Cybercrime and the Cyber Underground Economy

Richard A. Kemmerer
Computer Security Group
Department of Computer Science
University of California, Santa Barbara

http://seclab.cs.ucsb.edu
Evolution of Internet Threats

- Viruses
- Worms
- DDOS
- Phishing
- Scamming
- Botnets
- Targeted Attacks
- Cyberwar

Damage:
- Billions
- Millions
- Hundreds of Thousands
- Thousands
- Hundreds
Perfect Storm

• Hackers realize potential monetary gains associated with Internet fraud
  – Shift from “hacking for fun” to “hacking for profit”
• Traditional crime organizations realize the potential of the Internet for their endeavors
• Integration of hacker’s sophisticated computer attacks with organized crime’s well-established fraud attacks results in an underground economy similar to legitimate economies
Cybercrime

• General definition
  – Using a computer connected to the Internet for criminal purposes
    • Financial
    • Theft of intellectual property
    • Service disruption

• Organized cybercrime
  – Professional organizations that make a living off of electronic crimes
  – Affiliate networks
  – Commercial services
Financial Impacts of Cybercrime

- Businesses lose an estimated $1 trillion dollars annually (source McAfee)
- An estimated loss of $599.7 million dollars reported to the FBI in 2009
Cybercrime Impact

Norton Study Calculates Cost of Global Cybercrime: $114 Billion Annually
One of World’s Largest Cybercrime Studies Reveals More Than One Million Victims a Day

March 16, 2011

CYBERSECURITY

Continued Attention Needed to Protect Our Nation’s Critical Infrastructure and Federal Information Systems
Cyber Underground Economy

• Trades compromised hosts, personal information, and services
• Makes it possible to significantly increase the scale of the frauds carried out on the Internet
• Allows criminals to reach millions of potential victims
• Criminals are using
  – IRC channels to verify stolen credit cards
  – e-casinos to launder money
  – fast-flux networks to create attack-resilient services
UCSB Computer Security Group
Cybercrime Research

- Phishing
- Spam / Botnets / Drive-by-Downloads
- Social Network Fraud
- Fake Antivirus
- Online Ad Fraud
UCSB Computer Security Group’s Approach to Cybercrime Research

• Developing novel techniques and tools to analyze the underground economy
• Goal is to obtain a comprehensive picture of the complete criminal process
• Create models of
  – Cyber-underground market
  – Actors in the market
  – Processes and interactions between actors
  – Underlying infrastructure
• Leverage these models and develop techniques to disrupt parts of the criminal process
How to Steal a Botnet and What Can Happen When You Do
Botnet Terminology

• **Bot**
  – an application that performs some action or set of actions on behalf of a remote controller
  – installed on a victim machine (zombie)
  – modular (plug in your functionality/exploit/payload)

• **Botnet**
  – network of infected machines controlled by a malicious entity

• **Control channel**
  – required to send commands to bots and obtain results and status messages
  – usually via IRC, HTTP, HTTPs, or Peer-to-Peer

• **Bot Herder**
  – aka botmaster or controller
  – owns control channel, sends commands to botnet army
  – motivations are usually power or money
Torpig

• Trojan horse
  – distributed via the Mebroot “malware platform”
  – injects itself into 29 different applications as DLL
  – steals sensitive information (passwords, HTTP POST data)
  – HTTP injection for phishing
  – uses “encrypted” HTTP as C&C protocol
  – uses domain flux to locate C&C server

• Mebroot
  – spreads via drive-by downloads
  – sophisticated rootkit (overwrites master boot record)
Torpig: Behind the scenes

- Innocent victim
  - Commands + torpig
  - STOLEN DATA

- "Hacked" web servers
  - Mebroot download

- "Drive-By Download" server
  - Injection server
  - Mebroot C&C
  - Torpig C&C
Torpig HTML Injection

- Domains of interest (~300) stored in configuration file
- When domain of interest visited
  - Torpig issues request to injection server
  - Server specifies a *trigger page* on target domain and a URL on injection server to be visited when user visits trigger page
- When user visits the trigger page
  - Torpig requests injection URL from injection server
  - Torpig injects the returned content into the user’s browser
- Content is usually HTML phishing form that asks for sensitive data
  - Reproduces look and style of target web site
Example Phishing Page
Example Phishing Page
Domain Flux

• Taking down a single bot has little effect on botmaster

• C&C servers are vulnerable to take down
  – if you use a static IP address, people will block or remove host
  – if you use a DNS name, people will block or remove domain name

• Domain flux
  – idea is to have bots periodically generate new C&C domain names
  – often, use local date (system time) as input
  – botmaster needs to register one of these domains
    and respond properly so that bots recognize valid C&C server
  – defenders must register all domains to take down botnet
Torpig Domain Flux

• Each bot has
  – same domain generation algorithm (DGA)
  – three fixed domains to be used if all else fails

• DGA generates
  – weekly domain name (wd)
  – daily domain name (dd)

• Every 20 minutes bot attempts to connect (in order) to
  – wd.com, wd.net, wd.biz
  – if all three fail, then dd.com, dd.net, dd.biz
  – if they also fail, then the three fixed domains

• Criminals normally registered wd.com (and wd.net)
Sinkholing Torpig C&C Overview

- Reverse engineered name generation algorithm and C&C protocol
- Observed that domains for 01/25 – 02/15 unregistered
- Registered these domains ourselves
- Unfortunately, Mebroot pushed new Torpig binary on 02/04
- We controlled the botnet for ~10 days
- Data
  - 8.7 GB Apache logs
  - 69 GB pcap data (contains stolen information)
Sinkholing Torpig C&C

- Purchased hosting from two different hosting providers known to be unresponsive to complaints
- Registered wd.com and wd.net with two different registrars
  - One was suspended 01/31 due to abuse complaint
- Set up Apache web servers to receive bot requests
- Recorded all network traffic
- Automatically downloaded and removed data from our hosting providers
- Enabled hosts a week early
  - immediately received data from 359 infected machines
Data Collection Principles

• Principle 1: the sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized
  – always responded with okn message
  – never sent new/blank configuration file
  – removed data from servers regularly
  – stored data offline in encrypted form

• Principle 2: the sinkholed botnet should collect enough information to enable notification and remediation of affected parties
  – worked with law enforcement (FBI and DoD Cybercrime units)
  – worked with bank security officers
  – worked with ISPs
Data Collection

• Bot connects to Torpig C&C every 20 minutes via HTTP POST

• Sends a header
  – timestamp, IP address, proxy ports, OS version, locale, nid, Torpig build and version number

• nid
  – 8 byte value, used for encrypting header and data
  – derived from hard disk information or volume serial number
  – serves as a convenient, unique identifier
  – allows one to detect VMware machines

• Optional body data
  – stolen information (accounts, browser data, …)
Size Estimation

- Count number of infections
  - usually based on unique IP addresses
  - problematic: DHCP and NAT effects (we saw 1.2M unique IPs)
  - our count based on header information: ~180K hosts (nids) seen
Size Estimation

- Cummulative number of infections
  - linear for unique IP addresses
  - decayed quickly for unique nids
  - more than 75% of unique nids were observed in first 48 hours
Threats

• Theft of financial data
• Denial of service
• Proxy servers
• Privacy threats
Theft of Financial Information

- 8,310 unique accounts from 410 financial institutions
  - Top 10: PayPal (1,770), Poste Italiane, Capital One, E*Trade, Chase, Bank of America, UniCredit, Postbank, Scottrade, Wells Fargo
  - 38% of credentials stolen from browser’s password manager

- 1,660 credit cards
  - Top 5: Visa (1,056), Mastercard, American Express, Maestro, Discover
  - US (49%), Italy (12%), Spain (8%)
  - typically, one CC per victim, but there are exceptions …
Value of the Financial Information

- Symantec [2008] estimates
  - Credit card value at $.10 to $25.00
  - Bank account at $10.00 to $1,000.00
- Using Symantec estimates, 10 days of Torpig data valued at $83K to $8.3M
Threats: Denial of Service

- More than 60,000 active hosts at any given time
- Determine network speed from ip2location DB
  - cable and DSL make up 65% of infected hosts
  - used 435 kbps conservative upstream bandwidth
  - yields greater than 17 Gbps just from DSL/cable
    - corporate networks make up 22% of infected hosts

- Potential for a massive DDOS attack
Threats: Proxy Servers

- Torpig opens SOCKS and HTTP proxy
- 20% of infected machines are publicly reachable
- Only 2.45% of those marked by Spamhaus blacklist
- Could be abused for spamming
Threats: Privacy

- Web mail, web chat, and forum messages
- Focused on 6,542 messages in English that were 250 characters or longer
- Zeitgeist of the Torpig network
  - 14% are about jobs/resumes
  - 7% discuss money
  - 6% are sports fans
  - 5% prepare for exams and worry about grades
  - 4% partners/sex online
- Online security is a concern, but users think they are clean
  - 10% specifically mention security/malware
Password Analysis

• 297,962 unique credentials used on 368,501 web sites (domains)
  – mostly web mail (Google, live, Yahoo) and social networking sites (Facebook, MySpace, netlog.com)
  – 28% of the victims reused their password on multiple domains

• Used John the Ripper to assess the strength of the passwords
  – 173,686 unique passwords
  – 56,000 in < 65 minutes using permutation, substitution, etc.
  – 14,000 in next 10 minutes using large wordlist (i.e., 40% cracked in less than 75 minutes)
  – another 30,000 in next 24 hours
Password Analysis

John the Ripper, dictionary with 5800000 entries cracking 173686 unique passwords (HEX, 1 salt)
What about?

- Criminal retribution
- Law enforcement
- Repatriating the data
- Ethics, IRB, etc.
Criminal Retribution

• Big concern on January 25
  – are the criminals going to come to get us?
• More realistically - when will they DDOS our servers?
• Biggest question – why did it take them 10 days to download a new DGA?
Law Enforcement

• We needed to inform law enforcement about this
  – who do we notify?
  – need someone knowledgeable so they don’t shut us down

• How do we get a hold of law enforcement?
  – US CERT gives you a form to fill out
  – contacted David Dagon at Ga Tech and got FBI contact
  – contacted FBI cybercrime unit
  – also contacted DoD defense criminal investigative services

• FBI was very good to work with and gave us lots of contacts for repatriation
Repatriating the Data

- 8,310 accounts from 410 financial institutions
- 1,660 credit cards from various financial institutions
- Need to mine the information from the raw data files
- Cannot just cold call a bank and say I have information that you might want, send me your BINs
- Need introductions from trusted individuals or groups
- FBI and National Cyber-Forensics and Training Alliance (NCFTA) were very helpful
  - leads to individuals who could handle an entire country
Ethics

• Recall Principle 1: the sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized

• Collected sensitive data that potentially could threaten the privacy of victims

• Should emails be viewed at all?

• What about IRB approval?
  – not working with human subjects, why would we need it?
  – we didn’t plan on getting this kind of data
  – any data that can be used to identify an individual needs IRB
Conclusions

- Unique opportunity to understand
  - potential for profit and malicious activity of botnet’s creators
  - characteristics of botnet victims
- Previous evaluations of botnet sizes based on distinct IPs may be grossly overestimated
- Botnet victims are users with poorly maintained machines and choose easily guessable passwords to protect sensitive data
- Interacting with registrars, hosting facilities, victim institutions, and law enforcement can be a complicated process
Credits

• Brett Stone-Gross
• Marco Cova
• Lorenzo Cavallaro
• Bob Gilbert
• Martin Szydlowski
• Richard Kemmerer
• Chris Kruegel
• Giovanni Vigna
Questions?

UC Santa Barbara