
Freedom of Speech: Accountability



Sometimes People Misbehave

- Hacking
- Libel
- Copyright infringement
- Threats
- Child pornography
- Other illegal behavior

There's a Balance

- Last lecture, I said we needed anonymity
- Today, I'm saying there are reasons it can't be absolute
- Which is it?

Checks and Balances

- Few rights are absolute
- Who can track someone?
- Under what conditions can they track someone?
- Is it possible to bypass the restrictions?

Criminal Offenses

- Full power of wiretap law
- But—wiretaps are limited to certain serious offenses
- Also use pen registers, trap-and-trace, informants, bugs, etc.
- Must convince police or prosecutor that the offense is real and of sufficient magnitude to warrant prosecution: “de minimis non curat lex” (“the law does not care about trifles”)

Civil Offenses

- Can still get subpoenas, even against third parties
- But—you need a real case to get a subpoena
- De minimis non curat lex—and you generally can't get a subpoena until there's a real lawsuit

SLAPP

- SLAPP—*Strategic Lawsuit Against Public Participation*
- Sometimes filed by large organizations to harass opponents
- Force the opponents to spend a lot of money defending themselves, even if the lawsuit is preposterous
- Also—break their anonymity/pseudonymity

There Are Real Problems

- Can online commentary or harrassment be actionable?
 - Certainly—though usually it isn't
 - The standards for libel online are the same as offline
 - Anonymity (or perceived anonymity) seems to breed irresponsible behavior
- 👉 Of course, free speech applies online, too

Who Should Be Liable?

- Should newspapers be liable for article comments?
- Should Twitter be liable for illegal tweets?
- Should YouTube be liable for copyright infringement? Terrorist videos?
- Should Snapchat be liable for underage sexting? (What about a similar service marketed to younger teens?)
- What is the proper balance between disintermediation of speech and accountability?

When Can You Trace a Connection?

- Some tracing can be done by individuals; other forms generally require legal process
- 👉 Many parties involved will not turn over data unless compelled to—and sometimes, this refusal is *required* by law
- Legal process generally requires some legal cause of action: libel, threats, (perhaps) harrassment, hacking, etc.
- Sometimes, though, the point of the legal process is just to identify the “culprit”; there may not actually be any real follow-through contemplated

Prenda Law

- They looked for people downloading a single instance of a pornographic movie, filed a lawsuit, and used that to trace the IP address
- They asked for damages a bit below what a “bare-bones defense” would cost, figuring that people would pay up rather than be exposed as downloading porn
- If someone did fight back, they’d drop the case—they never really intended to fight it out in court
- (They also forged evidence, lied to the court, and committed sufficiently many other expenses)
- “It was when the Court realized Plaintiffs engaged their cloak of shell companies and fraud that the Court went to battlestations.”

(<https://ia902603.us.archive.org/17/items/gov.uscourts.cacd.543744/gov.uscourts.cacd.543744.130.0.pdf>)

Tracing a Connection

- Available to recipient (e.g., in mail headers)
- Log files
- Higher layers (e.g., cookies)

Log Files: Mail

```
Feb 22 21:20:26 machshav postfix/smtpd[28530]: connect from  
brinza.cc.columbia.edu[128.59.29.8]  
Feb 22 21:20:26 machshav postfix/smtpd[28530]: 45ECC52D4E9:  
client=brinza.cc.columbia.edu[128.59.29.8]  
Feb 22 21:20:26 machshav postfix/cleanup[8850]: 45ECC52D4E9:  
message-id=<4D03745C-C345-41A8-95E2-EF43F771A045@cs.columbia.edu>  
Feb 22 21:20:26 machshav postfix/qmgr[23733]: 45ECC52D4E9:  
from=<smb@cs.columbia.edu>, size=1023, nrcpt=1 (queue active)  
Feb 22 21:20:26 machshav postfix/smtpd[28530]: disconnect  
from brinza.cc.columbia.edu[128.59.29.8]
```

(recipient not shown here because of spam filter)

What's Interesting?

- IP address of the immediate (but not original) sender
- Timestamp—but no time zone...
- DNS hostname of sender—a spam clue...

```
Feb 22 21:31:53 machshav postfix/smtpd[19642]: connect  
from unknown[222.252.161.130]
```

```
Feb 22 21:31:53 machshav postfix/smtpd[19642]: NOQUEUE:  
reject: RCPT from unknown[222.252.161.130]: 550 5.1.1  
<easycert@machshav.com>: Recipient address rejected:  
User unknown in local recipient table;  
from=<happenedb33@ldbrewer.com> to=<easycert@machshav  
proto=ESMTP helo=<localhost>
```

Web Server Logs

```
209.2.227.65 - - [22/Feb/2010:21:45:07 -0500] "GET /
  HTTP/1.1" 200 401 "-" "Mozilla/5.0 (Macintosh;
  U; Intel Mac OS X 10.6; en-US; rv:1.9.1.8) Gecko/20100202
  Firefox/3.5.8"
209.2.227.65 - - [22/Feb/2010:21:45:07 -0500] "GET /favicon.i
  HTTP/1.1" 404 328
```

Note all of the information about the browser version

Third Party Web Logs

http://images.pcworld.com/shared/graphics/cms/bizdev_msfttout_070609.jpg
http://images.pcworld.com/shared/graphics/cms/bizdev_acer_tout.jpg
<http://ad.doubleclick.net/ad/pcw.main.trackingpixel/DellDHS;sz=1x1>
<http://images.pcworld.com/shared/graphics/cms/DellDealMeetingsmall.jpg>
<http://images.pcworld.com/images/common/adMods/deals2.gif>
http://images.pcworld.com//shared/graphics/cms/LenovoRC_ThinkPadT500.jpg
<http://ad.doubleclick.net/adj/pcw.main.trackingpixel/LenovoDealsModule;sz=1x1>
<http://images.pcworld.com//shared/graphics/cms/Lenovo-ThinkPad-X200.gif>
<http://images.pcworld.com/images/common/v3/mod-header-drkgray.gif>
http://images.pcworld.com/images/common/leftnav_main_bg_sel.png
<http://images.pcworld.com/images/common/v3/shopping/backgrounds/productS...>
http://i.pgcdn.com/pi/73/94/24/739424541_75.jpg

Ads on Web Sites

- Remember that many ads on web sites are from third-party sites
- Each site has a log
- Each log has its own set of IP addresses
- Collect and correlate, especially for attacks on web sites...

Using an IP Address

- We now have the bad guy's IP address
- What we want, though is a person
- How do we track down the target?

Address Registries

```
$ whois -a 128.59.0.0
```

```
OrgName:      Columbia University
OrgID:        COLUMB
Address:      612 W 115TH ST
City:         NEW YORK
StateProv:    NY
PostalCode:   10025
Country:      US

NetRange:     128.59.0.0 - 128.59.255.255
CIDR:         128.59.0.0/16
NetName:      CU-NET
NetHandle:    NET-128-59-0-0-1
Parent:       NET-128-0-0-0-0
NetType:      Direct Assignment
...
```

Contact information is in there, too—does CUIT know the owner?

IP Address Assignment

- Two types, static and dynamic
- Static: simple; consult a file
- Dynamic addresses: handed out for a short time; reclaimed and reassigned later
- Simple: unauthenticated *DHCP*
- More complex: based on some form of authentication, perhaps done by underlying hardware

DHCP

- DHCP—Dynamic Host Configuration Protocol
- Assigns a *lease* to some IP address to the preferred *MAC address*
- A MAC address is manufactured into your network hardware
- It can be overridden, but most people don't know how to
- Most DHCP servers log the lease
- Who owns a given MAC address?

MAC Addresses

- Who owns a given MAC address?
- No a priori way to tell, though the first 3 bytes indicate the manufacturer of the network card
- If the machine is seized, its MAC address can be compared to the DHCP logs
- Some sites require MAC addresses to be registered
- Other sites divert you to a login page

CUIT Can Do This

- They keep logs of connections
- Yes, they can trace abuse
- Also: if they detect a virus-infected machine, its DHCP status is changed to put the machine on an isolated net—download patches and A/V software only

Hackers

- Good hackers steal or make up IP and MAC addresses
- Even if they don't do that, even bad hackers use other people's machines as stepping stones
- Many have “botnets” of thousands—many thousands—of machines belonging to innocent people
- Conclusion: address-tracing goes only so far in locating the real guilty party

Switch Logs

- The site's network hardware can log which IP addresses and which MAC addresses appear on a given port
- Helpful if the attacker is stealing IP and MAC addresses
- For wired networks, can trace the occurrence to a particular wall jack
- Not nearly as useful for WiFi networks; an access point can reach up to 100 meters—more if the attacker has a good antenna

Authenticating Devices

- For some networks, especially wireless ones, the device itself authenticates to the network
- The network provider then has logs associating a user with an IP address
- Again, this is a short-term (but generally renewable) lease

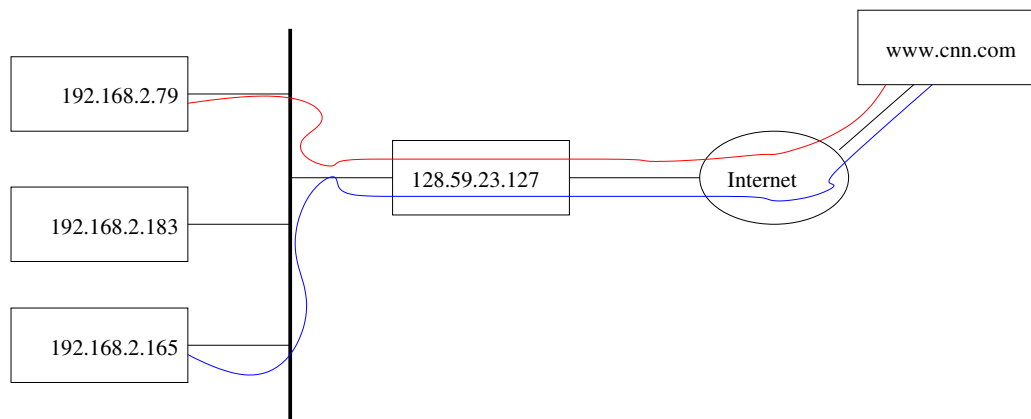
But...

- How long are the DHCP and switch logs retained?
- (What about the mail and web server logs?)
- Are the clocks properly synchronized?
- What time zone are the different logs in?

Network Address Translators

- We're basically out of IP addresses—there aren't enough to go around
- Most homes and many companies use *private address space* (sometimes known as *RFC 1918 space*)
- A *NAT* box (Network Address Translator) at the border translates from private space to a very few public addresses

NAT



Outbound packets will always have the public address of the NAT box. Because there can be multiple connections to a single destination, the source port number is also changed to allow disambiguation and routing of return packets.

Translations

- A packet from 192.168.2.79:2345 is sent to www.cnn.com:80
- Another machine sends traffic from 192.168.2.165:7890 to the same place
- After translation, they appear to be from 128.59.23.127:45678 and 128.59.23.127:46324
- The translation is reversed on inbound packets

Logs: Lost in Translation

- Most NAT boxes do not keep logs of translations
- They can't—it would have to be one per TCP connection
- Even if they did, it wouldn't help—receiving site logs do not include port numbers
- Attacks can be traced to the NAT, but rarely beyond it

Other Means of Attribution

- Remember all of those third-party web ads?
- They all have cookies and logs, and cookies pass unchanged through NATs and Tor networks
- Maybe one of those ads also appears on some site where the bad guy has an account

Example: Cookie Crumb Tracing

- The bad guy attacks a web site via a page that has a Doubleclick cookie
- Doubleclick also serves ads on a NY Times page that person visits
- The NY Times registration is tied to the attacker's home subscription to the paper edition of the Times
- That, of course, is tied to a physical address

Buts...

- You have to get logs from three different web sites to establish the linkage
- You have to get address data from a site that has no connection to the attack
- It takes persistence and court orders—and money...

Who Can Do All This?

- Law enforcement, with search warrants
- Plaintiffs in civil suits—if they have deep pockets or expect to win a big settlement
- Anti-terrorism investigators, with “National Security Letters”?
- What are the limits?

The Limits of Traceability

- In ideal circumstances—good logs, no evasive action, one jurisdiction, etc.—it’s generally feasible to trace connections to a building
- Tracing past there can be difficult; you may need subsidiary evidence
- Hackers generally use “stepping stones” to launch real attacks
- (Poorly protected WiFi networks can be abused by outsiders)
- It can take significant effort, though, and there are often breaks that you can’t go past
- Without legal process or an application-level leak, tracing can be difficult or impossible