Master’s degrees awarded to nonresident alien (4960 to US citizens)

<table>
<thead>
<tr>
<th>Table M3. Master’s Degrees Awarded by Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Nonresident Alien</strong></td>
</tr>
<tr>
<td>CS 4,123 62.3%</td>
</tr>
<tr>
<td>CE 544 69.3%</td>
</tr>
<tr>
<td>I 397 19.8%</td>
</tr>
<tr>
<td>Total 5,064 53.8%</td>
</tr>
<tr>
<td><strong>Amer Indian or Alaska Native</strong></td>
</tr>
<tr>
<td>CS 10 0.2%</td>
</tr>
<tr>
<td>CE 1 0.1%</td>
</tr>
<tr>
<td>I 9 0.4%</td>
</tr>
<tr>
<td>Total 20 0.2%</td>
</tr>
<tr>
<td><strong>Asian</strong></td>
</tr>
<tr>
<td>CS 484 7.3%</td>
</tr>
<tr>
<td>CE 52 6.6%</td>
</tr>
<tr>
<td>I 213 10.6%</td>
</tr>
<tr>
<td>Total 749 8.0%</td>
</tr>
<tr>
<td><strong>Black or African-American</strong></td>
</tr>
<tr>
<td>CS 123 1.9%</td>
</tr>
<tr>
<td>CE 8 1.0%</td>
</tr>
<tr>
<td>I 122 6.1%</td>
</tr>
<tr>
<td>Total 253 2.7%</td>
</tr>
<tr>
<td><strong>Native Hawaiian/Pac Island</strong></td>
</tr>
<tr>
<td>CS 9 0.1%</td>
</tr>
<tr>
<td>CE 0 0.0%</td>
</tr>
<tr>
<td>I 0 0.0%</td>
</tr>
<tr>
<td>Total 9 0.1%</td>
</tr>
<tr>
<td><strong>White</strong></td>
</tr>
<tr>
<td>CS 1,725 26.1%</td>
</tr>
<tr>
<td>CE 161 20.5%</td>
</tr>
<tr>
<td>I 1,144 57.0%</td>
</tr>
<tr>
<td>Total 3,030 32.2%</td>
</tr>
<tr>
<td><strong>Multiracial, not Hispanic</strong></td>
</tr>
<tr>
<td>CS 22 0.3%</td>
</tr>
<tr>
<td>CE 1 0.1%</td>
</tr>
<tr>
<td>I 25 1.2%</td>
</tr>
<tr>
<td>Total 48 0.5%</td>
</tr>
<tr>
<td><strong>Hispanic, any race</strong></td>
</tr>
<tr>
<td>CS 123 1.9%</td>
</tr>
<tr>
<td>CE 18 2.3%</td>
</tr>
<tr>
<td>I 96 4.8%</td>
</tr>
<tr>
<td>Total 237 2.5%</td>
</tr>
<tr>
<td><strong>Total Residency &amp; Ethnicity Known</strong></td>
</tr>
<tr>
<td>CS 6,619</td>
</tr>
<tr>
<td>CE 785</td>
</tr>
<tr>
<td>I 2,006</td>
</tr>
<tr>
<td>Total 9,410</td>
</tr>
<tr>
<td><strong>Resident, ethnicity unknown</strong></td>
</tr>
<tr>
<td>CS 285</td>
</tr>
<tr>
<td>CE 78</td>
</tr>
<tr>
<td>I 144</td>
</tr>
<tr>
<td>Total 507</td>
</tr>
<tr>
<td><strong>Residency unknown</strong></td>
</tr>
<tr>
<td>CS 558</td>
</tr>
<tr>
<td>CE 15</td>
</tr>
<tr>
<td>I 28</td>
</tr>
<tr>
<td>Total 601</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
</tr>
<tr>
<td>CS 7,462</td>
</tr>
<tr>
<td>CE 878</td>
</tr>
<tr>
<td>I 2,178</td>
</tr>
<tr>
<td>Total 10,518</td>
</tr>
</tbody>
</table>

—*We should let them stay in the country after they graduate*
50% of PhDs awarded in 2012 to nonresident aliens

We did not train Russia’s weapons scientists at MIT during the Cold War.

Table D3. PhDs Awarded by Ethnicity

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>CE</th>
<th>I</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonresident Alien</td>
<td>763 51.3%</td>
<td>99 55.3%</td>
<td>32 26.9%</td>
<td>894 50.1%</td>
</tr>
<tr>
<td>Amer Indian or Alaska Native</td>
<td>1 0.1%</td>
<td>0 0.0%</td>
<td>1 0.8%</td>
<td>2 0.1%</td>
</tr>
<tr>
<td>Asian</td>
<td>168 11.3%</td>
<td>32 17.9%</td>
<td>27 22.7%</td>
<td>227 12.7%</td>
</tr>
<tr>
<td>Black or African-American</td>
<td>27 1.8%</td>
<td>1 0.6%</td>
<td>7 5.9%</td>
<td>35 2.0%</td>
</tr>
<tr>
<td>Native Hawaiian/Pac Islander</td>
<td>5 0.3%</td>
<td>0 0.0%</td>
<td>0 0.0%</td>
<td>5 0.3%</td>
</tr>
<tr>
<td>White</td>
<td>496 33.4%</td>
<td>45 25.1%</td>
<td>51 42.9%</td>
<td>592 33.2%</td>
</tr>
<tr>
<td>Multiracial, not Hispanic</td>
<td>5 0.3%</td>
<td>0 0.0%</td>
<td>0 0.0%</td>
<td>5 0.3%</td>
</tr>
<tr>
<td>Hispanic, any race</td>
<td>22 1.5%</td>
<td>2 1.1%</td>
<td>1 0.8%</td>
<td>25 1.4%</td>
</tr>
<tr>
<td>Total Residency &amp; Ethnicity Known</td>
<td>1,487</td>
<td>179</td>
<td>119</td>
<td>1,785</td>
</tr>
<tr>
<td>Resident, ethnicity unknown</td>
<td>25</td>
<td>1</td>
<td>5</td>
<td>31</td>
</tr>
<tr>
<td>Residency unknown</td>
<td>94</td>
<td>14</td>
<td>5</td>
<td>113</td>
</tr>
<tr>
<td>Grand Total</td>
<td>1,606</td>
<td>194</td>
<td>129</td>
<td>1,929</td>
</tr>
</tbody>
</table>

—We did not train Russia’s weapons scientists at MIT during the Cold War.
Just 67 / 1275 (5%) PhDs went into Information Assurance
15 professors & postdocs; 48 to industry & government

Security should be taught to everyone, but we need specialists
“...the problem may not be that there are too few STEM qualified college graduates, but rather that STEM firms are unable to attract them.

Highly qualified students may be choosing a non-STEM job because it pays better, offers a more stable professional career, and/or perceived as less exposed to competition from low-wage economies.”
Bureau of Labor Statistics puts CS as 12th highest paying profession, after...

Highest paying occupations:

- Oral Surgeons > $166,400
- Orthodontists > $166,400
- Physicians and Surgeons > $166,400
- CEOs: $165,080
- Dentists: $161,020
- Judges: $119,260
- Architectural & Eng. Mgrs. $119,260
- Prosthodontists $118,400
- Podiatrists $118,030
- Natural Sci. Mgrs. $116,020
- Computer Scientists: $115,070
- Petroleum Engineers $114,080
- Marketing Managers $112,800
- Lawyers: $112,760
Manufacturing policy —
The US did not buy WW2 aircraft from Germany

Boeing Whichata B-29 Assembly Line, 1944
But we buy nearly all of our computers from China.

It’s easy to put backdoors in hardware and software.
There is no obvious way to secure cyberspace.

We *trust* computers…
—*but we cannot make them trustworthy.*
(A “trusted” system is a computer that can violate your security policy.)

We know a lot about building secure computers…
—*but we do not use this information when building and deploying them.*

We know about usable security…
—*but we can’t make any progress on usernames and passwords*

We should design with the assumption that computers will fail…
—*but it is cheaper to design without redundancy or resiliency.*

Despite the new found attention to cyber security, our systems seem to be growing more vulnerable every year.
Backup Slides: HCI-SEC
Major Themes in HCI-SEC Academic Research

User Authentication
- Text Passwords
- Graphical Authentication
- Biometrics
- Token-based Authentication
- CAPTCHAs

Email Security and PKI
- Automatic, Transparent Encryption

Anti-Phishing Technology
Password Managers
Device Pairing
Web Privacy
Policy Specification and Interaction
Security Experts
Mobile Security and Privacy
- Location Privacy
- Application platforms
- Mobile authentication

Social Media Privacy
HCI-SEC Lessons and Challenges

Lessons Learned:
• Users need better information, not more information
• To make good decisions, users require clear context
• Plain Language Works, Even if it is less precise
• Where Possible, Reduce Decisions and Configuration Options
• Education Works, but cannot overcome economics

Research Challenges
• Authentication Challenges
• Administration Challenges
• Privacy Challenges
• Challenge of Modelling the Adversary
• The Challenge of Social Media and Social Computing
• Teaching Challenges
HCI-SEC Conclusion: The Next 10 years

More HCI-SEC Research Centers
More HCI-SEC Research Targets
Increased Researching on Nudges and Persuasion
Increased Emphasis on Offensive Work
Increased demand for HCI-SEC from non-technical sectors
Backup Slides: Insider Threat
DETECTING THREATENING INSIDERS WITH LIGHTWEIGHT MEDIA FORENSICS

Naval Postgraduate School &
The University of Texas at San Antonio

Dr. Simson Garfinkel (NPS) & Dr. Nicole Beebe (UTSA)

8am, Wednesday November 13th, 2013
Team Profile

Naval Postgraduate School

• Simson L. Garfinkel
  Assoc. Prof
  Computer Science
  —simsong@acm.org
  —+1.202.649.0029

The University of Texas at San Antonio

• N. Beebe, Asst. Prof.
  Info Systems/Cyber Security
  —Nicole.Beebe@utsa.edu
  —+1.210.269.5647
The current approaches for finding hostile insiders are based on “signatures.”

Sample signature to find a problem employee:

(CERT 2011)

- if the mail is from a departing insider
- and the message was sent in last 30 days
- and the recipient is not in organization’s domain
- and the total bytes summed by day is more than X,

→ send an alert to security operator

These signatures are typically hand written.

— Brittle
— Don’t scale
— Miss new patterns
We propose a new approach for finding threatening insiders—storage profile anomalies.

Hypothesis 1:
Some insiders hoard before exfiltration

- Manning
- Snowden
Hypothesis 2:
Some illegal activity has storage indicators:
• Contraband software (hacking tools) and data
• Large amount of:
  — graphics
  — PII; PHI; account numbers
  — Encrypted data
• Stolen documents

Illegal employee activity is:
• Bad for business
• Exploitation threat
• Fraud risk
Our plan: look for storage devices that are different than their peers.

We build a “storage profile” from features:

- # of credit card numbers, phone #s; SSNs, DOBs, etc.
- % pictures; %video
- % Doc files; %PDFs;

“Different” relative to:

- User’s history
- User’s organization
- Others in role.

Our approach: Collect “storage profiles” and look for outliers.

We profile storage on the hard drive/storage device:
- Allocated & “deleted” files; Unallocated space (file fragments)

Statistical profile is collected:
- Frequently, at “random” times
- Securely — by going to raw media
- Centrally — at management console
We cluster the storage profiles to find “outliers.”

What’s an outlier?

- Something that’s different from its peers
- Something different from its own history
Outlier detection should have significant benefits:

- Not signature based
- Not reliant on access patterns
- Not reliant on policy definition, discovery, auditing

Design constraints:
- Agent must be scalable and cannot interfere with operations
  - Desktop: background process, samples disk data
  - Network load: small, aggregated data transfer
  - Management console: scalable algorithms used

- Must work with isolated systems
- Must be OS agnostic
- Must includes deleted data in collection/analysis
Our system has three parts:

1. Sample disk to collect desired data
   - bulk_extractor
     — *a lightweight media forensics tool*

2. Client-server, enterprise response framework
   - Google Rapid Response (GRR)

3. Anomaly detection agent
   - Univariate and multivariate outlier detection
Random sampling is a great way to analyze data.

Simple random sampling can determine % free space

Garfinkel, Simson, Vassil Roussev, Alex Nelson and Douglas White, *Using purpose-built functions and block hashes to enable small block and sub-file forensics*, DFRWS 2010, Portland, OR

Data characterization can determine the *kind* of stored data

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Audio</th>
<th>Photos</th>
<th>Other</th>
<th>Free</th>
</tr>
</thead>
<tbody>
<tr>
<td>148.87 GB</td>
<td>2.25 GB</td>
<td>8.18 GB</td>
<td>20.25 GB</td>
<td>118.20 GB</td>
</tr>
</tbody>
</table>

Audio Data reported by iTunes: 2.25 GB, 2.42 GB
MP3 files reported by file system: 2.39 GB
Estimated MP3 usage with random sampling: 2.49 GB, 2.71 GB, 10,000 random samples, 5,000 random samples

Sector hashing can identify specific target files

It takes 3.5 hours to read a 1TB hard drive.

In 5 minutes you can read:

- 36 GB in one strip
- 100,000 randomly chosen 64KiB strips (assuming 3 msec/seek)

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Data</th>
<th># Seeks</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>208</td>
<td>1 TB</td>
<td>1</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>36 GB</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>5</td>
<td>6.5 GB</td>
<td>100,000</td>
<td>0.65%</td>
</tr>
</tbody>
</table>

Tuesday, March 25, 14
The statistics of a randomly chosen sample predict the statistics of a population.

US elections can be predicted by sampling thousands of households:

Hard drive contents can be predicted by sampling thousands of sectors:

The challenge is identifying likely voters.

The challenge is identifying the sector content that is sampled.
We think of computers as devices with *files*.
This heatmap of anomalies let an analyst easily identify clusters and outliers.
Current status —

bulk_extractor updated v1.4 just released
  • Added features & GRR integration preparation

Sceadan data type classifier updated v1.2 released

Extraction, transformation, loading of datasets
  • M57 Patents (digitalcorpora.org) case

Progress on anomaly detection algorithm
  • Real Data Corpus extraction, translation and loading near complete
  • Theoretical development
  • Empirical data descriptive analyses (test assumptions)
  • Univariate anomaly detection performing well on synthetic data set
We are in year 1 of a 3-year effort.

<table>
<thead>
<tr>
<th>Year</th>
<th>NPS Lead</th>
<th>UTSA Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>bulk_extractor upgrades</td>
<td>Outlier detection algorithm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Synthetic data experimentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Real Data Corpus experimentation</td>
</tr>
<tr>
<td>Year 2</td>
<td>Integrate GRR</td>
<td>Develop/test data outlier detection</td>
</tr>
<tr>
<td></td>
<td>Develop/test management console</td>
<td>Develop/test visualization component</td>
</tr>
<tr>
<td>Year 3</td>
<td>Large-scale testing on partner net</td>
<td>Final dev. of outlier detection algorithm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final dev. of visualization agent</td>
</tr>
</tbody>
</table>
Many challenges remain.

“Anomalous” suggests “normal” exists

- Large, diverse, dislocated organizations
- High fluidity and variety in workforce
- Remote, mobile, multi-device access requirements
- Uninterruptible, critical computational operations

Clustering algorithm selection/development

- Accuracy and speed trade-off of extant algorithms
- Develop combinatorial algorithm to improve accuracy
- Need for automated parameter selection amidst noise
- Feature selection

Engineering of visualization component
In conclusion, we are developing a system that uses “lightweight media forensics” to find hostile insiders.

We use random sampling to build a storage profile of media.

We collect these profiles on a central server.

We cluster & data mine to find outliers.

Contact:
- Simson L. Garfinkel simsong@acm.org
- Nicole Beebe Nicole.Beebe@utsa.edu