Botnets and Applications

Nick Feamster
CS 7260
March 6, 2006

Administrivia

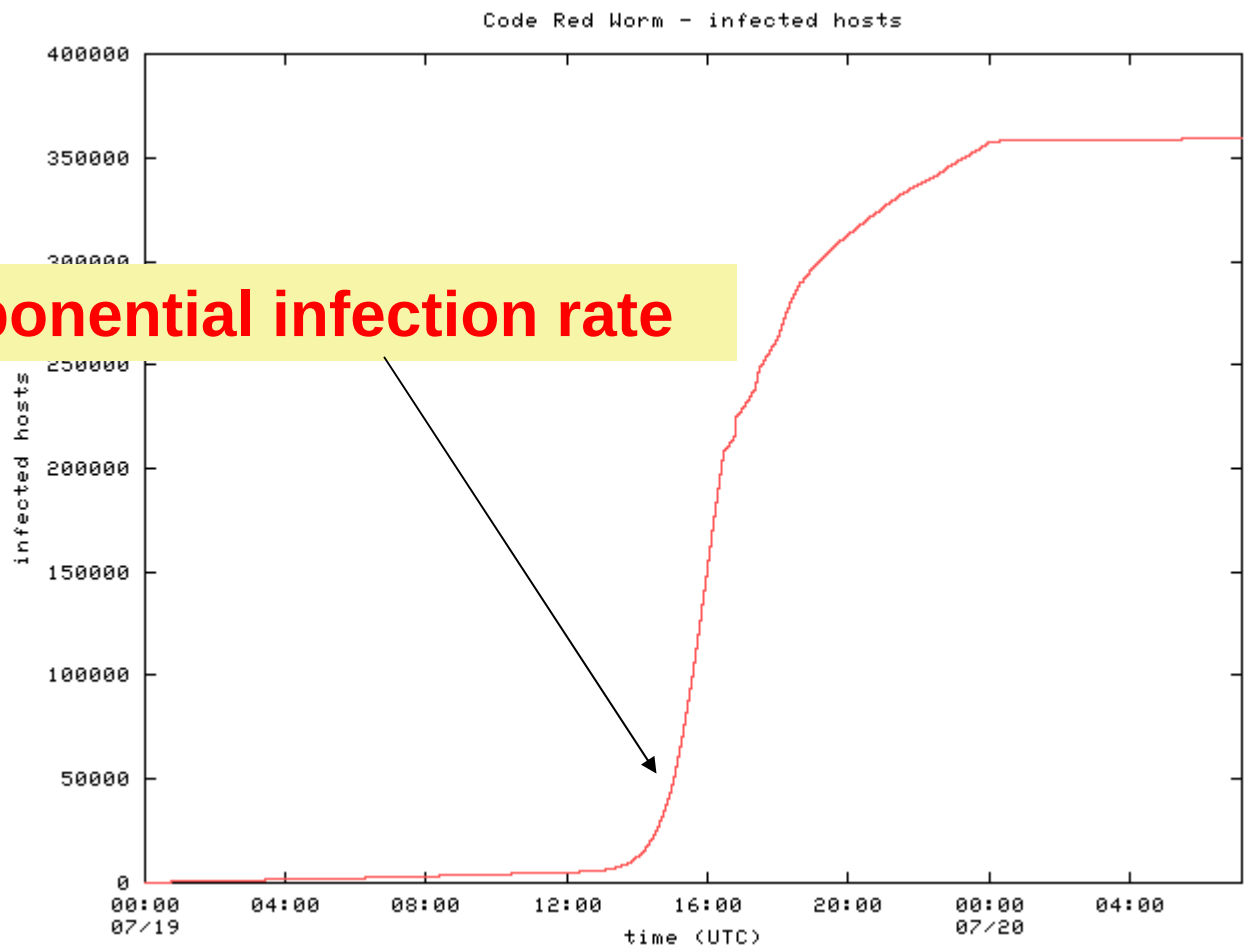
- Quizzes are all graded, will go back Wednesday
- Interim reports and presentations: 3 weeks
  - Turn in a writeup that includes:
    - Draft of related work section
    - Current status
    - Any results so far
  - 10-minute presentation that includes the same

Today's Lecture: Botnets

- DDoS Wrapup
- What is a botnet?
- How is a botnet controlled?
- Utilities: Rootkits
- Propagation
  - Dagon *et al., Modeling Botnet Propagation using Timezones*
- What can be done with a botnet?
  - Spam
  - Phishing
  - Click fraud
  - Identity theft
  - DDoS
- Detection, Tracking, and Mitigation
  - Honeynets

Code Red: Host Infection Rate

Measured using backscatter technique



Exponential infection rate

Designing Fast-Spreading Worms

- **Hit-list scanning**
  - Time to infect first 10k hosts dominates infection time
  - **Solution:** Reconnaissance (stealthy scans, etc.)
- **Permutation scanning**
  - **Observation:** Most scanning is redundant
  - **Idea:** Shared permutation of address space. Start scanning from own IP address. Re-randomize when another infected machine is found.
- **Internet-scale hit lists**
  - *Flash worm:* complete infection within 30 seconds

Recent Advances: Slammer

- February 2003
- Exploited vulnerability in MS SQL server
- Exploit fit into a single UDP packet
  - ***Send and forget!***
- Lots of damage
  - BofA, Wash. Mutual ATMs unavailable
  - Continental Airlines ticketing offline
  - Seattle E911 offline

Scary recent advances: Witty

- March 19, 2004
- Single UDP packet exploits flaw in the *passive analysis* of Internet Security Systems products.
- “Bandwidth-limited” UDP worm ala’ Slammer.
- Initial spread seeded via a *hit-list*.
- All 12,000 vulnerable hosts infected within 45 mins
- **Payload:** slowly corrupt random disk blocks

Why Denial-of-Service "Works"

- **Asymmetry:** generating a request is cheaper than formulating a response
- One attack machine can generate a lot of requests, and effectively multiply its power
- Not always possible to achieve this asymmetry

Why does *Distributed* DoS work?

- Simplicity
- “On by default” design
- Readily available zombie machines
- **Attacks look like normal traffic**
- Internet’s federated operation obstructs cooperation for diagnosis/mitigation

Research: DoS-Resistant Architectures

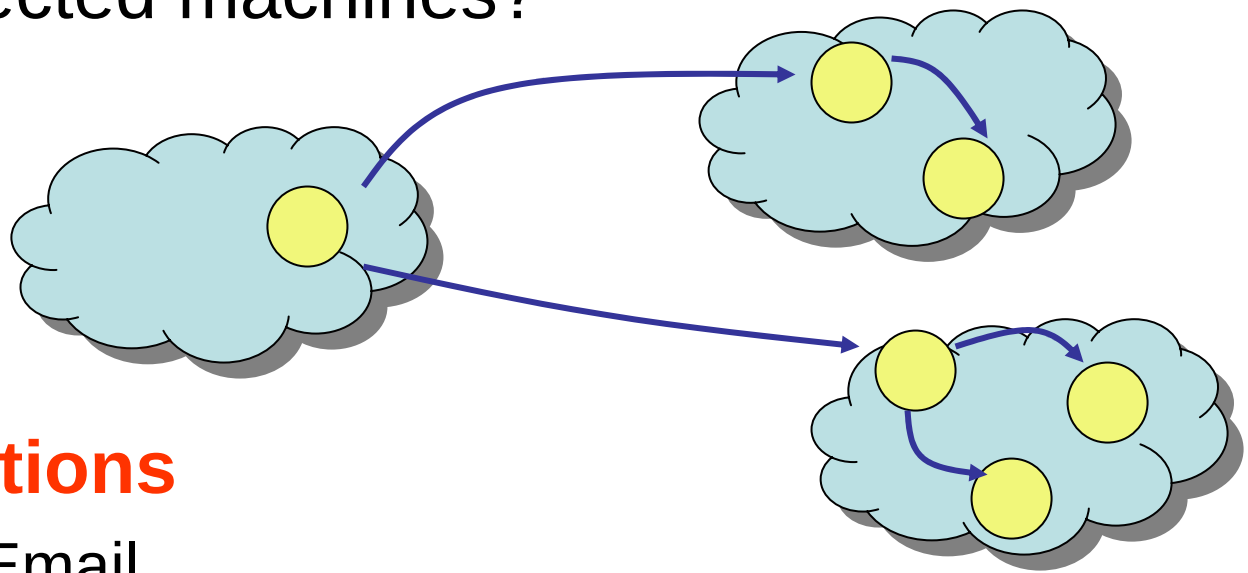
- **Identity:** better notion of who is sending traffic
- **Resource allocation:** controlling and limited access to network resources
- **Detection:** tracking and mitigating malicious flows and “unwanted packets”
- **Accountability:** assigning responsibility for bad packets or packet streams
- **Epidemiology:** tracking machine infections

Botnets

- **Bots:** Autonomous programs performing tasks
- Plenty of “benign” bots
  - *e.g.,* weatherbug
- **Botnets:** group of bots
  - Typically carries malicious connotation
  - Large numbers of infected machines
  - Machines “enlisted” with infection vectors like worms (last lecture)
- Available for **simultaneous control** by a master
- *Size:* up to 350,000 nodes (from today’s paper)

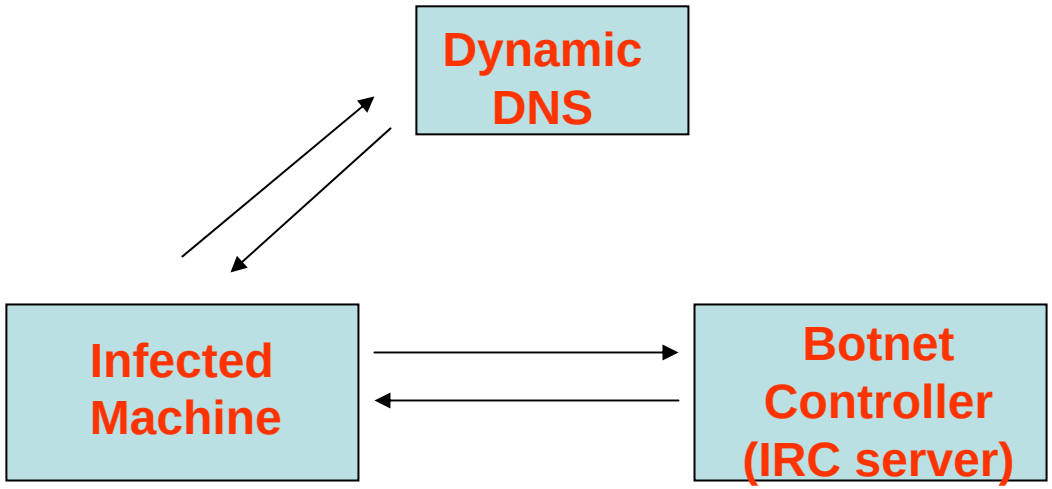
“Rallying” the Botnet

- Easy to combine worm, backdoor functionality
- **Problem:** how to learn about successfully infected machines?



- **Options**
  - Email
  - Hard-coded email address

Botnet Control



- Botnet master typically runs some IRC server on a well-known port (*e.g.*, 6667)
- Infected machine contacts botnet with pre-programmed DNS name (*e.g.*, big-bot.de)
- **Dynamic DNS:** allows controller to move about freely

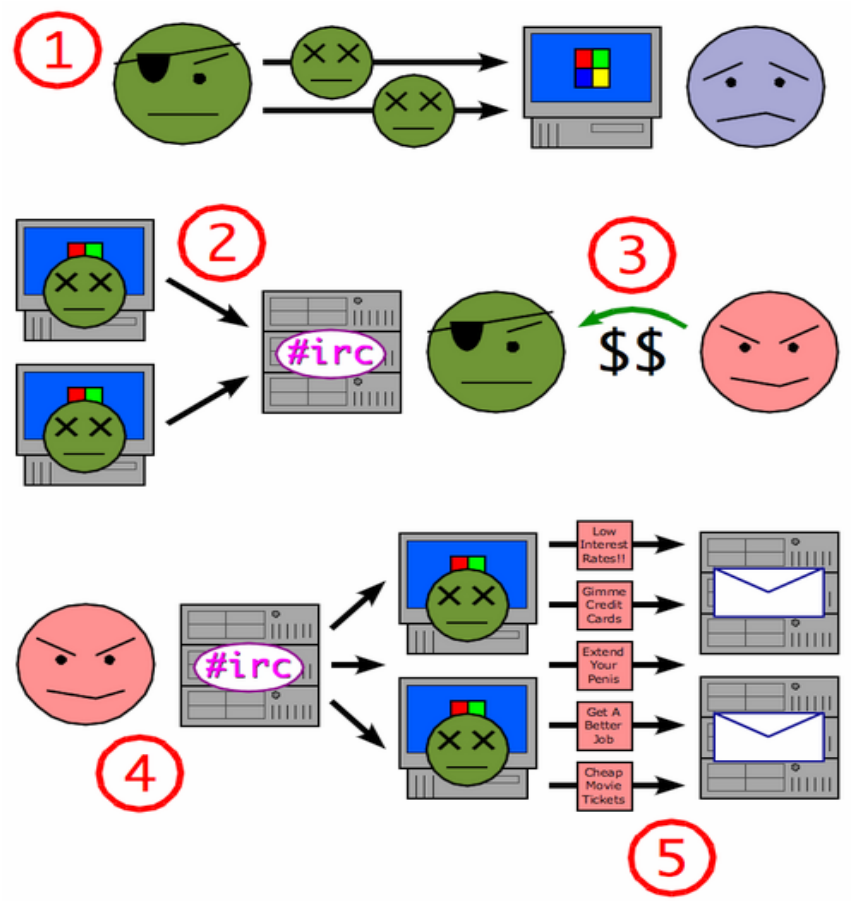
Botnet History: How we got here

- **Early 1990s:** IRC bots
  - eggdrop: automated management of IRC channels
- **1999-2000:** DDoS tools
  - Trinoo, TFN2k, Stacheldraht
- **1998-2000:** Trojans
  - BackOrifice, BackOrifice2k, SubSeven
- **2001- :** Worms
  - Code Red, Blaster, Sasser

Fast spreading capabilities pose big threat

Put these pieces together and add a controller…

Putting it together



1. Miscreant (botherd) launches worm, virus, or other mechanism to infect Windows machine.
2. Infected machines contact botnet controller via IRC.
3. Spammer (sponsor) pays miscreant for use of botnet.
4. Spammer uses botnet to send spam emails.

Top sources of botnet controllers

ASN	AS Name	Unique C&Cs
6517	YIPESCOM - Yipes Communication	60
21840	SAGONET-TPA - Sago Networks	90
25761	STAMINUS-COMM - Staminus Commu	86
4766	KIXS-AS-KR Korea Telecom	43
13680	AS13680 Hostway Corporation Ta	22
21698	NEBRIX-CA - Nebrix Communicati	24
13301	UNITEDCOLO-AS Autonomous Syste	27
21788	NOC - Network Operations Cente	29
29415	EUROWAN-ASN OVANET - EuroWan d	16
13749	EVERYONES-INTERNET - Everyones	24
30083	SERVER4YOU - Server4You Inc.	21
25700	SWIFTDESK - SWIFTDESK VENTURE	13
23522	CIT-FOONET - CREATIVE INTERNET	14
27595	ATRIVO-AS - Atrivo	31
13237	LAMBDANET-AS European Backbone	11

As of June 7, 2005, data from Prof. Randall Vaughnposted to NANOG.

Tools: Rootkits

- Software to set up and maintain an environment on a compromised machine
  - **Binary:** Replace system files with trojan counterpart
  - **Kernel:** Uses kernel components
  - **Library:** Uses system library trojans
- **Key idea:** *conceal presence*

Binary Rootkits: Defeat Auditing

- Auditing commands would show nefarious activity
  - **last:** what accounts intruders were using, where they were coming from, and when they were in your system.
  - **ls:** files
  - **ps:** the sniffer, password cracking program, and anything else being run by the intruders
  - **netstat:** the current network connections and ports on which listening for incoming connections
  - **ifconfig:** if the ethernet interface was in promiscuous mode
- Rootkit “trojans” these commands
  - Usually precompiled for particular platform
  - Script places the binaries over the old one

Binary Kits

- Rootkit's tools deploy in a hidden directory
- Some common locations found

  /dev/.hdd

  /dev/.lib

  /usr/src/.poop

  /usr/src/linux/arch/alpha/lib/.lib/.lproc
- Invisible special characters used in dir names to make the detection and deletion harder
- Tools to adjust timestamps and sizes of trojans to match the original (touch files)

Kernel Rootkits

- First reported in 1997
- Modify system calls
  - Applications run in user mode
  - Hardware device interaction happens in kernel mode
  - Hence the severity
- **Example**
  - If open() call meant “ get to disk and open file from this location” could be changed to “ get to disk and open file from this location unless its name is “rootkit” “
- Whole operating system becomes untrustworthy

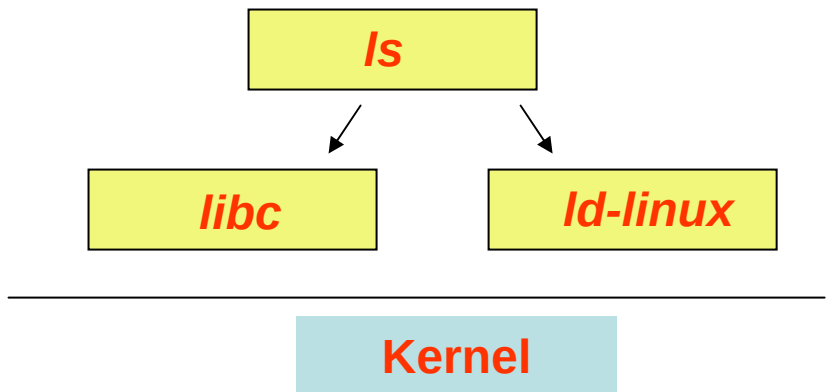
Mitigation: Primitive State-of-the-Art

- Filtering port 6667
- Command and control filtering
  - Port numbers
  - IP addresses
  - **Problem:** need to locate command and control

Library Kits

- Replaces the standard system library
- **Interposition:** Can be positioned in such a way that they will be loaded before the system libraries

```
% ldd /bin/ls
        linux-gate.so.1 =>  (0x00c3a000)
        librt.so.1 => /lib/librt.so.1 (0x0020f000)
        libacl.so.1 => /lib/libacl.so.1 (0x001e8000)
        libselinux.so.1 => /lib/libselinux.so.1 (0x00ac1000)
        libc.so.6 => /lib/libc.so.6 (0x0092b000)
        libpthread.so.0 => /lib/libpthread.so.0 (0x008ae000)
        /lib/ld-linux.so.2 (0x0090d000)
        libattr.so.1 => /lib/libattr.so.1 (0x001e2000)
```



Digression: Other Recent Rootkits

"Most people, I think, don't even know what a rootkit is, so why should they care about it?"

-Thomas Hesse

- Sony DRM Rootkit
  (identified by Mark Russinovich, Oct 31, 2005)
  - Stealth installation and concealment
  - Removal rendered CD unplayable
  - masks files whose filenames start with "sys"
- **Nov. 10:** taken advantage of by virus writers
- **Nov. 10:** class action lawsuits
- **Nov. 16:** massive recall of 2+ million CDs
- **Nov. 17:** DRM uninstaller found to be worse than rootkit
  - Uninstaller installs ActiveX component that can be exploited by attacker sites

Botnet Propagation

- Email (social engineering)
- Remote vulnerabilities
- Webpages
- “Seed” botnets
  - Faster scanning of unpatched systems
- **Worms**

Botnet Detection and Tracking

- Network Intrusion Detection Systems (*e.g.*, Snort)
  - **Signature:** alert tcp any any -> any any (msg:"Agobot/Phatbot Infection Successful"; flow:established; content:"221
- **Honeynets:** gather information
  - Run unpatched version of Windows
  - Usually infected within 10 minutes
  - **Capture binary**
    - determine scanning patterns, etc.
  - **Capture network traffic**
    - Locate identity of command and control, other bots, etc.

Detection: In-Protocol

- Snooping on IRC Servers
- Email (*e.g.*, CipherTrust ZombieMeter)
  - > 170k new zombies per day
  - 15% from China
- Managed network sensing and anti-virus detection
  - Sinkholes detect scans, infected machines, etc.

- **Drawback:** Cannot detect botnet structure

Using DNS Traffic to Find Controllers

- Different types of queries may reveal info
  - Repetitive A queries may indicate bot/controller
  - MX queries may indicate spam bot
  - PTR queries may indicate a server
- Usually 3 level: hostname.subdomain.TLD
- Names and subdomains that just look rogue
  - (*e.g.,* irc.big-bot.de)

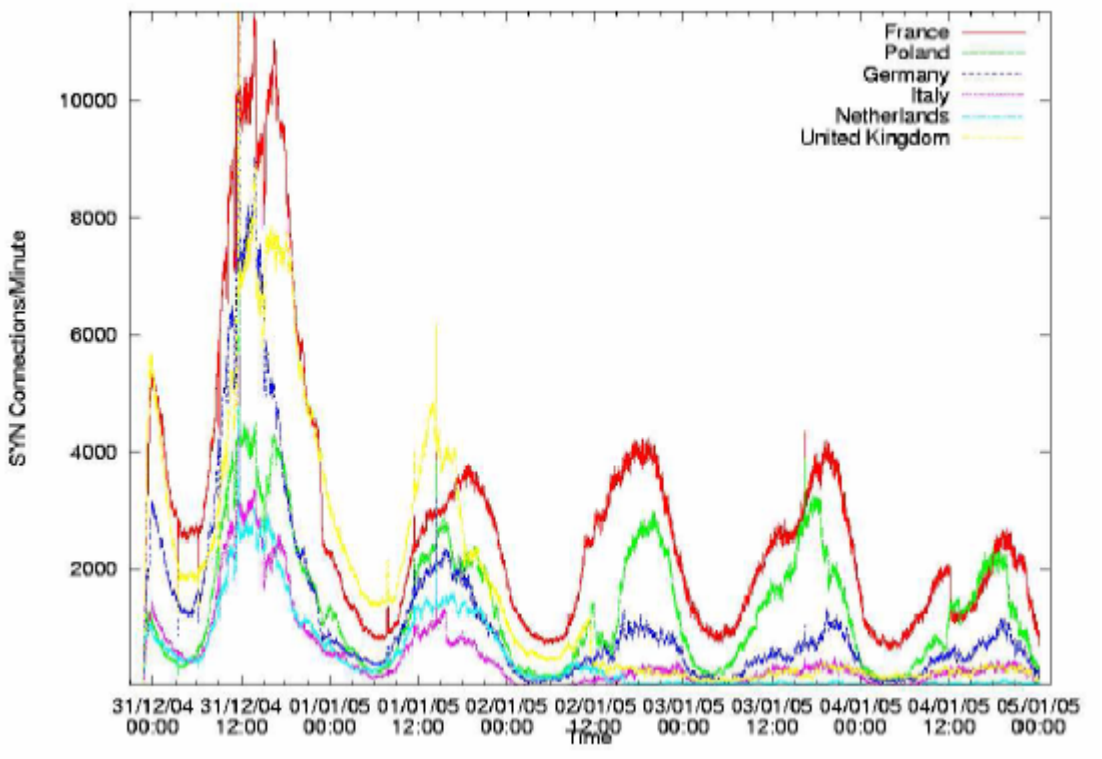
DNS Monitoring

- Command-and-control hijack
  - **Advantages:** accurate estimation of bot population
  - **Disadvantages:** bot is rendered useless; can't monitor activity from command and control

- Complete TCP three-way handshakes
  - Can distinguish distinct infections
  - Can distinguish infected bots from port scans, etc.

Modeling Botnet Propagation

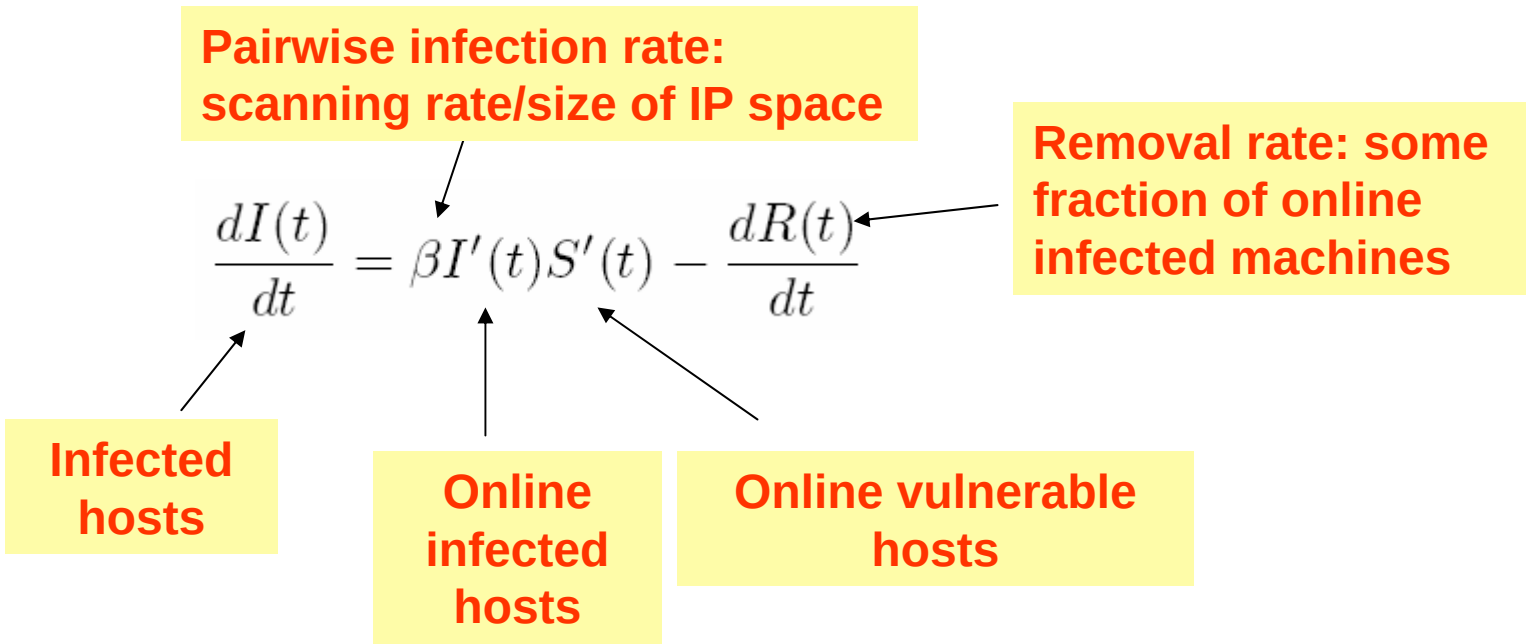
- Heterogeneous mix of vulnerabilities
- Diurnal patterns



Diurnal patterns can have an effect on the rate of propagation

Can model spread of the botnet based on short-term propagation.

Modeling Propagation: Single TZ



$$S(t) = N(t) - I(t) - R(t)$$

- Useful for modeling the spread of “regional worms”
- **Question:** How common is this?
- Extension to multiple timezones is (reasonably) straightforward

Spread across multiple timezones

$$\frac{dI_i(t)}{dt} = \alpha_i(t)[N_i(t) - I_i(t) - R_i(t)] \cdot \sum_{j=1}^{K} \beta_{ji}\alpha_j(t)I_j(t) - \gamma_i\alpha_i(t)I_i(t)$$

Newly infected hosts in timezone *i*

Online vulnerable hosts in timezone *i*

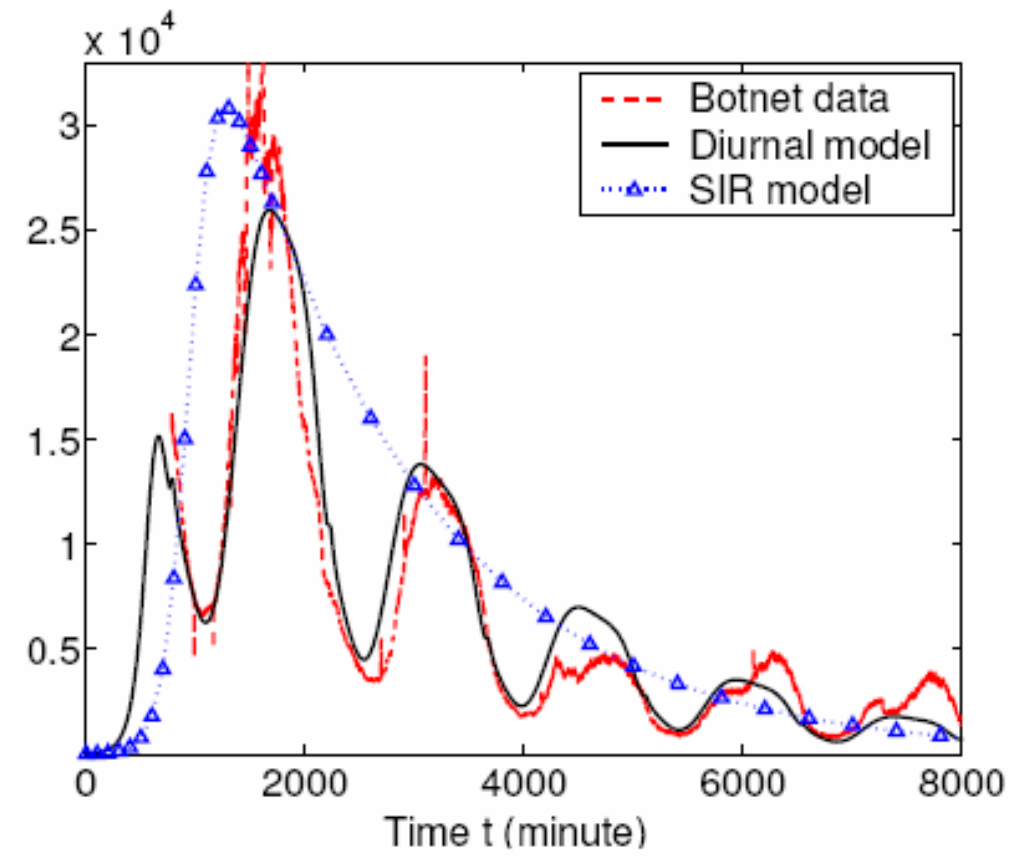
Infection from zone *j* to *i*

- **Question:** What assumption is being made regarding scanning rates and timezones?

Experimental Validation

- How to capture various parameters?
  - Derive diurnal shaping function by country
  - Monitor scanning activity per hour, per day (24 bins)
  - Normalize each day to 1 and curve-fit
- How to estimate N(t) per timezone?

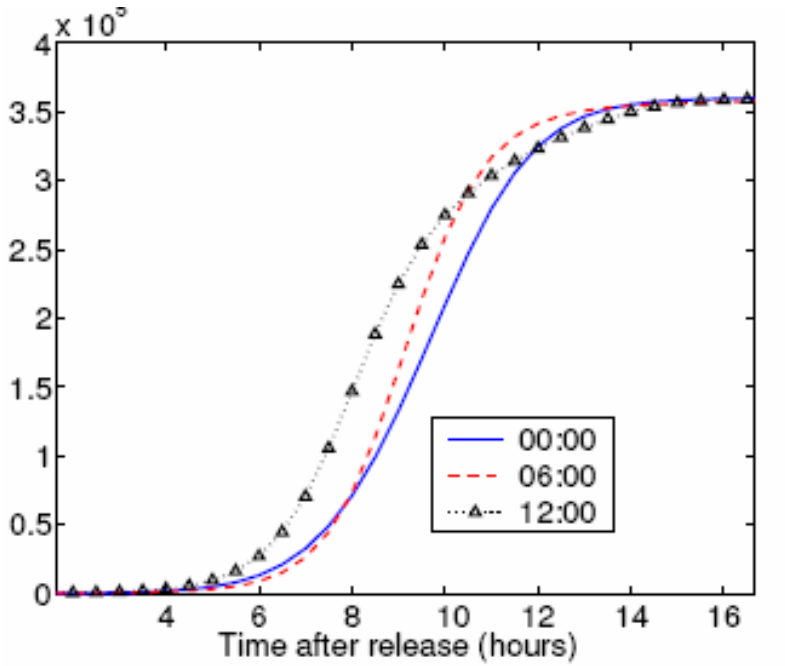
Fitting the model to the data



Diurnal shaping function yields more accurate model.

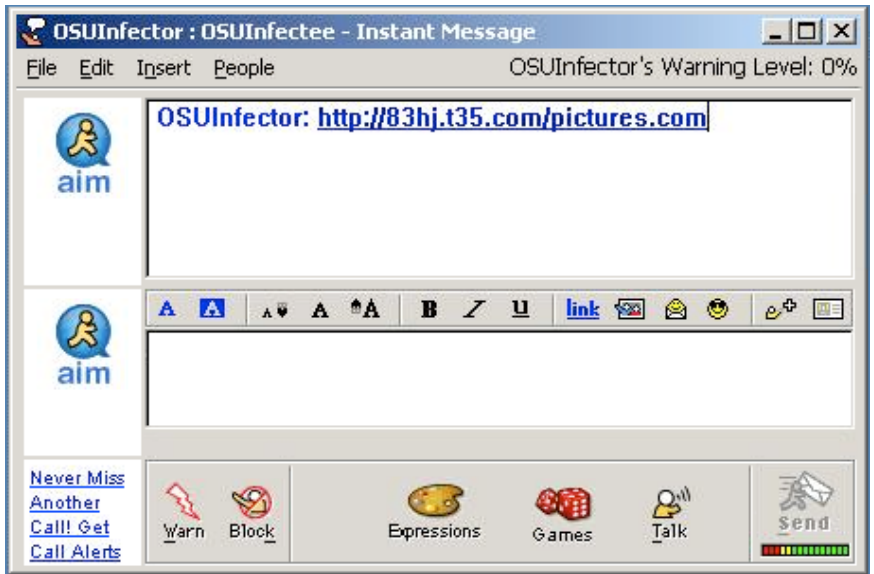
Applications of the model

- Forecasting the spread of botnets
- Improved monitoring and response capabilities
  - A faster spreading worm may be “stealth” depending on the time of day that the worm was released



New Trend: Social Engineering

- Bots frequently spread through AOL IM
  - A bot-infected computer is told to spread through AOL IM
  - It contacts all of the logged in buddies and sends them a link to a malicious web site
  - People get a link from a friend, click on it, and say “sure, open it” when asked



Early Botnets: AgoBot (2003)

- Drops a copy of itself as svchost.exe or syschk.exe
- Propagates via Grokster, Kazaa, etc.
- Also via Windows file shares

Botnet Operation

- General
  - Assign a new random nickname to the bot
  - Cause the bot to display its status
  - Cause the bot to display system information
  - Cause the bot to quit IRC and terminate itself
  - Change the nickname of the bot
  - Completely remove the bot from the system
  - Display the bot version or ID
  - Display the information about the bot
  - Make the bot execute a .EXE file
- IRC Commands
  - Cause the bot to display network information
  - Disconnect the bot from IRC
  - Make the bot change IRC modes
  - Make the bot change the server Cvars
  - Make the bot join an IRC channel
  - Make the bot part an IRC channel
  - Make the bot quit from IRC
  - Make the bot reconnect to IRC
- Redirection
  - Redirect a TCP port to another host
  - Redirect GRE traffic that results to proxy PPTP VPN connections
- DDoS Attacks
  - Redirect a TCP port to another host
  - Redirect GRE traffic that results to proxy PPTP VPN connections
- Information theft
  - Steal CD keys of popular games
- Program termination

PhatBot (2004)

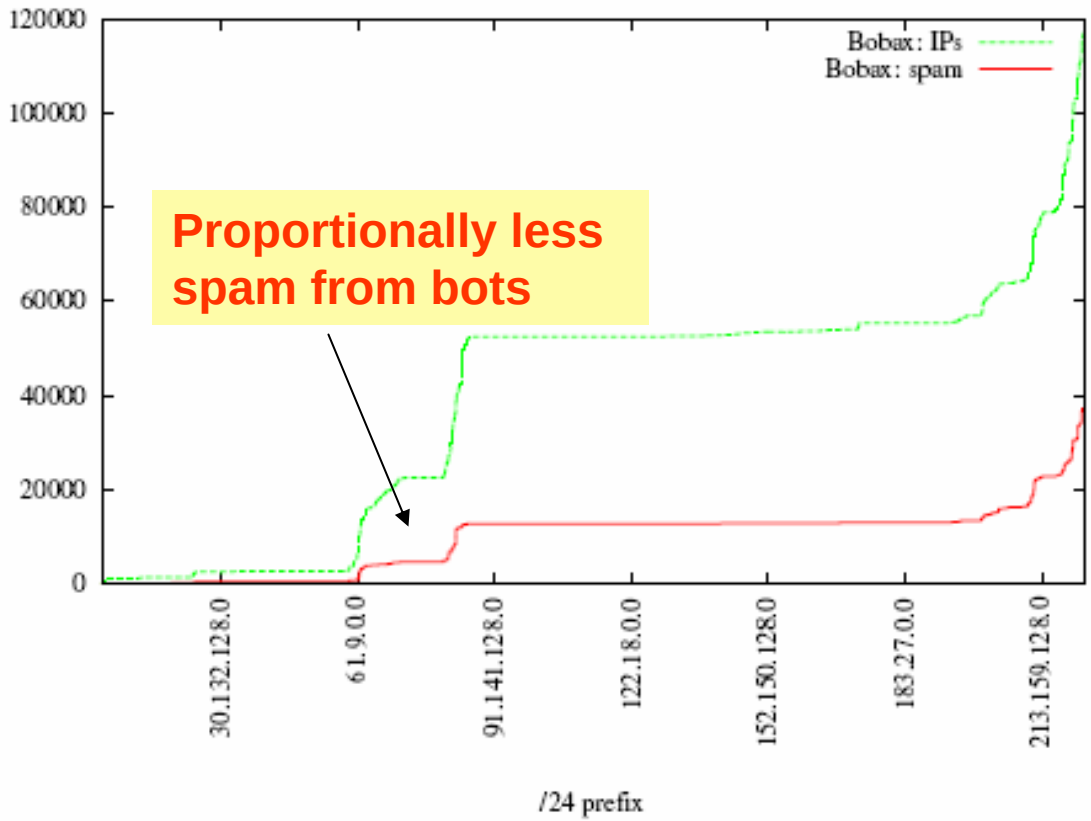
- Direct descendent of AgoBot
- More features
  - Harvesting of email addresses via Web and local machine
  - Steal AOL logins/passwords
  - Sniff network traffic for passwords
- Control vector is peer-to-peer (not IRC)

Peer-to-Peer Control

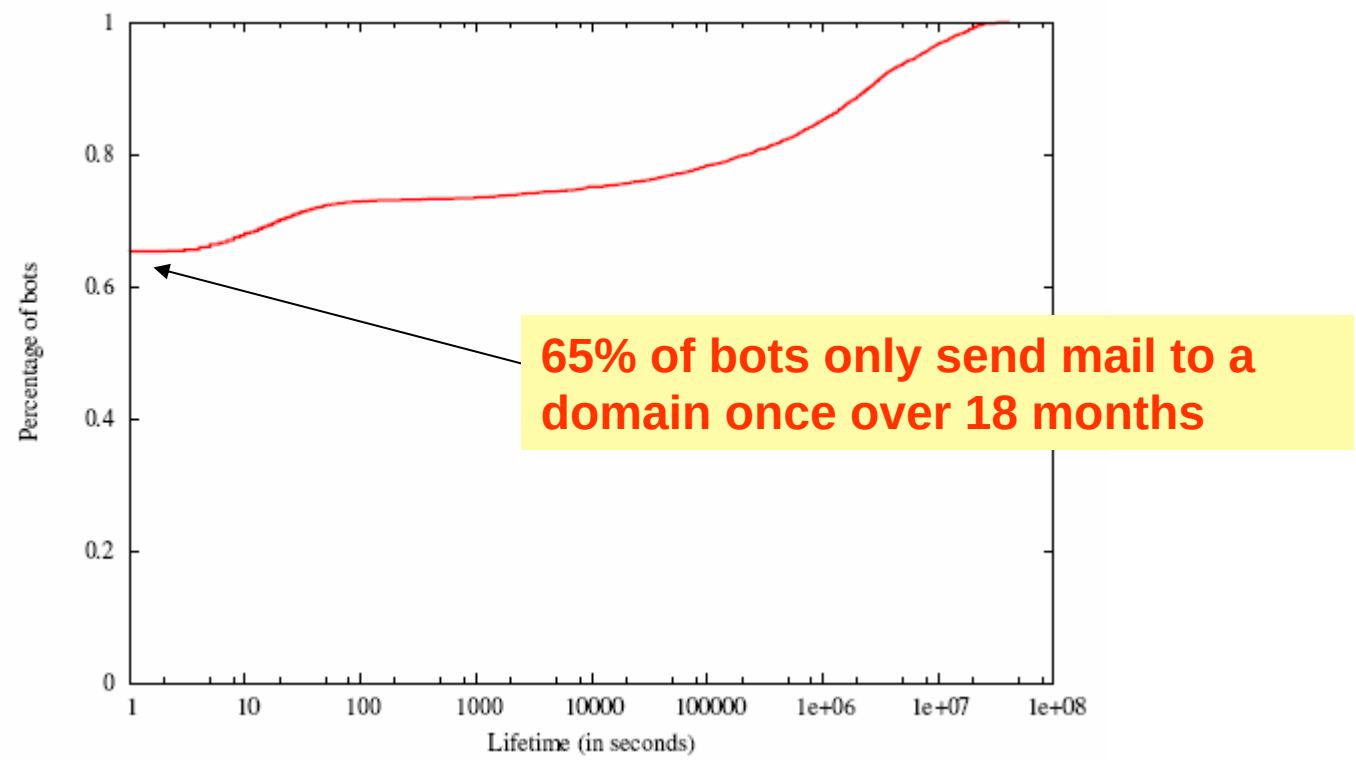
- **Good**
  - distributed C&C
  - possible better anonymity
- **Bad**
  - more information about network structure directly available to good guys IDS,
  - overhead,
  - typical p2p problems like partitioning, join/leave, etc

Botnet Application: Spam

- **Example:** Bobax
  - Approximate size: 100k bots

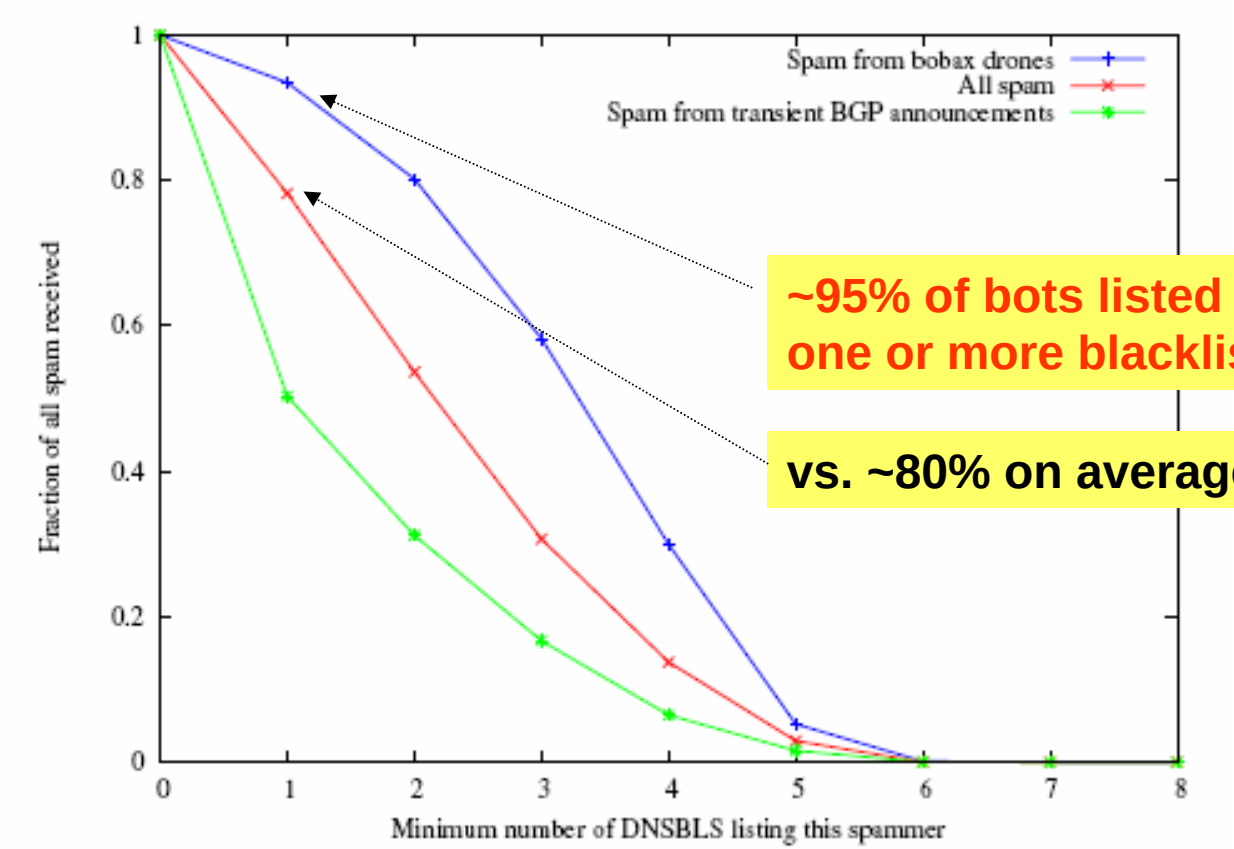


Most Bot IP addresses do not return



Collaborative spam filtering seems to be helping track bot IP addresses

Blacklisting Seems to Work Pretty Well



- Premium for spamming bots that are not blacklisted

Botnet Application: Phishing

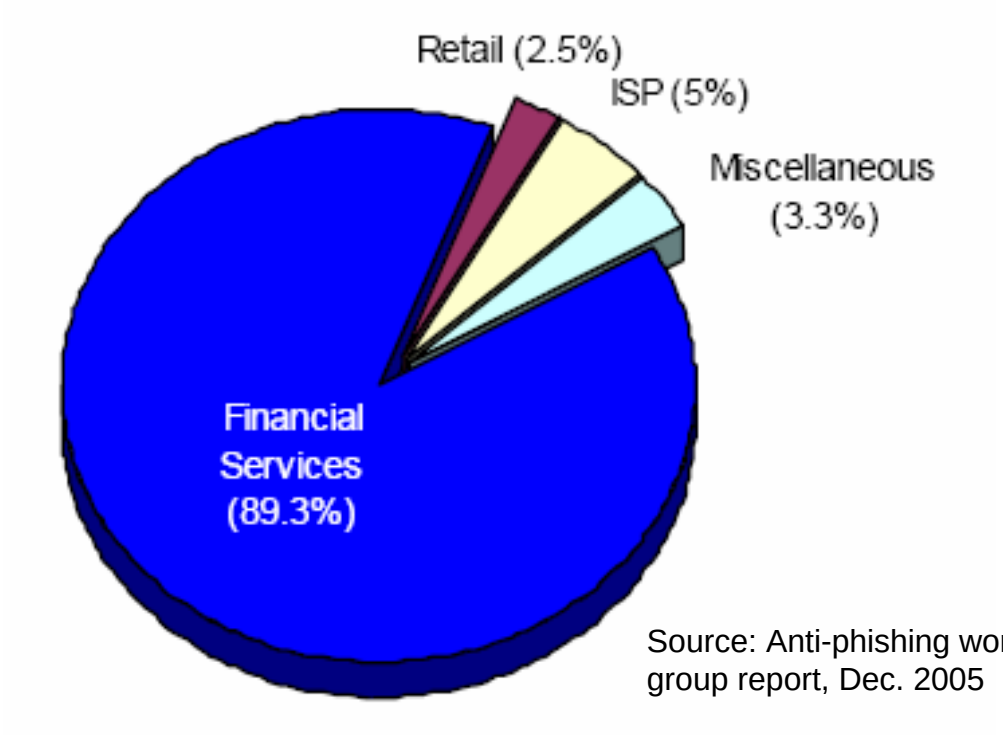
"Phishing attacks use both **social engineering** and **technical subterfuge** to steal consumers' personal identity data and financial account credentials." -- Anti-spam working group

- Social-engineering schemes
  - Spoofed emails direct users to counterfeit web sites
  - Trick recipients into divulging financial, personal data

- Anti-Phishing Working Group Report (Oct. 2005)
  - 15,820 phishing e-mail messages 4367 unique phishing sites identified.
  - 96 brand names were hijacked.
  - Average time a site stayed on-line was 5.5 days.

Question: What does phishing have to do with botnets?

Which web sites are being phished?



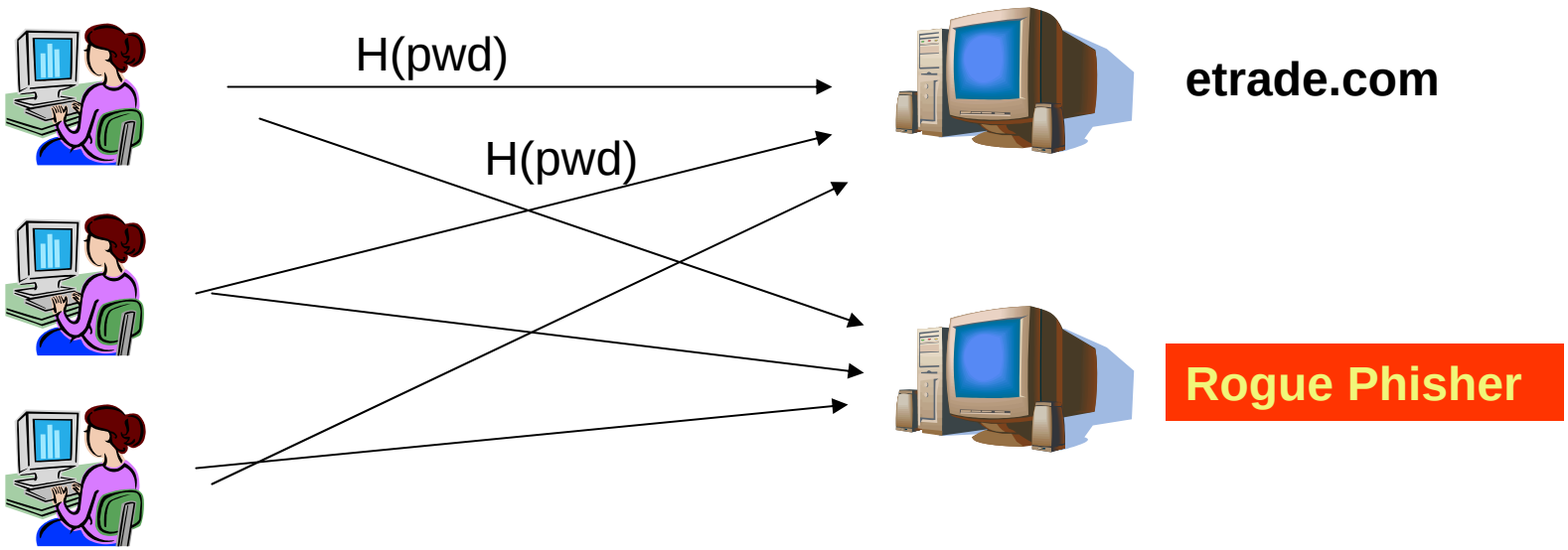
Source: Anti-phishing working group report, Dec. 2005

- Financial services by far the most targeted sites

New trend: **Keystroke logging…**

Phishing: Detection and Research

- **Idea:** Phishing generates sudden uptick of password re-use at a brand-new IP address



Distribution of password harvesting across bots can help.

Botnet Application: Click Fraud

- Pay-per-click advertising
  - **Publishers** display links from **advertisers**
  - **Advertising networks** act as middlemen
    - Sometimes the same as publishers (*e.g.*, Google)
- **Click fraud:** botnets used to click on pay-per-click ads
- **Motivation**
  - Competition between advertisers
  - Revenue generation by bogus content provider

Open Research Questions

- Botnet membership detection
  - Existing techniques
    - Require special privileges
    - Disable the botnet operation
  - Under various datasets (packet traces, various numbers of vantage points, etc.)
- Click fraud detection
- Phishing detection