Your Botnet is My Botnet: Analysis of a Botnet Takeover

Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, Giovanni Vigna

Presenter: Farhan Jiva

Botnets are a nuisance

- They steal your credentials
- They steal your banking information
- They steal your bandwidth
- They open backdoors on your computer
- They own your system

Related Work

- Your computer is now stoned (...again!)

[http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/your_computer_is_now_stoned.pdf]

  - System level analysis of Mebroot

- Analysis of Sinowal

[http://web17.webbpro.de/index.php?page=analysis-of-sinowal]

  - System level analysis of Torpig

- Kraken Botnet Infiltration

[http://dvlabs.tippingpoint.com/blog/2008/04/28/kraken-botnet-infiltration]

  - Contacted by 65,000 unique IP addresses during infiltration
  - Various size estimates of 185,000 and 600,000 infected hosts

- A Foray into Conficker's Logic and Rendezvous Points

[http://www.usenix.org/event/leet09/tech/full_papers/porras/porras.pdf]

  - Multi-million infected hosts

Let's take a look at one

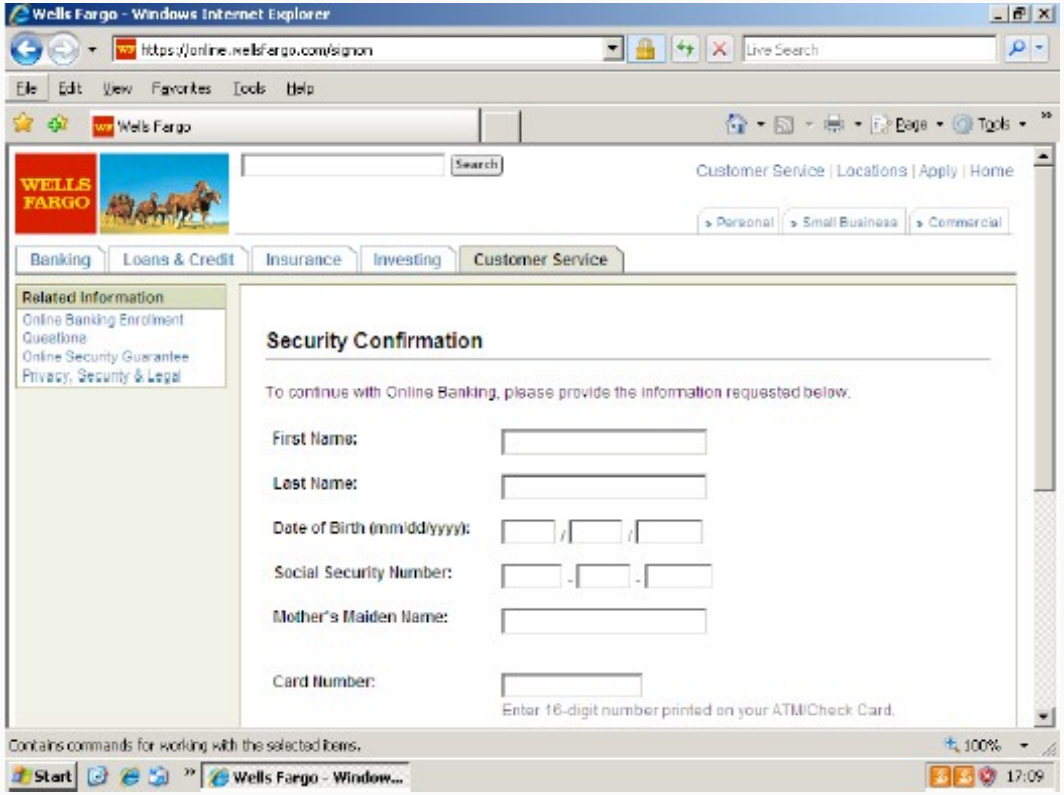
- Torpig aka Sinowal aka Anserin
  - "One of the advanced pieces of crimeware ever created"

- Piggybacks on Mebroot

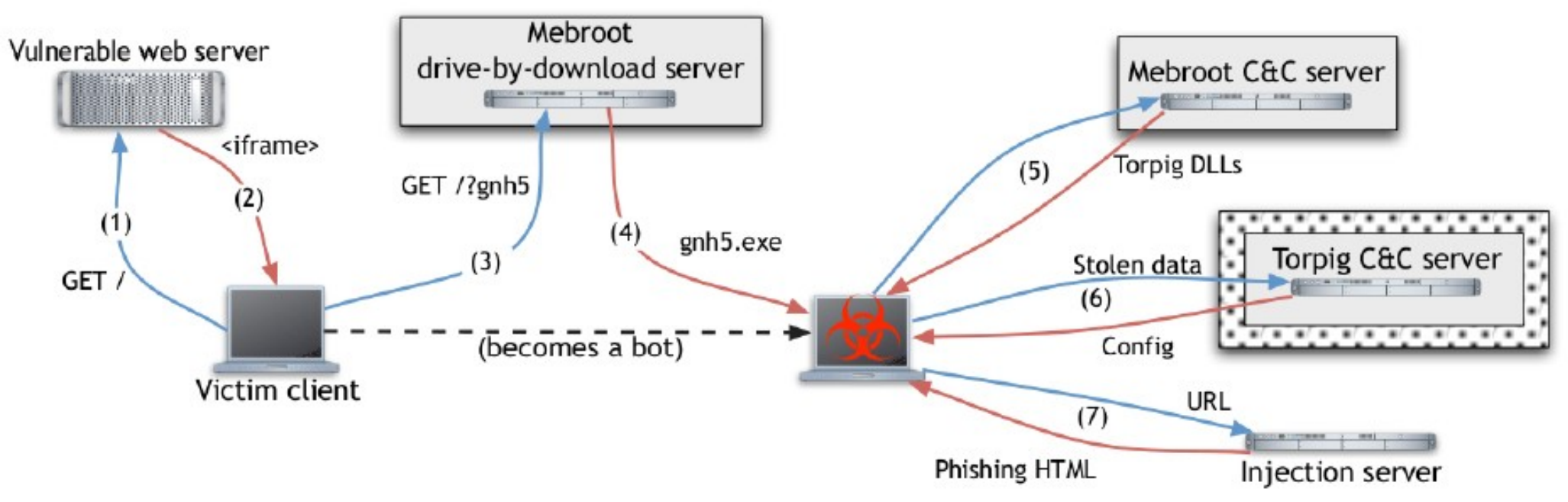
- Uses Mebroot-injected dlls to phish you

- Estimated to contain over 180 thousand infections

- Steals a **LOT** of your information



Torpig's Infection Vector



Get your very own infection in a just <u>few easy steps</u>!

1: Access a vulnerable web site

2,3,4: Drive-by-download to get Mebroot

5: Get Torpig via Mebroot

6: Send Torpig your personal data

7: Get special HTML for phishing purposes

8: Enjoy getting your identity back!

What is domain flux?

•A bot must keep in contact with the botmaster to be useful

•A botmaster must be coordinated with his bots to be efficient

Flux Capacitor

•Hard coding domains or IPs in the bots are a bad idea
  - What if they are taken down?

•Domain flux: Have the bots use an algorithm to generate domains to use on a daily/weekly basis
  - If the domains go bad? No problem, move on to the next one

Torpig's Domain Generation Algorithm

- Seeded by current date
- Weekly domains
  - Torpig will generate some string *dw* to use for the week
  - dw.com, dw.net, dw.biz
- If weekly domains fail, torpig will generate a daily domain *dd*
  - dd.com, dd.net, dd.biz
- If daily domains fail, torpig will use a hard-coded domain from the latest configuration file received from the C&C

```
suffix =  ["anj", "ebf", "arm", "pra", "aym", "unj",
    "ulj", "uag", "esp", "kot", "onv", "edc"]

def generate_daily_domain():
    t = GetLocalTime()
    p = 8
    return generate_domain(t, p)

def scramble_date(t, p):
    return (((t.month ^ t.day) + t.day) * p) +
        t.day + t.year

def generate_domain(t, p):
    if t.year < 2007:
        t.year = 2007
    s = scramble_date(t, p)
    c1 = (((t.year >> 2) & 0x3fc0) + s) % 25 + 'a'
    c2 = (t.month + s) % 10 + 'a'
    c3 = ((t.year & 0xff) + s) % 25 + 'a'
    if t.day * 2 < '0' ||  t.day * 2 > '9':
        c4 = (t.day * 2) % 25 + 'a'
    else:
        c4 = t.day % 10 + '1'
    return c1 + 'h' + c2 + c3 + 'x' + c4 +
        suffix[t.month - 1]
```

Time to steal a botnet

- Researchers bought 2 domains and some hosting

- During the ten days they had control
  - captured 69GBs of pcap data
  - collected 8.7GBs of Apache log data
- Data was encrypted using 256-bit AES
- Some data collection principles were set prior to gathering data
  - Don't intentionally cause damage to the hosts on the botnet
  - Collect enough information to notify and remediate those who are affected by the data gathered

Format of Torpig's data transmission

```
POST /A15078D49EBA4C4E/qxoT4B5uUFFqw6c35AKDYFpdZHdKLCNn...AaVpJGoSZG1at6E0AaCxQg6nIGA

ts=1232724990&ip=192.168.0.1:&sport=8109&hport=8108&os=5.1.2600&cn=United%20States&
nid=A15078D49EBA4C4E&bld=gnh5&ver=229
```

```
[gnh5_229]
[MSO2002-MSO2003:pop.smith.com:John Smith: john@smith.com]
[pop3://john:smith@pop.smith.com:110]
[smtp://:@smtp.smith.com:25]
```

```
[gnh5_229]
POST /accounts/LoginAuth
Host: www.google.com
POST_FORM:
Email=test@gmail.com
Passwd=test
```

- Torpig communicates via HTTP POST
  - The URL contains a bot identifier and a submission header
  - The body of the POST request contains the stolen data
  - Both are encrypted using base64 and XOR (with the key sent as plaintext)
- Submission header contains identifying information about the specific bot
  - Timestamp, IP, sport (SOCKS proxy), hport (HTTP proxy), operating system, country name, node-id, build (customer purchasing stolen data?), version
- Torpig steals your email client's credentials, email address list, form data you submit on webpages, your windows password and more

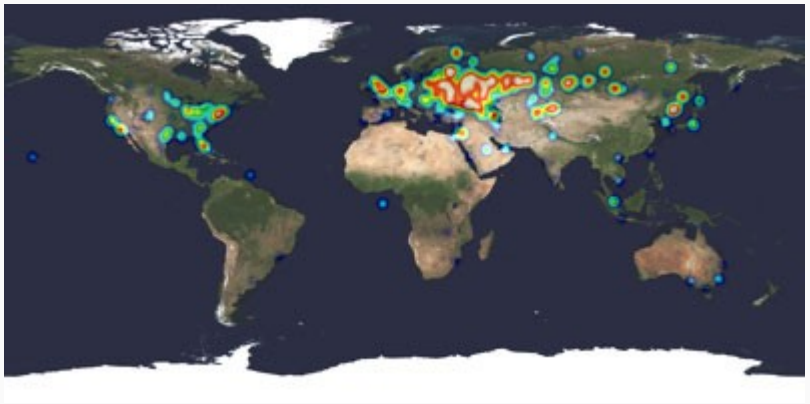
Mailbox account	54,090
Email	1,258,862
Form data	11,966,532
HTTP account	411,039
FTP account	12,307
POP account	415,206
SMTP account	100,472
Windows password	1,235,122

Problem of botnet sizing

- Sizing botnets is a difficult task
- There tends to be many disagreements regarding the sizes of some botnets
- Why not just count the number of unique IP addresses?
  - Many computers are behind a NAT
  - DHCP might assign you a new IP when your lease is up

Sizing Torpig

- Use some values in the submission header to determine Torpig's footprint
- *nid* is a value based on your hard drive's serial number
  - Appears to be unique, however there were around 2,000 *nid*s shown to have the same value
- Found that the tuple (nid, os, cn, bld, ver) remained unique
- 182,800 unique tuples identified with the server
- 1,247,642 unique IPs identified with the server
  - Assuming #unique IPs == #unique bots would have been a gross overestimation



New IPs and Bots per hour

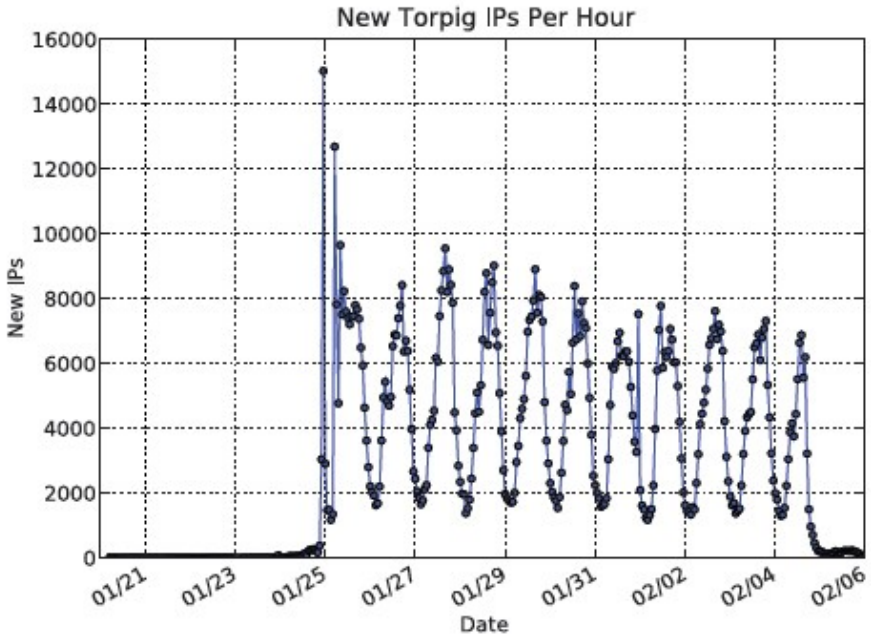


Figure 5: New unique IP addresses per hour.

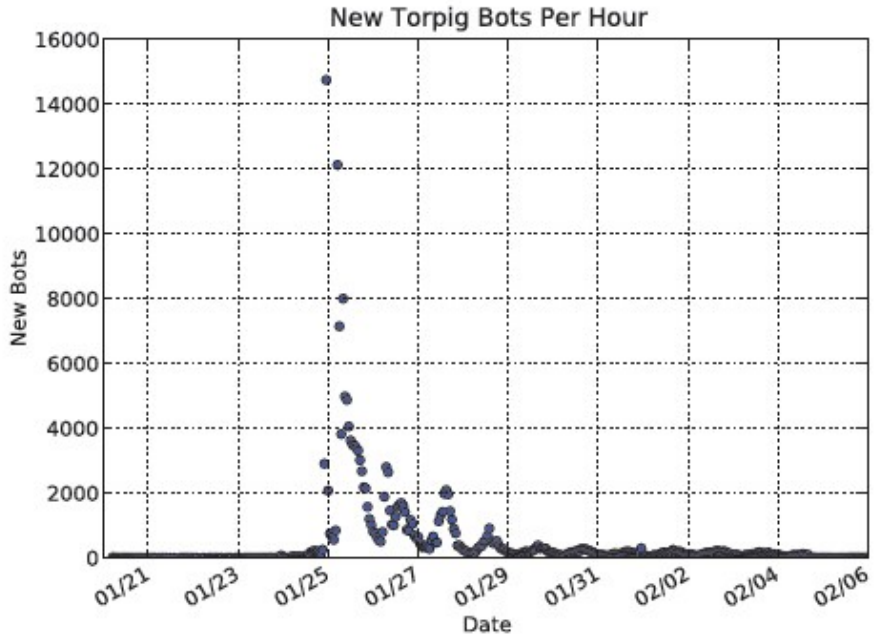


Figure 6: New bots per hour.

- After initial spike, consistent diurnal pattern
- Averaging 4,690 new IPs per hour

- After initial spike, rapid drop-off
- Averaging 705 new bots per hour

Cumulative IPs and Bots per hour

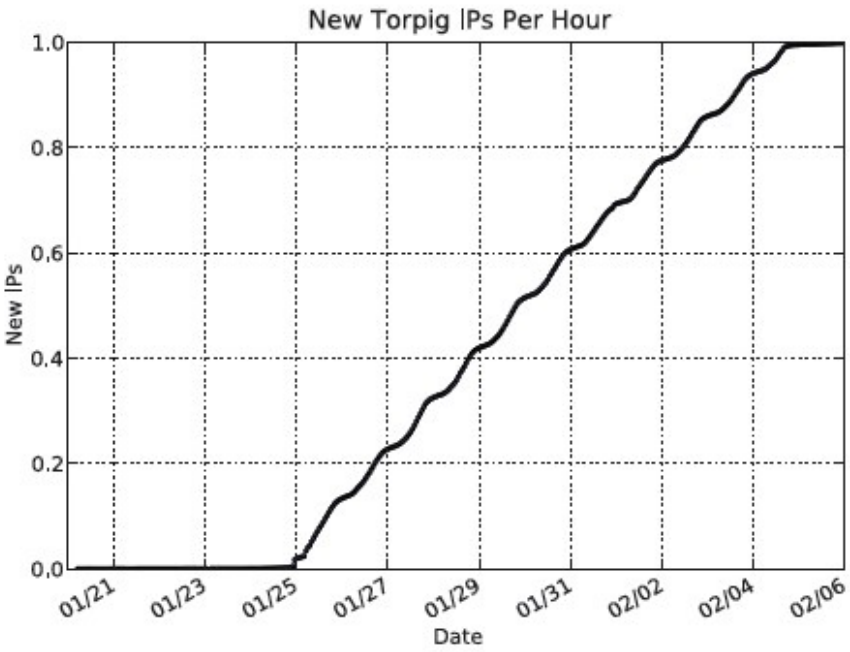


Figure 7: CDF – New unique IP addresses per hour.

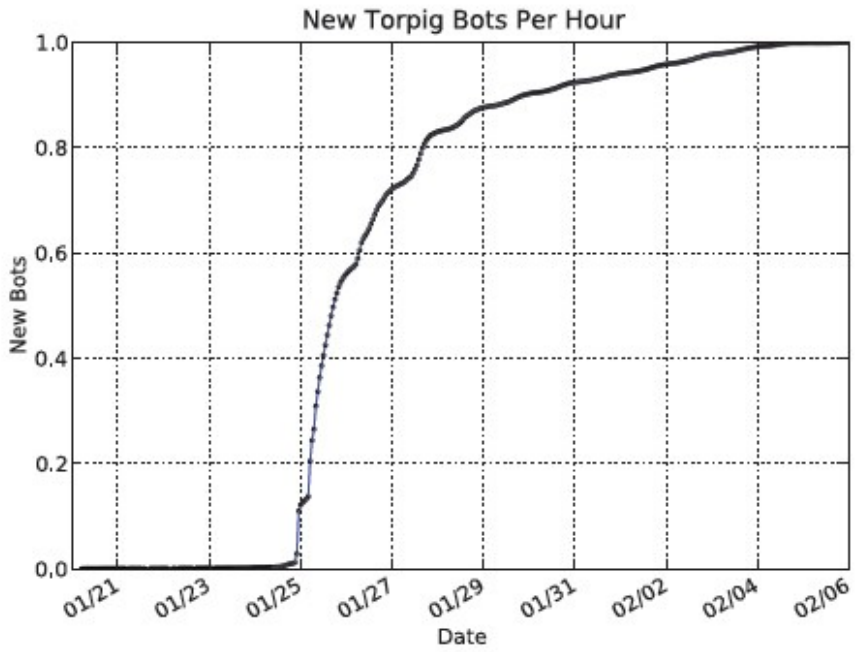


Figure 8: CDF – New bots per hour.

- Number of cumulative new IPs increased linearly

- Number of cumulative bots decayed quickly
- More than 75% of all new bots during the ten day study were observed in the first 48 hours

Using IP addresses to size Torpig

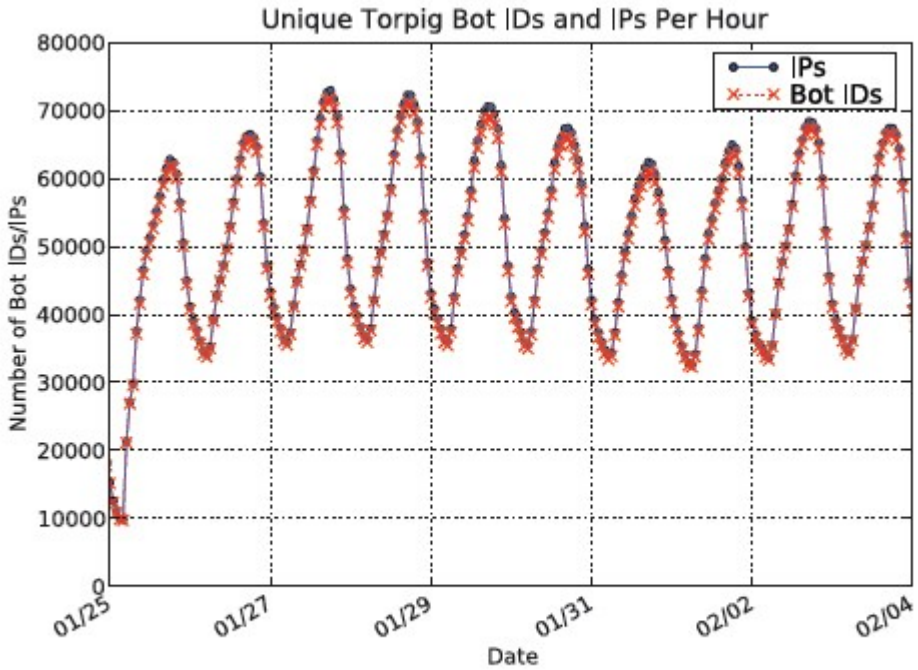


Figure 9: Unique Bot IDs and IP addresses per hour.

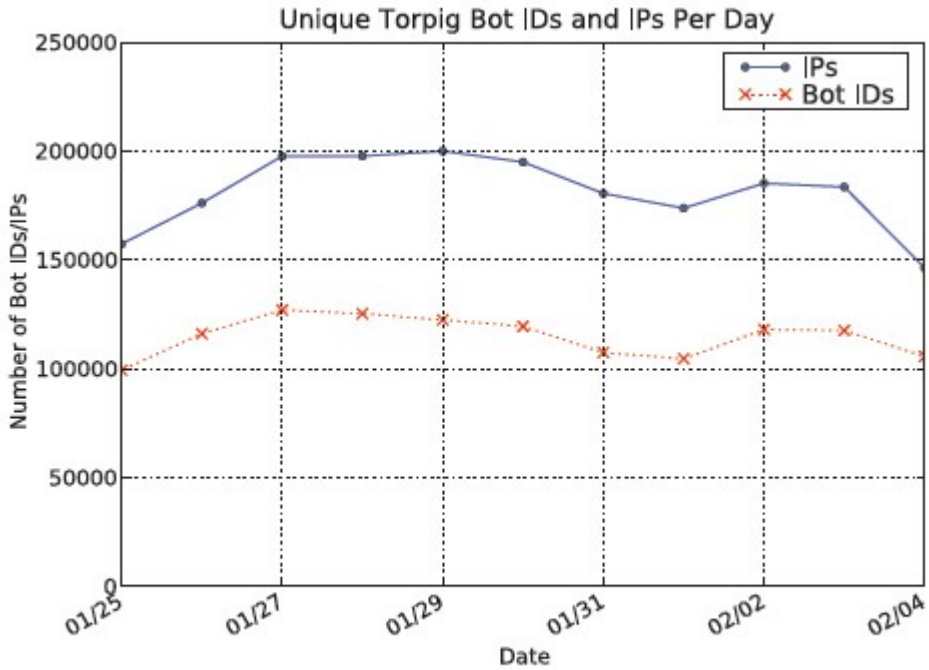


Figure 10: Unique Bot IDs and IP addresses per day.

- Number of unique bot IDs per hour and number of unique IPs per hour are nearly identical
- Number of unique bot IDs per day does not reflect the number of unique IPs per day

This difference is a consequence of the bots contacting the C&C every 20 minutes, which occurs more frequently than the rate of DHCP churn

Observing DHCP Churn

Country	IP Addresses (Raw #)	Bot IDs	DHCP Churn Factor
US	158,209	54,627	2.90
IT	383,077	46,508	8.24
DE	325,816	24,413	13.35
PL	44,117	6,365	6.93
ES	31,745	5,733	5.54
GR	45,809	5,402	8.48
CH	30,706	4,826	6.36
UK	21,465	4,792	4.48
BG	11,240	3,037	3.70
NL	4,073	2,331	1.75
Other	180,070	24,766	7.27
Totals:	1,247,642	182,800	6.83

Table 2: Top 10 infected hosts by country.

- DHCP allocation is dynamic
  - Not guaranteed to receive the same IP each time you connect
- DHCP Churn Factor: about how many different IPs each host received throughout the 10 day study
- In one instance, a single host changed its IP address 694 time in a ten day period

New Torpig infections over time

- Recall that the submission header contained a timestamp field
  - Timestamp of the most recently received configuration file from C&C
- By counting the number of bots who had timestamp == 0, can determine new infections
- 49,294 new infections while the botnet was under the control of the researchers

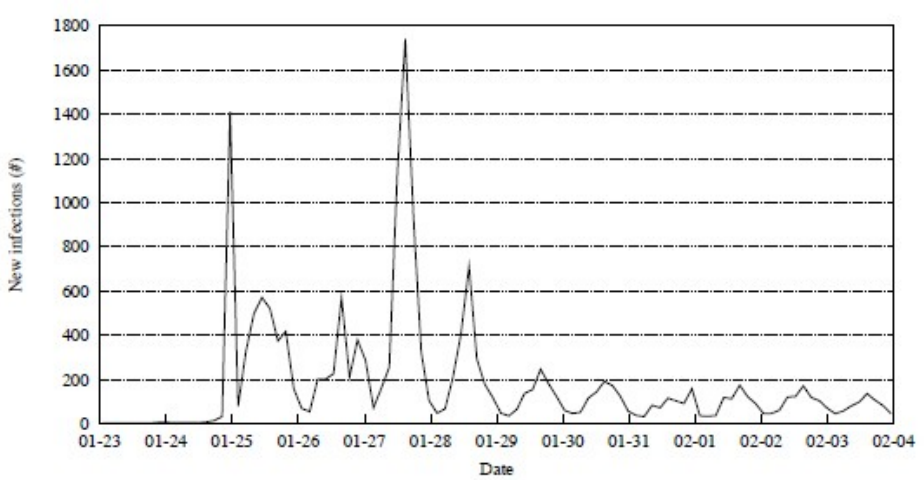


Figure 11: New infections over time.

Botnets as a service

- Recall that in the submission header, there was a *build* field
- The researchers believe this field corresponded to a “customer” id
  - Each customer would receive the data stolen which contained their customer id
- 12 different values for build
  - dxtrbc, eagle, gnh1, gnh2, gnh3, gnh4, gnh5, grey, grobin, grobin1, mentat, zipp

Stealing Financial Data

- In just ten days, Torpig stole 8,310 accounts from 410 institutions
  - Paypal: 1,770 accounts
  - Poste Italiane: 765 accounts
  - Capital One: 314 accounts
  - E*Trade: 304 accounts
  - Chase: 217 accounts
  - ...

Country	Institutions (#)	Accounts (#)
US	60	4,287
IT	34	1,459
DE	122	641
ES	18	228
PL	14	102
Other	162	1,593
Total	410	8,310

Table 3: Accounts at financial institutions stolen by Torpig.

How much money are we talking about?

- 1,600 unique credit and debit card numbers were obtained
- Quantifying the value of financial information is difficult
- The researchers estimated that the botmasters profit anywhere from $83K to $8.3M in the span of ten days

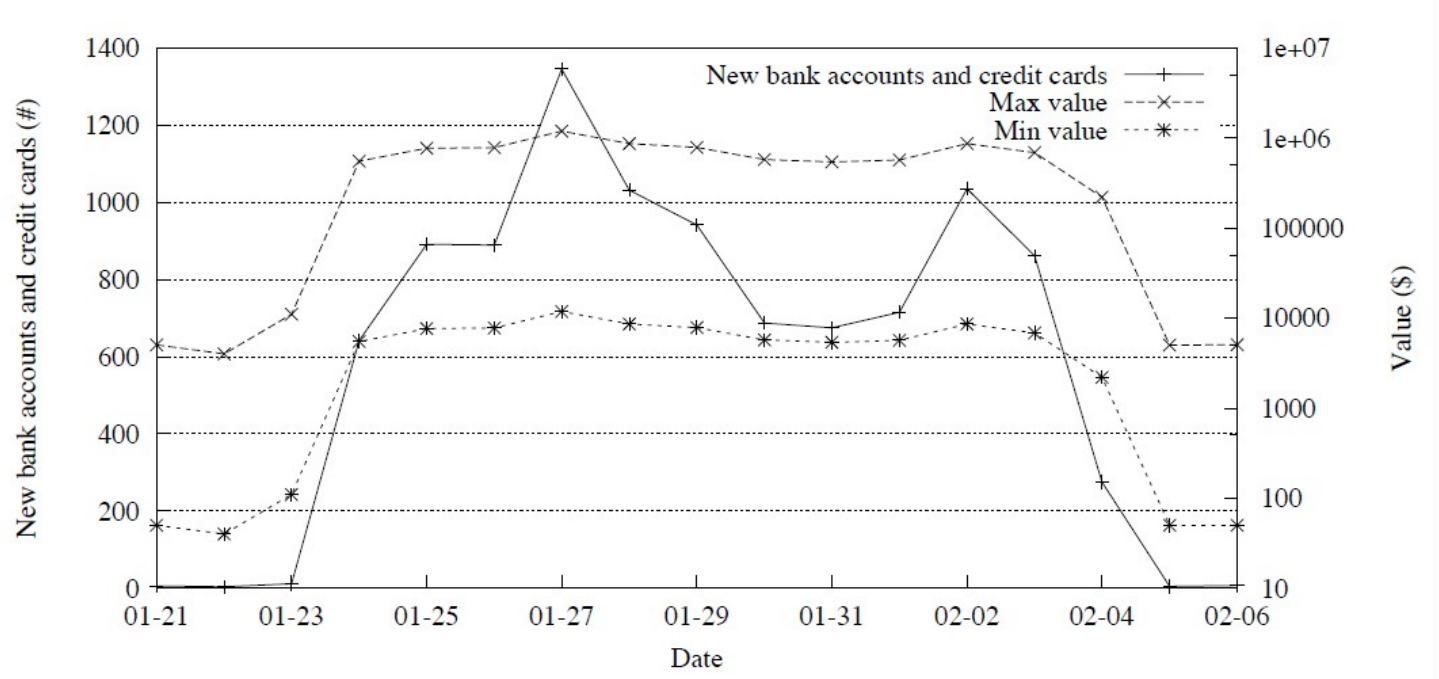
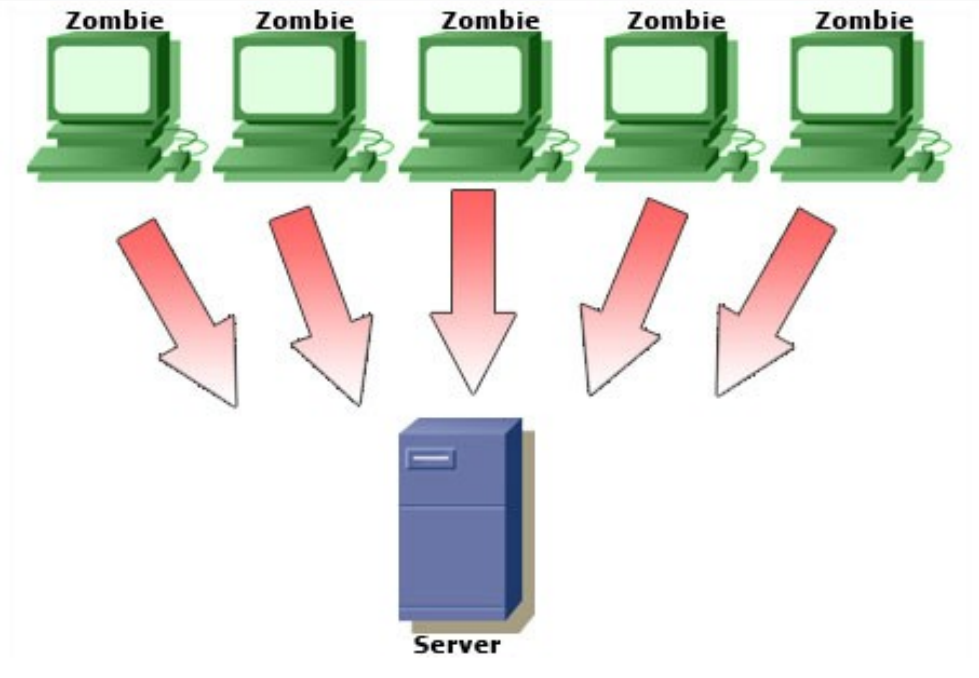


Figure 12: The arrival rate of financial data.

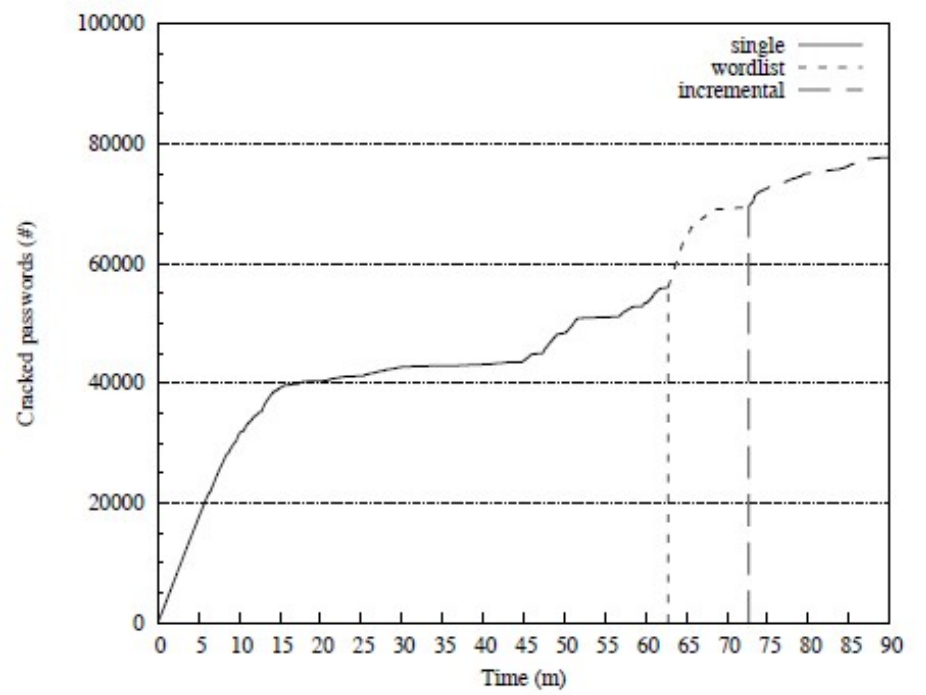
Potential for Distributed Denial-of-Service

- During peak intervals, there were around 70,000 live hosts on Torpig
- Conservative estimate of 435 kbps upstream bandwidth for each host
- Roughly 17 Gbps of bandwidth available to the botmasters.



Password Analysis

- Torpig stole 297,962 unique username/password pairs
- Researchers found that 28% of victims reused credentials for 368,501 web sites
- Strength test:
  - Created a UNIX-like password file using the unique passwords (about 174,000 of them)
  - Fed into John the Ripper
  - Cracked around 100,000 passwords in 24 hours



Questions?