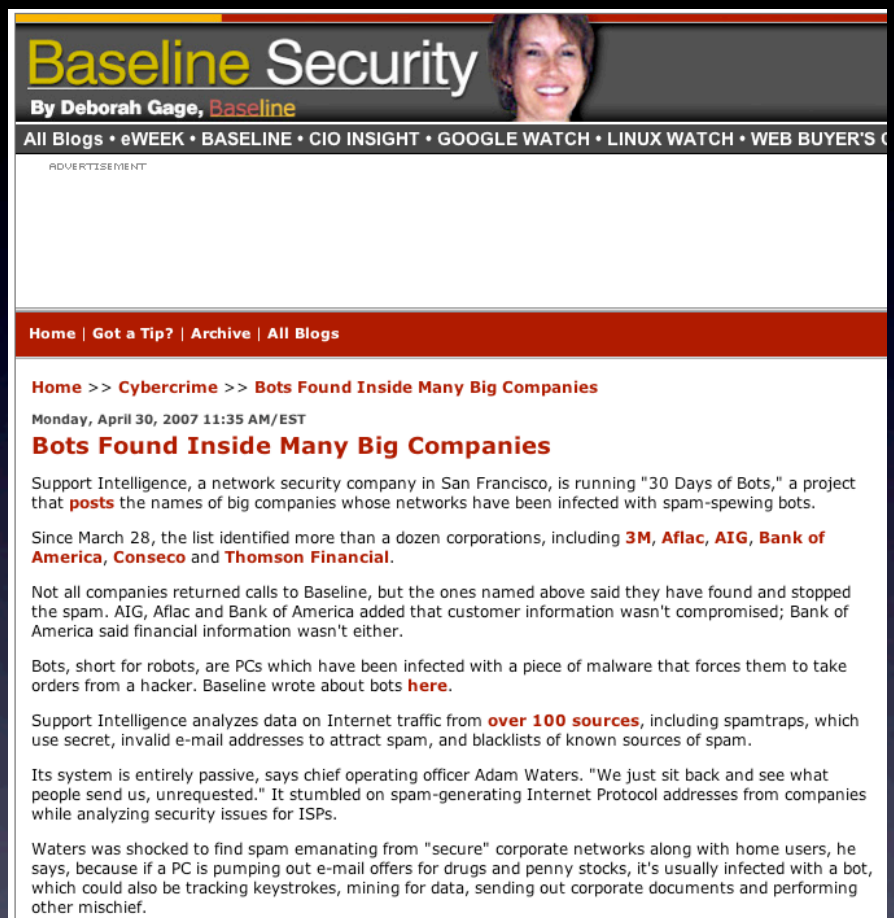
Studying Botnets Using *BotLab*

Arvind Krishnamurthy

Botnets: a Growing Threat



Baseline Security

By Deborah Gage, Baseline

All Blogs • eWEEK • BASELINE • CIO INSIGHT • GOOGLE WATCH • LINUX WATCH • WEB BUYER'S

ADVERTISEMENT

Home | Got a Tip? | Archive | All Blogs

Home >> Cybercrime >> Bots Found Inside Many Big Companies

Monday, April 30, 2007 11:35 AM/EST

Bots Found Inside Many Big Companies

Support Intelligence, a network security company in San Francisco, is running "30 Days of Bots," a project that **posts** the names of big companies whose networks have been infected with spam-spewing bots.

Since March 28, the list identified more than a dozen corporations, including **3M**, **Aflac**, **AIG**, **Bank of America**, **Conseco** and **Thomson Financial**.

Not all companies returned calls to Baseline, but the ones named above said they have found and stopped the spam. AIG, Aflac and Bank of America added that customer information wasn't compromised; Bank of America said financial information wasn't either.

Bots, short for robots, are PCs which have been infected with a piece of malware that forces them to take orders from a hacker. Baseline wrote about bots **here**.

Support Intelligence analyzes data on Internet traffic from **over 100 sources**, including spamtraps, which use secret, invalid e-mail addresses to attract spam, and blacklists of known sources of spam.

Its system is entirely passive, says chief operating officer Adam Waters. "We just sit back and see what people send us, unrequested." It stumbled on spam-generating Internet Protocol addresses from companies while analyzing security issues for ISPs.

Waters was shocked to find spam emanating from "secure" corporate networks along with home users, he says, because if a PC is pumping out e-mail offers for drugs and penny stocks, it's usually infected with a bot, which could also be tracking keystrokes, mining for data, sending out corporate documents and performing other mischief.

Botnets: a Growing Threat

Baseline Security

By Deborah Gage, Baseline

All Blogs • eWEEK • BASELINE • CIO INSIGHT • GOOGLE WATCH • LINUX WATCH •

ADVERTISEMENT

Home | Got a Tip? | Archive | All Blogs

Home >> Cybercrime >> Bots Found Inside Many Big Companies

Monday, April 30, 2007 11:35 AM/EST

Bots Found Inside Many Big Companies

Support Intelligence, a network security company in San Francisco, is running "30 Days of B
that posts the names of big companies whose networks have been infected with spam-spew

Since March 28, the list identified more than a dozen corporations, including 3M, Aflac, AIG
America, Conseco and Thomson Financial.

Not all companies returned calls to Baseline, but the ones named above said they have found
the spam. AIG, Aflac and Bank of America added that customer information wasn't comprom
America said financial information wasn't either.

Bots, short for robots, are PCs which have been infected with a piece of malware that forces t
orders from a hacker. Baseline wrote about bots here.

Support Intelligence analyzes data on Internet traffic from over 100 sources, including spa
use secret, invalid e-mail addresses to attract spam, and blacklists of known sources of spam

Its system is entirely passive, says chief operating officer Adam Waters. "We just sit back and
people send us, unrequested." It stumbled on spam-generating Internet Protocol addresses f
while analyzing security issues for ISPs.

Waters was shocked to find spam emanating from "secure" corporate networks along with h
says, because if a PC is pumping out e-mail offers for drugs and penny stocks, it's usually infe
which could also be tracking keystrokes, mining for data, sending out corporate documents a
other mischief.

bbc.co.uk Home | TV | Radio | Talk | Where I Live | A-Z Index | Search

UK version | International version | About the versions | Low graphics | Accessibility help

BBC NEWS | WATCH LIVE BBC News 24 | News services Your news when you want it

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science/Nature
Technology
Entertainment

Video and Audio

Have Your Say
Magazine
In Pictures
Country Profiles
Special Reports

RELATED BBC SITES
SPORT
WEATHER
CBBC NEWSROUND
ON THIS DAY
EDITORS' BLOG

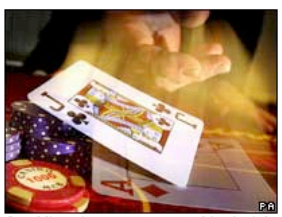
Last Updated: Friday, 4 May 2007, 11:57 GMT 12:57 UK

E-mail this to a friend | Printable version

Firms hit rivals with web attacks

By Mark Ward
Technology Correspondent, BBC News website

Legitimate businesses are turning to cyber criminals to help them cripple rival websites, say security experts.



Gambling sites were targeted in denial of service attacks

The rise in industrial sabotage comes as some suggest cyber criminals are turning away from using web-based attack tools in extortion rackets.

Experts suspect this is because of the risks involved in mounting such an attack on a web shop or retailer.

Instead the tools, usually hijacked home computers, are being used to pump out junk e-mail.

Cash call

Often these hijacked PCs, known as bots, are used for "Distributed Denial of Service" (DDoS) attacks that attempt to knock a site or server offline by bombarding it with huge amounts of data.

Online gambling sites were among the first to be threatened with DDoS attacks if they did not hand over significant sums of cash.

SEE ALSO
- Hi-tech crime: A glossary 05 Oct 06 | UK
- Caught in the net 05 Oct 06 | Technology
- Online service foils ransom plot 31 May 05 | Technology
- Rings of steel combat net attacks 13 Jan 05 | Technology
- Blackmailers target $1m website 18 Jan 06 | Technology
- Bookies suffer online onslaught 19 Mar 04 | Technology
- Bookies race to beat net attacks 02 Apr 04 | Technology

RELATED INTERNET LINKS
- Symantec
- Symantec blog entry on extortion via DDoS
- Prolexic

The BBC is not responsible for the content of external internet sites

TOP TECHNOLOGY STORIES
- Google searches web's dark side
- Bandwidth leap for British forces
- Children warned on web safety

News feeds

MOST POPULAR STORIES NOW

MOST E-MAILED | MOST READ

1 10m tuned in for Eurovision

Botnets: a Growing Threat

Baseline Security

By Deborah Gage, Baseline

All Blogs • eWEEK • BASELINE • CIO INSIGHT • GOOGLE WATCH • LINUX WATCH •

ADVERTISEMENT

Home | Got a Tip? | Archive | All Blogs

Home >> Cybercrime >> Bots Found Inside Many Big Companies

Monday, April 30, 2007 11:35 AM/EST

Bots Found Inside Many Big Companies

Support Intelligence, a network security company in San Francisco, is running "30 Days of B
that posts the names of big companies whose networks have been infected with spam-spew

Since March 28, the list identified more than a dozen corporations, including 3M, Aflac, AIG
America, Conseco and Thomson Financial.

Not all companies returned calls to Baseline, but the ones named above said they have found
the spam. AIG, Aflac and Bank of America added that customer information wasn't comprom

bbc.co.uk

Home | TV | Radio | Talk | Where I Live | A-Z Index | Search

UK version | International version | About the versions | Low graphics | Accessibility help

BBC NEWS

WATCH LIVE BBC News 24

News services
Your news when you want it

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science/Nature
Technology
Entertainment

Video and Audio

Have Your Say
Magazine
In Pictures

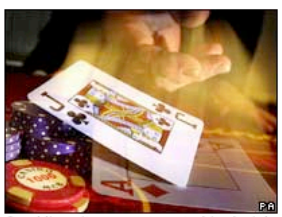
Last Updated: Friday, 4 May 2007, 11:57 GMT 12:57 UK

E-mail this to a friend | Printable version

Firms hit rivals with web attacks

By Mark Ward
Technology Correspondent, BBC News website

Legitimate businesses are turning to cyber criminals to help them cripple rival websites, say security experts.



Gambling sites were targeted in denial of service attacks

The rise in industrial sabotage comes as some suggest cyber criminals are turning away from using web-based attack tools in extortion rackets.

Experts suspect this is because of the risks involved in mounting such an attack on a web shop or retailer.

Instead the tools, usually hijacked home computers, are being used to pump out junk e-mail.

Cash call

Often these hijacked PCs, known as bots, are used for "Distributed Denial of Service" (DDoS) attacks that attempt to knock a site or server offline by bombarding it with huge amounts of data.

Online gambling sites were among the first to be threatened with DDoS attacks if they did not hand over significant sums of cash.

SEE ALSO
- Hi-tech crime: A glossary 05 Oct 06 | UK
- Caught in the net 05 Oct 06 | Technology
- Online service foils ransom plot 31 May 05 | Technology
- Rings of steel combat net attacks 13 Jan 05 | Technology
- Blackmailers target $1m website 18 Jan 06 | Technology
- Bookies suffer online onslaught 19 Mar 04 | Technology
- Bookies race to beat net attacks 02 Apr 04 | Technology

RELATED INTERNET LINKS
- Symantec
- Symantec blog entry on extortion via DDoS
- Prolexic

The BBC is not responsible for the content of external internet sites

TOP TECHNOLOGY STORIES
- Google searches web's dark side
- Bandwidth leap for British forces
- Children warned on web safety

News feeds

MOST POPULAR STORIES NOW

MOST E-MAILED | MOST READ

1 10m tuned in for Eurovision

NETWORKWORLD

Search / DocFinder

HOME
RESEARCH CENTERS
Security
Anti-Virus
Firewalls / VPN / Intrusion
Spam / Phishing
Wireless Security
+ LANs & Routers
+ VoIP & Convergence
+ Network Management
+ Wireless & Mobile
+ Operating Systems
+ Servers & Data Center
+ Applications
+ Storage
+ Wide Area Network
+ Small Business Networking

Security

NAC Cram Session | Anti-Virus | Firewalls / VPN / Intrusion | Spam / Phishing | Wirele

NetworkWorld.com > Security >

Antispam firm says it was victim of attack

By Jaikumar Vijayan, Computerworld, 05/05/06

The CEO of an antispam firm whose service was knocked offline by a spammer earlier this week claimed his company was the victim of a sophisticated attack carried out, in part, with the help of someone at a top-tier ISP.

Eran Reshef, CEO of Blue Security, an Israeli antispam firm, said that his company was attacked by a major spammer named PharmaMaster who used a combination of methods to knock

Other stories on this topic

Hackers hijack Windows Update's downloader
5/11/2007

Nevis announces free LAN

Botnets: a Growing Threat

What do Bots do?

- Steal personal information, install keyloggers
- Participate in "distributed denial of service" attacks
- Send spam
- Infect other machines
- Perform click fraud
- ...

Botnets still a mystery...

- Increasing awareness, but there is a dearth of hard facts especially in real-time
  - Meager network-wide cumulative statistics
  - Sparse information regarding individual botnets
  - Most analysis is post-hoc

Inconsistent Information

Big Honkin' Botnet - 1.5 Million!

Published: 2005-10-20,
Last Updated: 2005-10-20 21:13:23 UTC
by Ed Skoudis (Version: 1)

0 comment(s) Digg submit

A diligent reader from the Netherlands requesting anonymity (lots of folks doing that today) pointed this article about a recent botnet bust in the Netherlands. The article is in Dutch, but our reader tra it thusly:

"The botnet in the spotlight by the Dutch National Criminal Investigation unit in the Netherlands, abo weeks ago was found to comprise approximately 1.5 million hacked computers (instead of 100k repo earlier) . This has been discovered by GovCert.nl, the Dutch Computer Emercency Response Team, dismantling the network of computers infected with a Trojan Horse. Of the total number of infected computers, it was estimated that only 30,000 were located in the Netherlands.

Inconsistent Information

Big Honkin' Botnet - 1.5 Million!

Published: 2005-10-20,
Last Updated: 2005-10-20 21:13:23 UTC
by Ed Skoudis (Version: 1)

Botnet scams are exploding

Updated 17d ago | Comments 92 | Recommend 37 E-mail | Save | Print | Reprints & Permissions | RSS

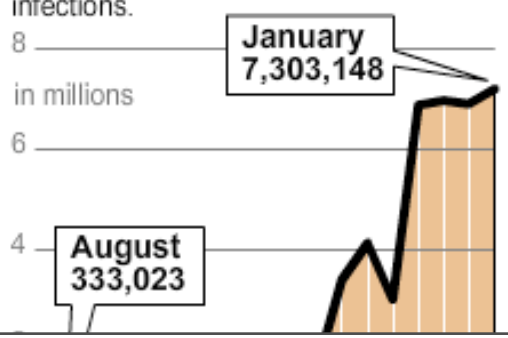
CYBERCRIME PAYS

The escalating number of botnets have helped feed a surge in various forms of online fraud.

Botnet deluge Virus rate
E-mail spam Phishing attacks

Botnet deluge

The average daily number of unique botnet communiqués to accept instructions from a controller, deliver spam, conduct phishing campaigns, click on ads to earn ad revenue, carry out denial-of-service attacks, steal data, scan for vulnerable computers, and spread infections.



By Byron Acohido and Jon Swartz, USA TODAY

Mixx it
Other ways to share:
Yahoo! Buzz
Digg
Newsvine
Reddit
Facebook
What's this?

SEATTLE — Two days after actor Heath Ledger died, e-mails began moving across the Internet purportedly carrying a link to a detailed police report divulging "the real reason" behind the actor's death. Ledger had been summarily drafted into the service of a botnet.

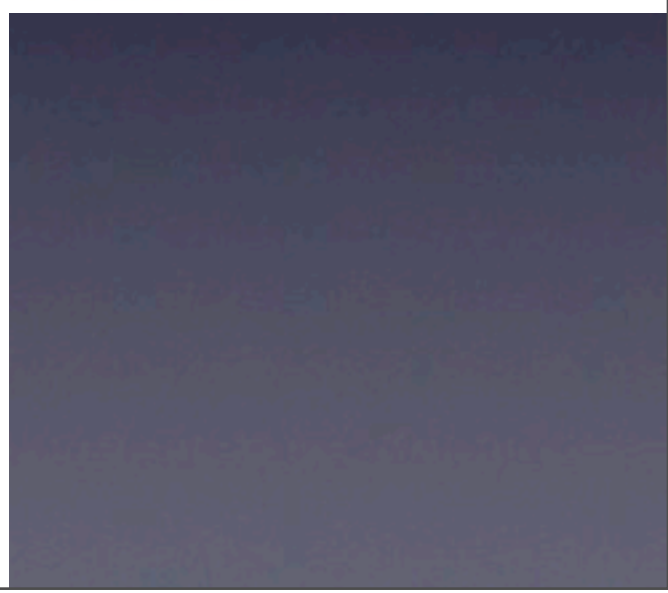
Bots are compromised computers controlled by profit-minded crooks. Those e-mails were spread by a network of thousands of bots, called a botnet. Anyone who clicked on the link got instantly absorbed into the fast-spreading Mega-D botnet, says security firm Marshal. Mega-D enriches its operators, mainly by distributing spam for male-enhancement pills.

BACKGROUND: Botnets can be used to blackmail targeted sites

Largely unnoticed by the [illegible] the Internet. On a typic[illegible] day, 40% of the 800 million computers connected to the Internet [illegible] spam, stealing sensitive data typed at banking and shopping websites, bombarding websites as part of extortionist denial-

ts of folks doing that today) pointed
rticle is in Dutch, but our reader tra

stigation unit in the Netherlands, abo
cked computers (instead of 100k repo
mputer Emercency Response Team,
rse. Of the total number of infected
e Netherlands.



Inconsistent Information

Big Honkin' Botnet - 1.5 Million!

Published: 2005-10-20,
Last Updated: 2005-10-20 2
by Ed Skoudis (Version: 1)

Botnet scams are explodi

Updated 17d ago | Comments 92 | Recommend 37

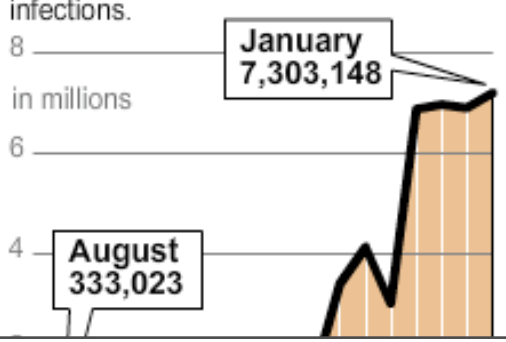
CYBERCRIME PAYS

The escalating number of botnets have helped feed a surge in various forms of online fraud.

Botnet deluge Virus rate
E-mail spam Phishing attacks

Botnet deluge

The average daily number of unique botnet communiqués to accept instructions from a controller, deliver spam, conduct phishing campaigns, click on ads to earn ad revenue, carry out denial-of-service attacks, steal data, scan for vulnerable computers, and spread infections.



By Byron Ac
TODAY

SEATTLE —
Ledger died,
the Internet p
detailed polic
reason" behir
been summar
botnet.

Bots are com
by profit-mind
network of th
clicked on the
Mega-D botn
operators, ma
pills.

BACKGROU
sites

Largely unnot
the Internet.
connected to
spam, stealin
websites, bon

ars technica
the art of technology

Main | Business IT | Apple | Gam

New face. Same Ars.
Learn more about how we're making Ars Technica even better than before.

Home | News | Articles | Guides | Journals | Search | GO

From the News Desk

Vint Cerf: one quarter of all computers part of a botnet

By Nate Anderson | Published: January 25, 2007 - 04:35PM CT

The World Economic Forum takes place this week in Davos, Switzerland, and leaders around the world gather to discuss issues like the Iraq war, global climate change, and globalization—along with the incredible prevalence of botnets.

The BBC's Tim Weber, who was in the audience of an Internet panel featuring Vint Cerf, Michael Dell, John Markoff of the *New York Times*, and Jon Zittrain of Oxford, came away most impressed by the botnet statistics. Cerf told his listeners that approximately 600 million computers are connected to the Internet, and that 150 million of them might be participants in a botnet—nearly all of them unwilling victims. Weber remarks that "in most cases the owners of these computers have not the slightest idea what their little beige friend in the study is up to."

If Cerf's estimate is accurate, that's one quarter of all machines connected to the Internet. So is the Internet doomed? Well, you're reading this, so no, not yet. But the botnet menace is no phantom, and it has been growing in strength for years. In September 2006, security research

Research Agenda

To build a *botnet monitoring platform* that can track the activities of the *most significant spamming botnets* currently operating in *real-time*

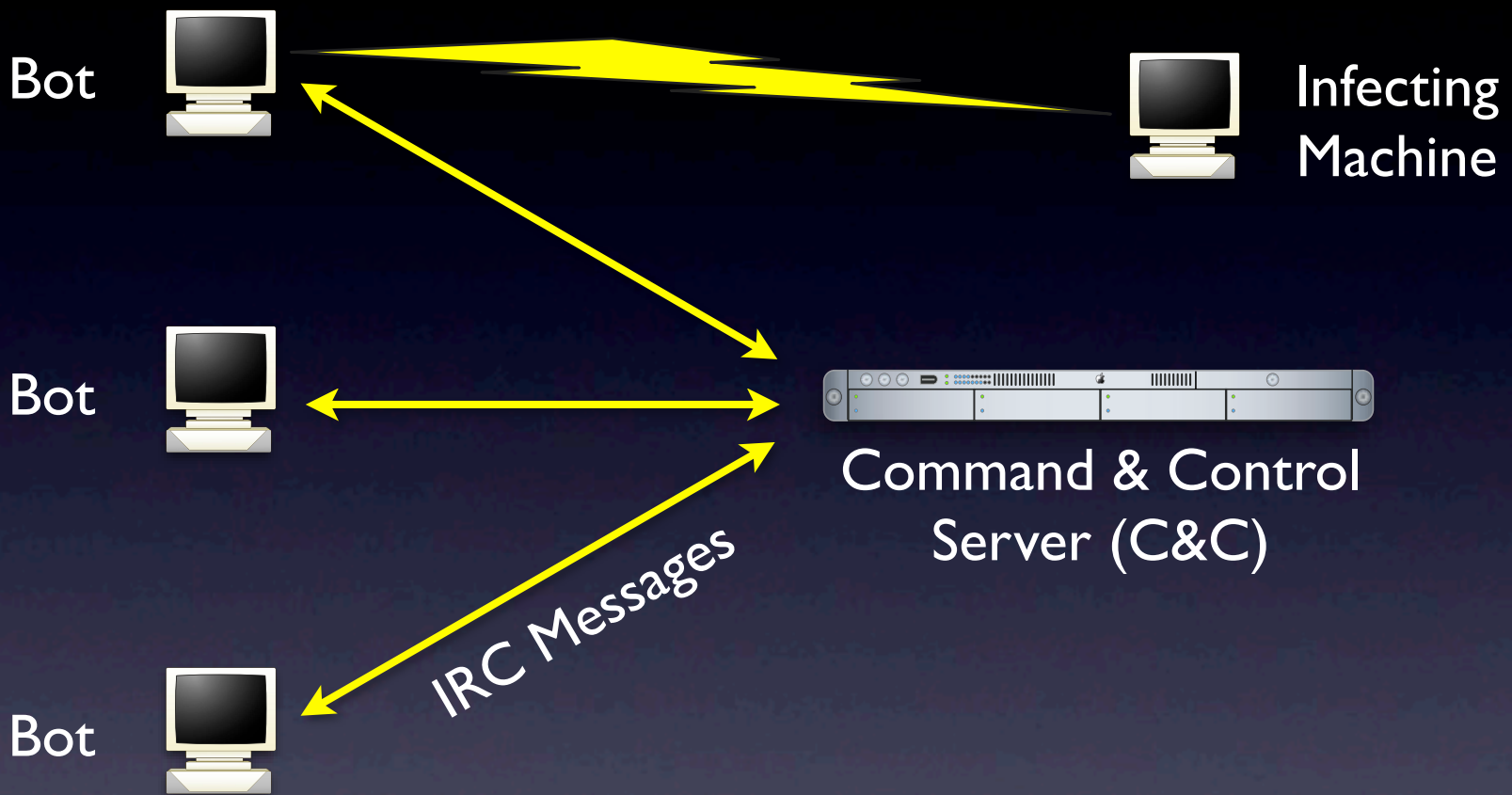
Botnet Lifecycle (Traditional View)

Infecting Machine

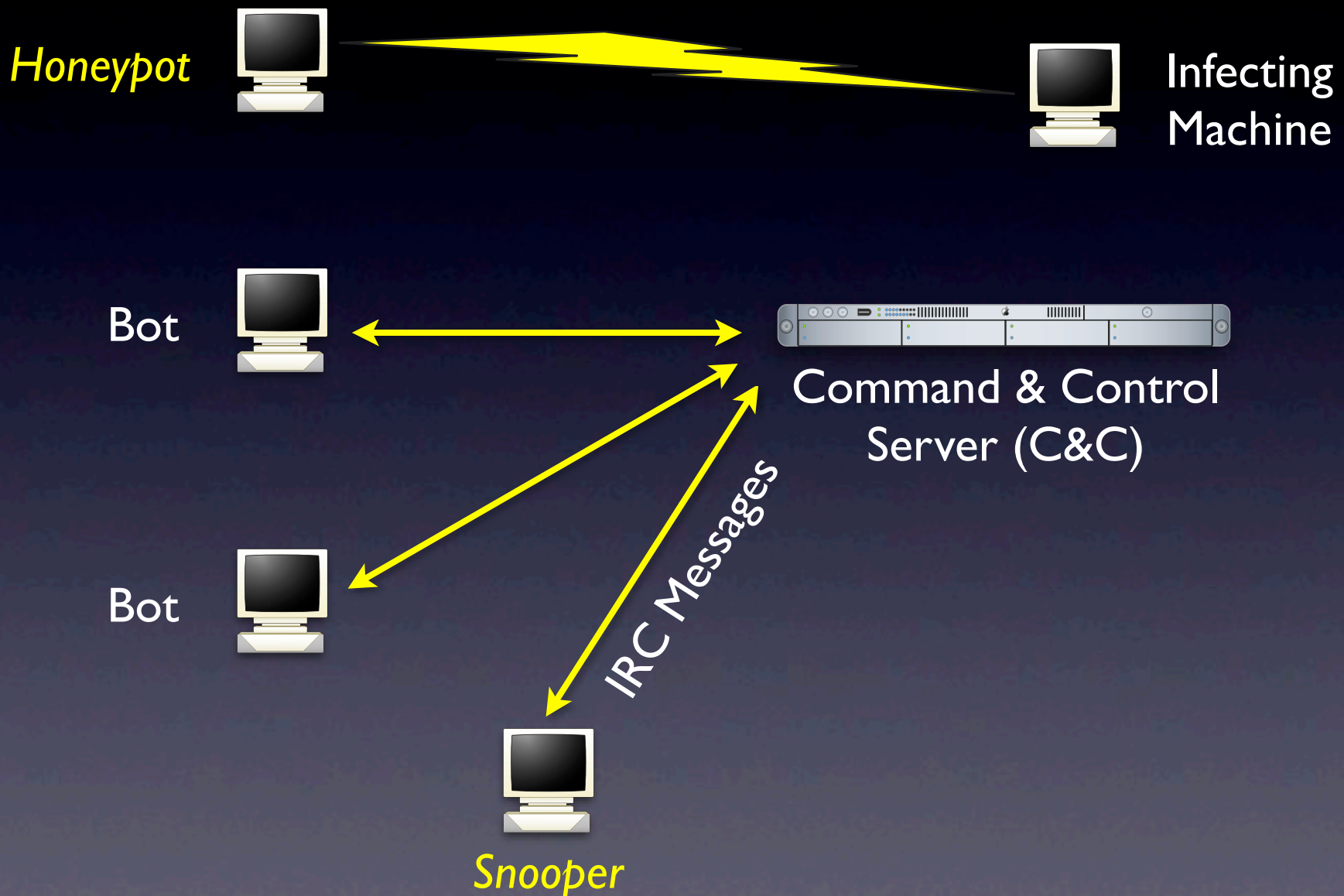
Botnet Lifecycle (Traditional View)

Infecting Machine

Botnet Lifecycle (Traditional View)



Tools for Monitoring



Botnet Operators' Response

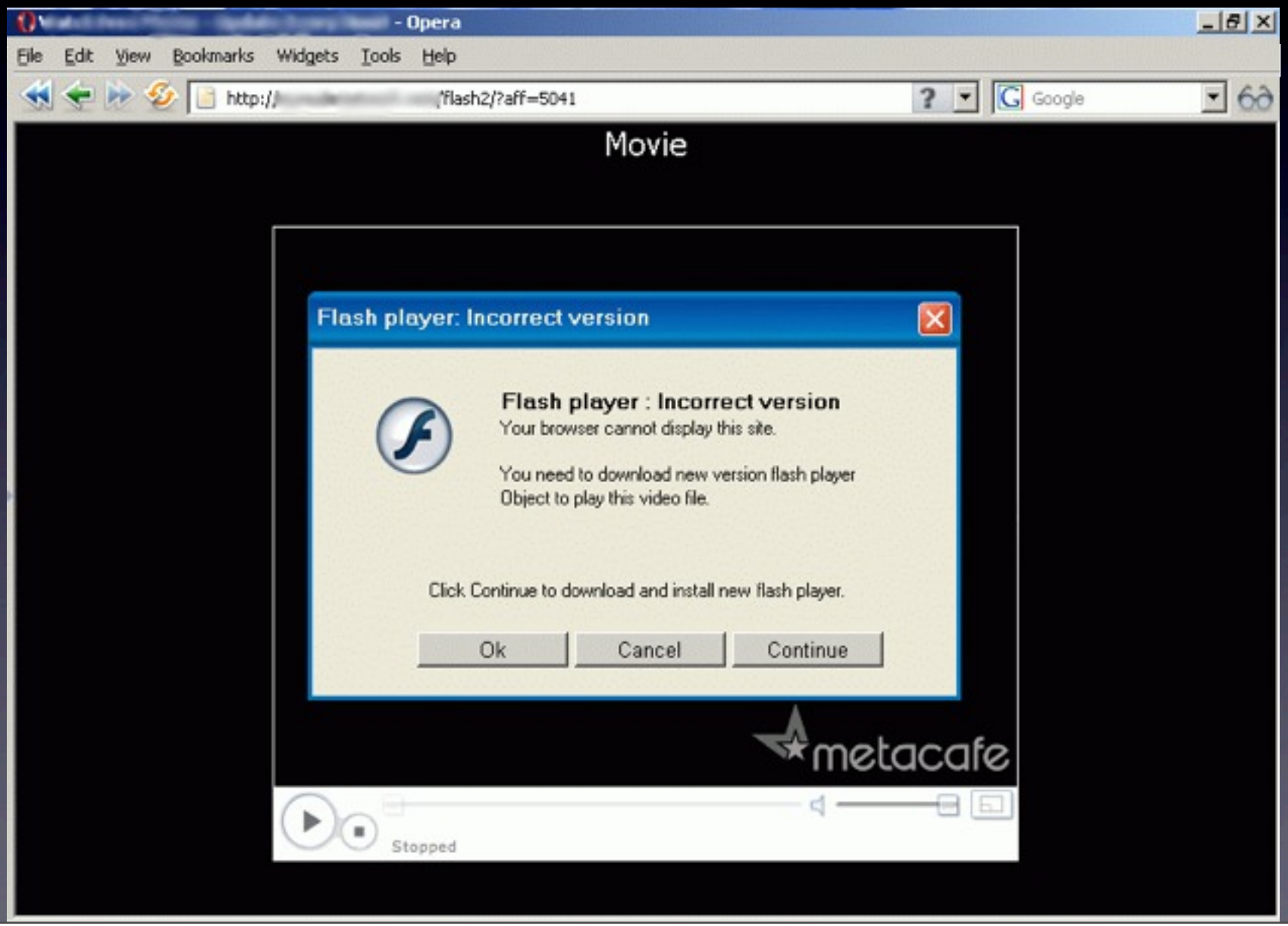
- Use *social engineering* techniques for infection
  - Cleverly crafted emails/websites induce users to download malicious programs

Botnet Operators' Response

- Use *social engineering* techniques for infection
  - Cleverly crafted emails/websites induce users to download malicious programs

Botnet Operators' Response

- Use *social engineering* techniques for infection



Botnet Operators' Response

- Use *social engineering* techniques for infection
  - Cleverly crafted emails/websites induce users to download malicious programs
- *Detect* virtualization
- Use *customized protocols* over HTTP
- Use *dynamic adaptation*
  - Malware binaries morph every few minutes
  - FastFlux DNS allows for fast redirection to new C&C
  - Change C&C protocols as well
- Serve malware/phishing from *compromised* websites

Finding vulnerable servers

- How are vulnerable servers found?
  - Brute force -- not very feasible
  - Use search to narrow scope
  - Lots of known bugs in php, asp, etc.
  - Underground sites post such vulnerabilities

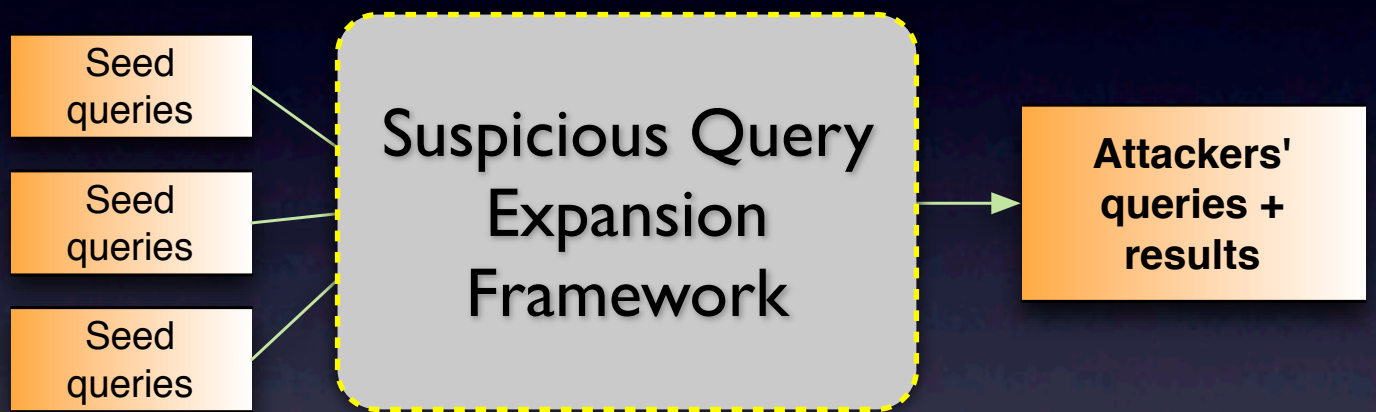
One such hacker site

A malicious query

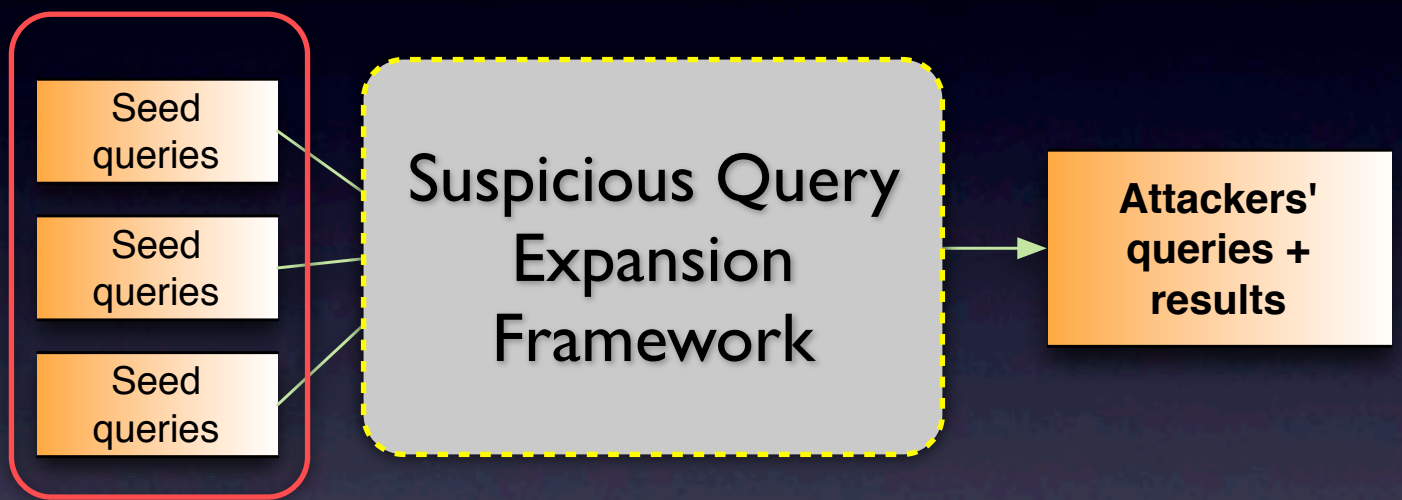
```
=================================================
DatalifeEngine 8.2 Remote File Inclusion Vulnerability
=================================================
++++++++++++++++++++++ Exploit ++++++++++++++++++++++
=================================================
<<->> google dork : Powered By DataLife Engine
<<->> Exploit ::

 >>> www.site/path /engine/api/api.class.php?dle_config_api=[shell.txt?]
```

Detecting Vulnerability Searches

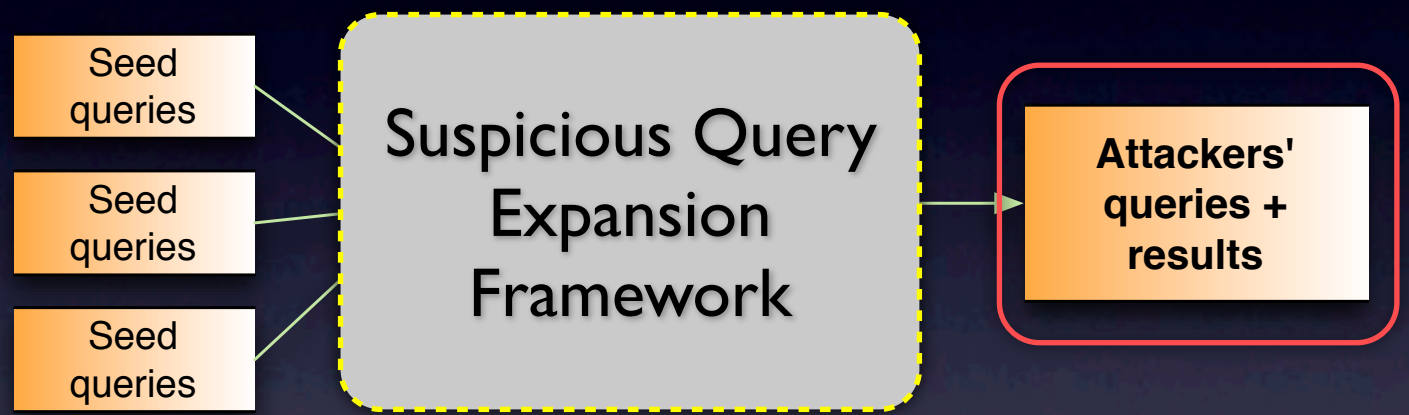


Detecting Vulnerability Searches



- 70 seed queries
- From `milw0rm`

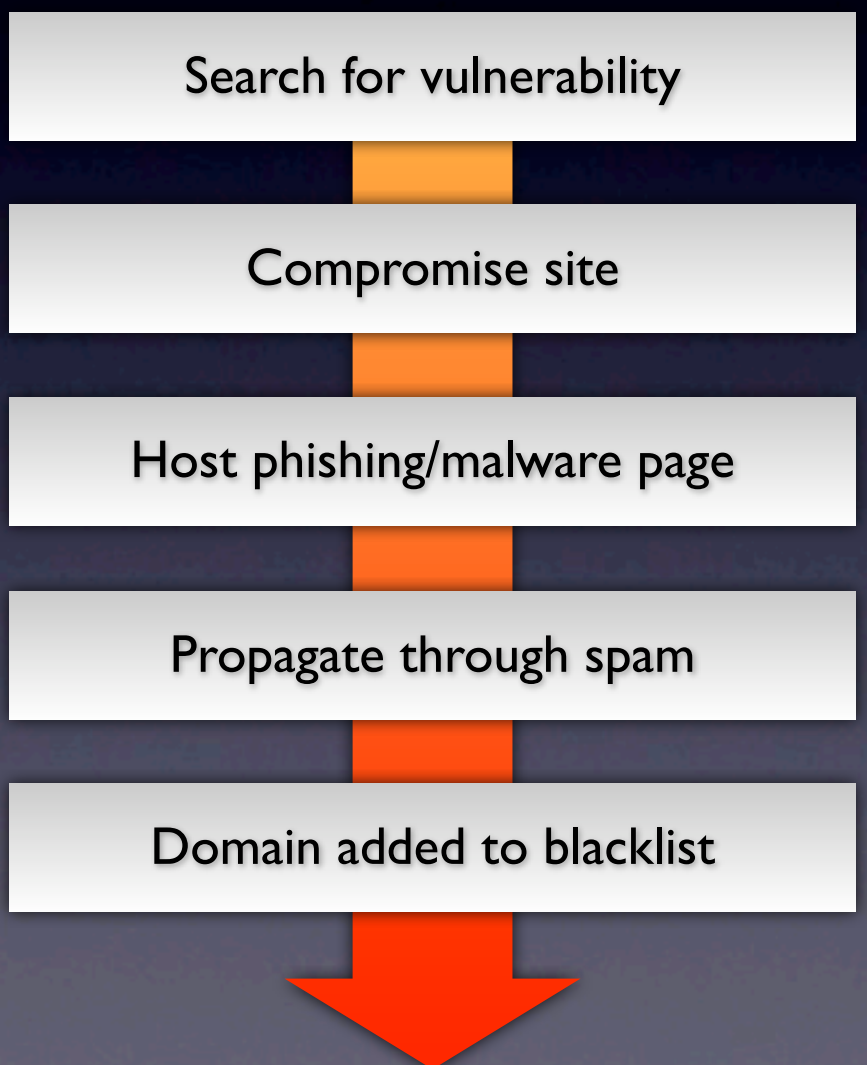
Detecting Vulnerability Searches



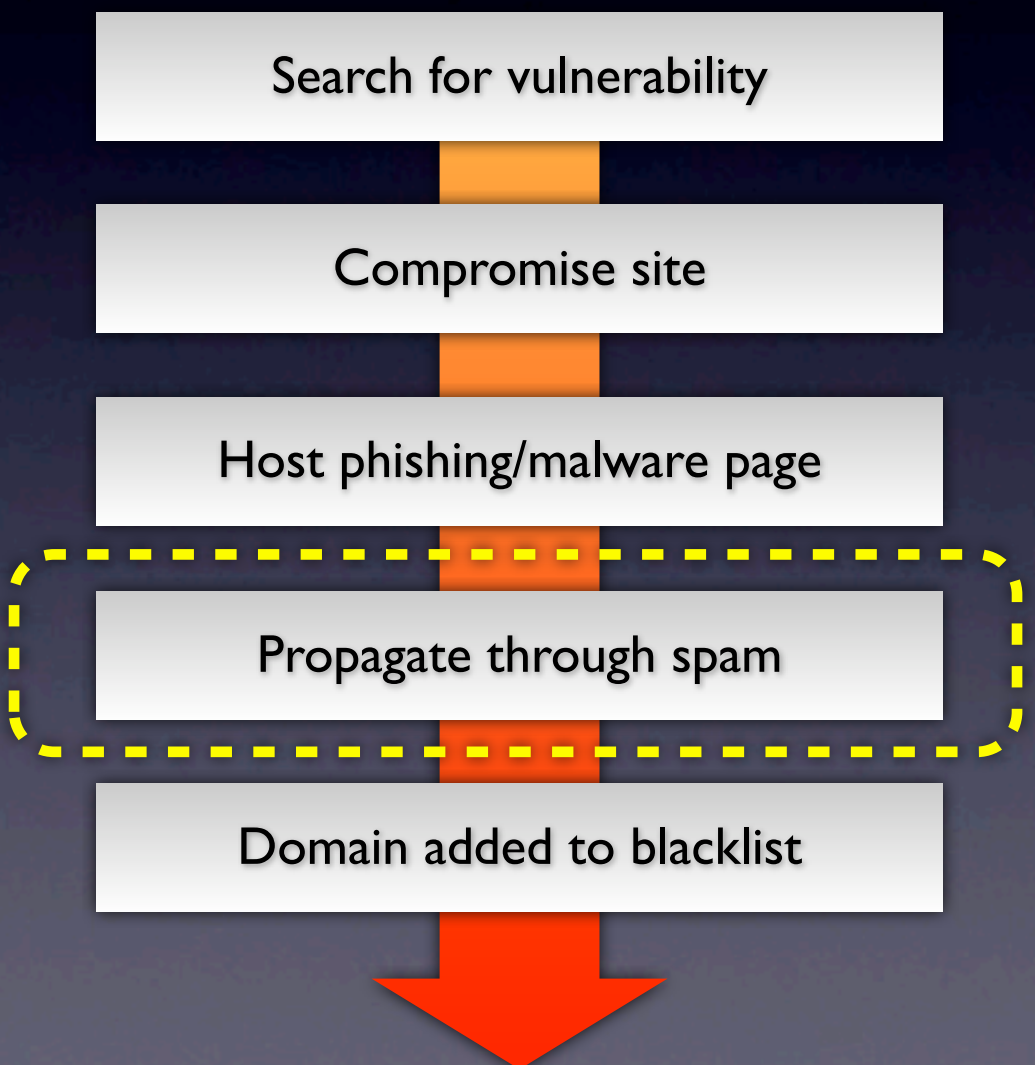
- 70 seed queries
- From `milw0rm`

- 1.2M searches
- 16k unique queries
- 436 IPs

An attacker's view

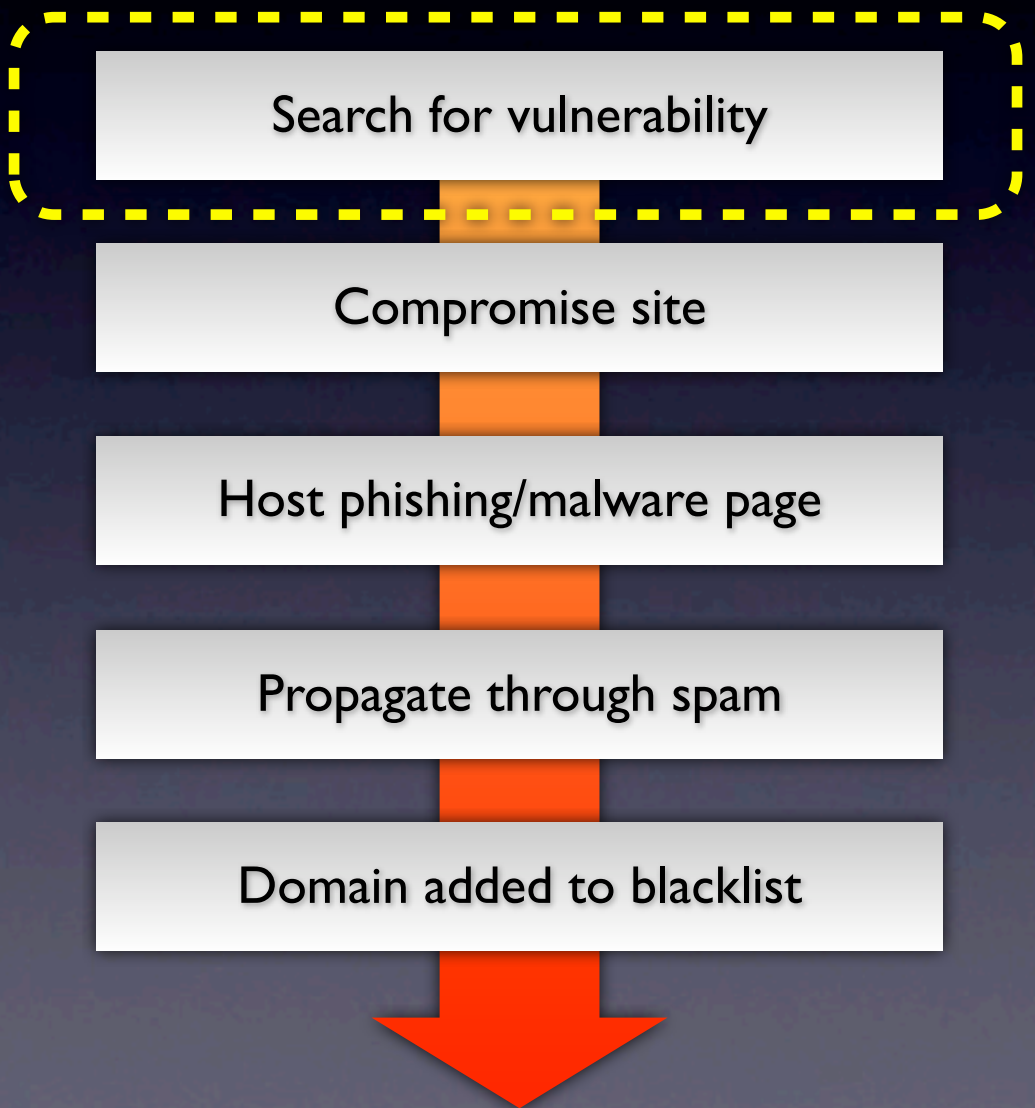


An attacker's view



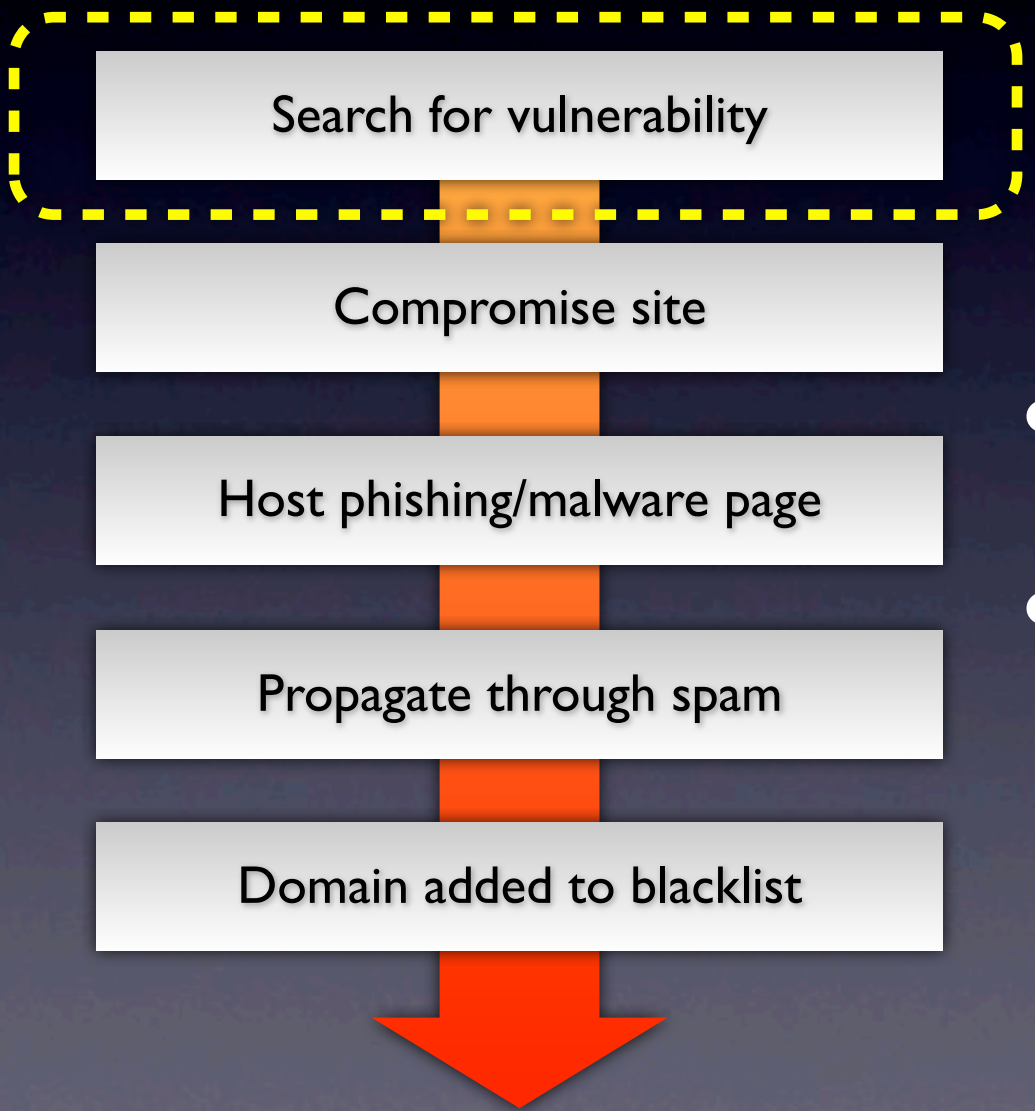
Detection

Defender's view



Possible detection

Defender's view



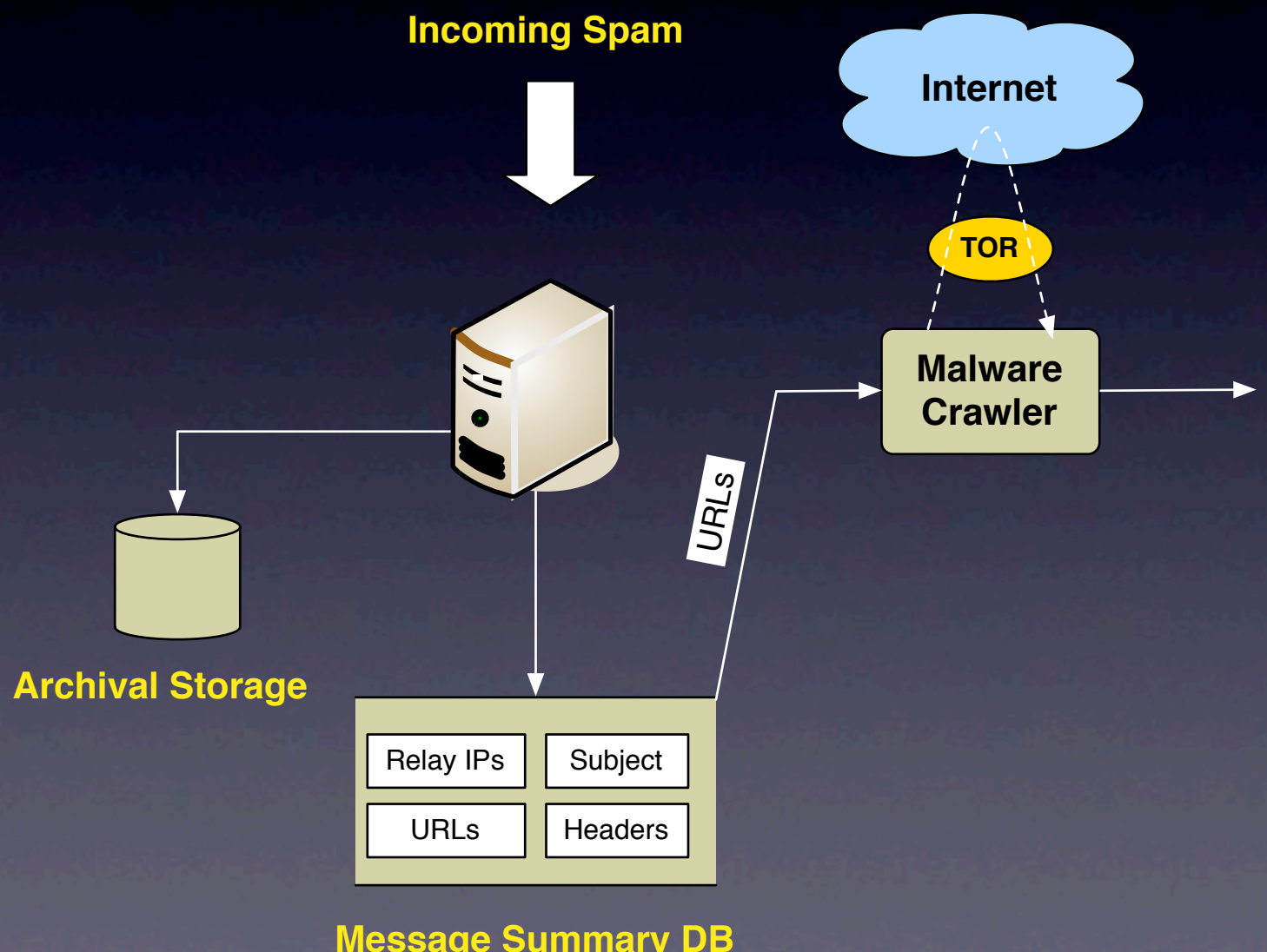
Possible detection

- Can proactively inform administrators
- Can predict which servers might be attacked

BotLab Design

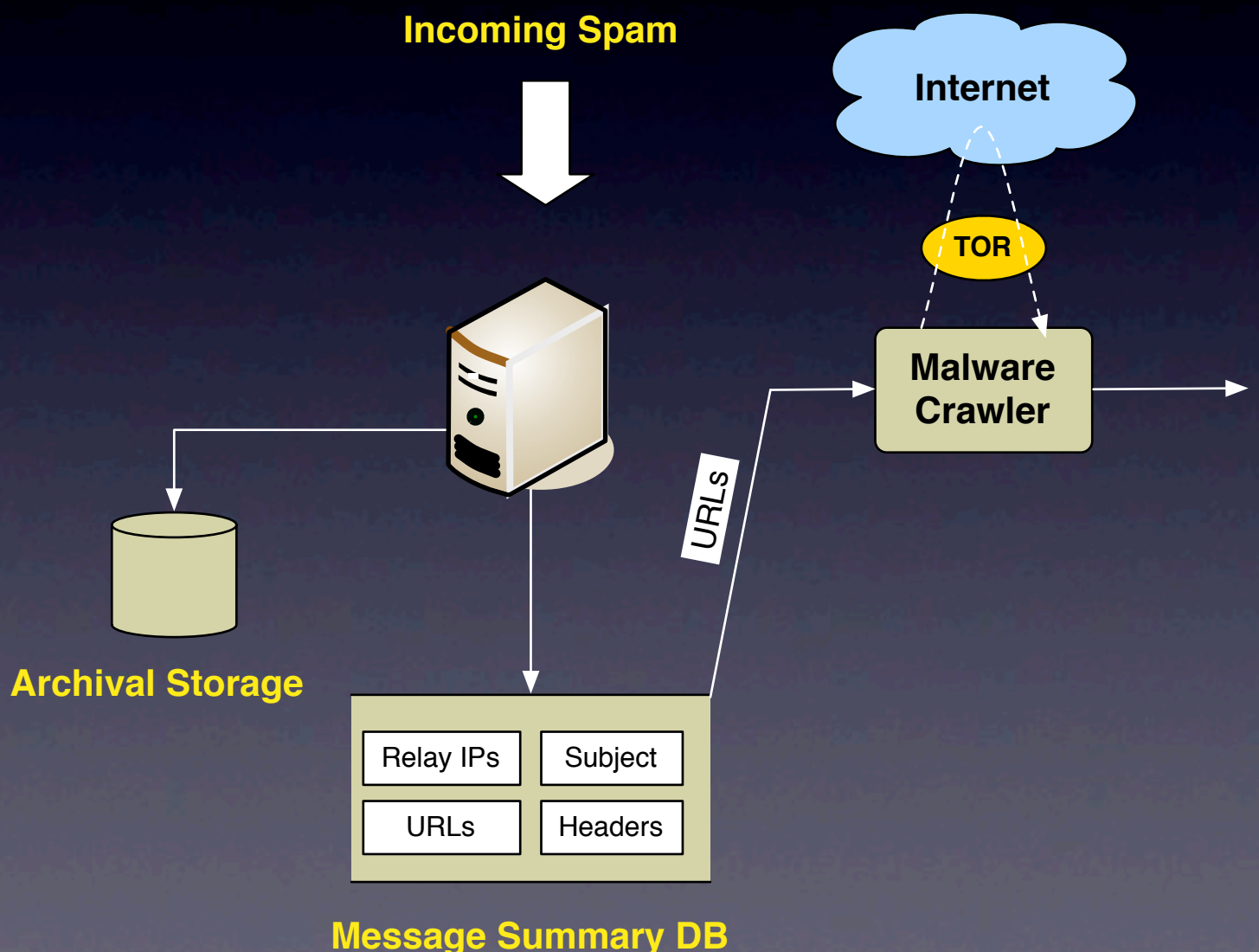
- *Active* as opposed to passive collection of binaries
- *Attribution:* run actual binaries and monitor behavior without causing harm
- *Scalably* identify duplicate binaries
- *Correlate* incoming spam with outgoing spam

1. Malware Collection



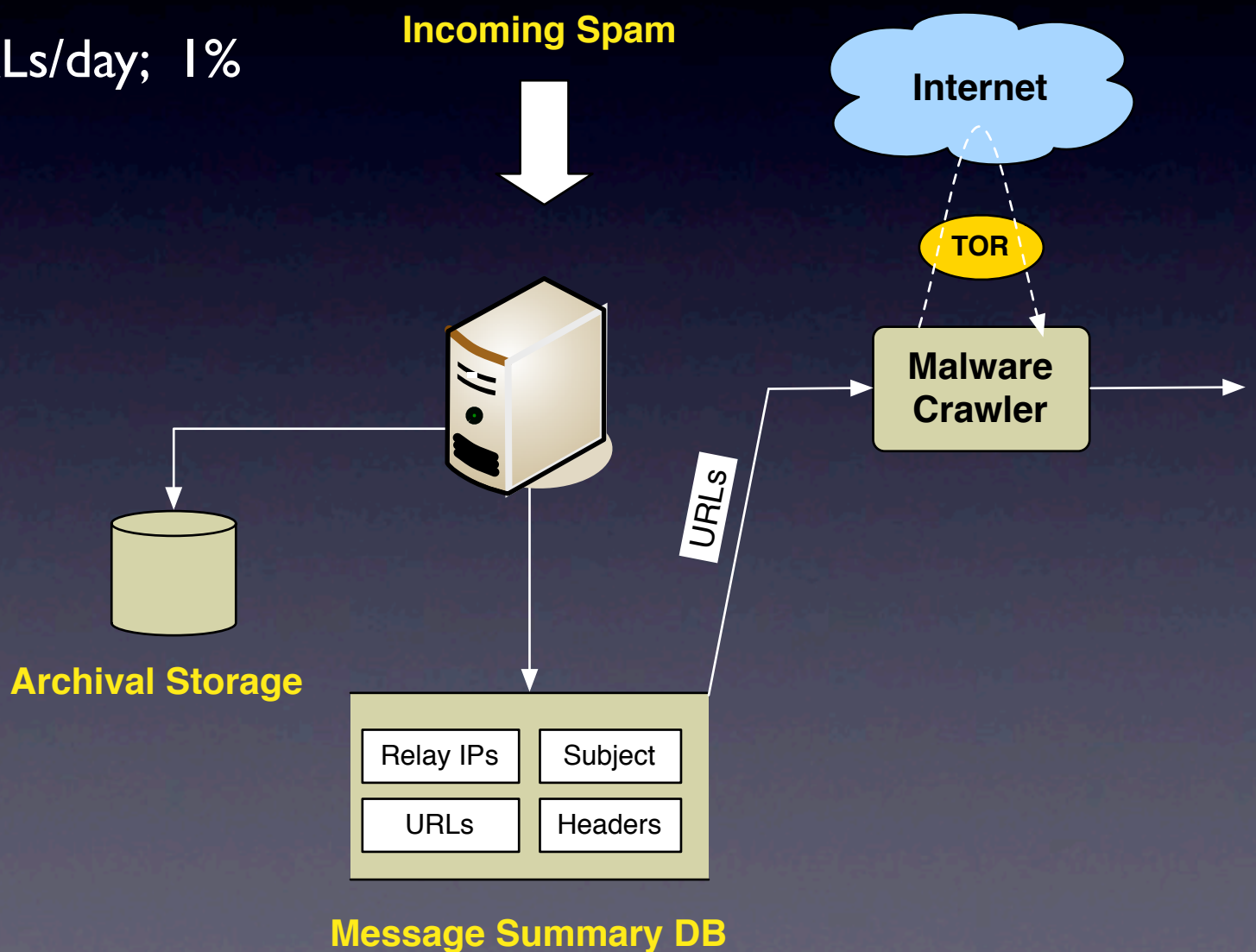
1. Malware Collection

- Augment honeypots with active crawling of spam URLs



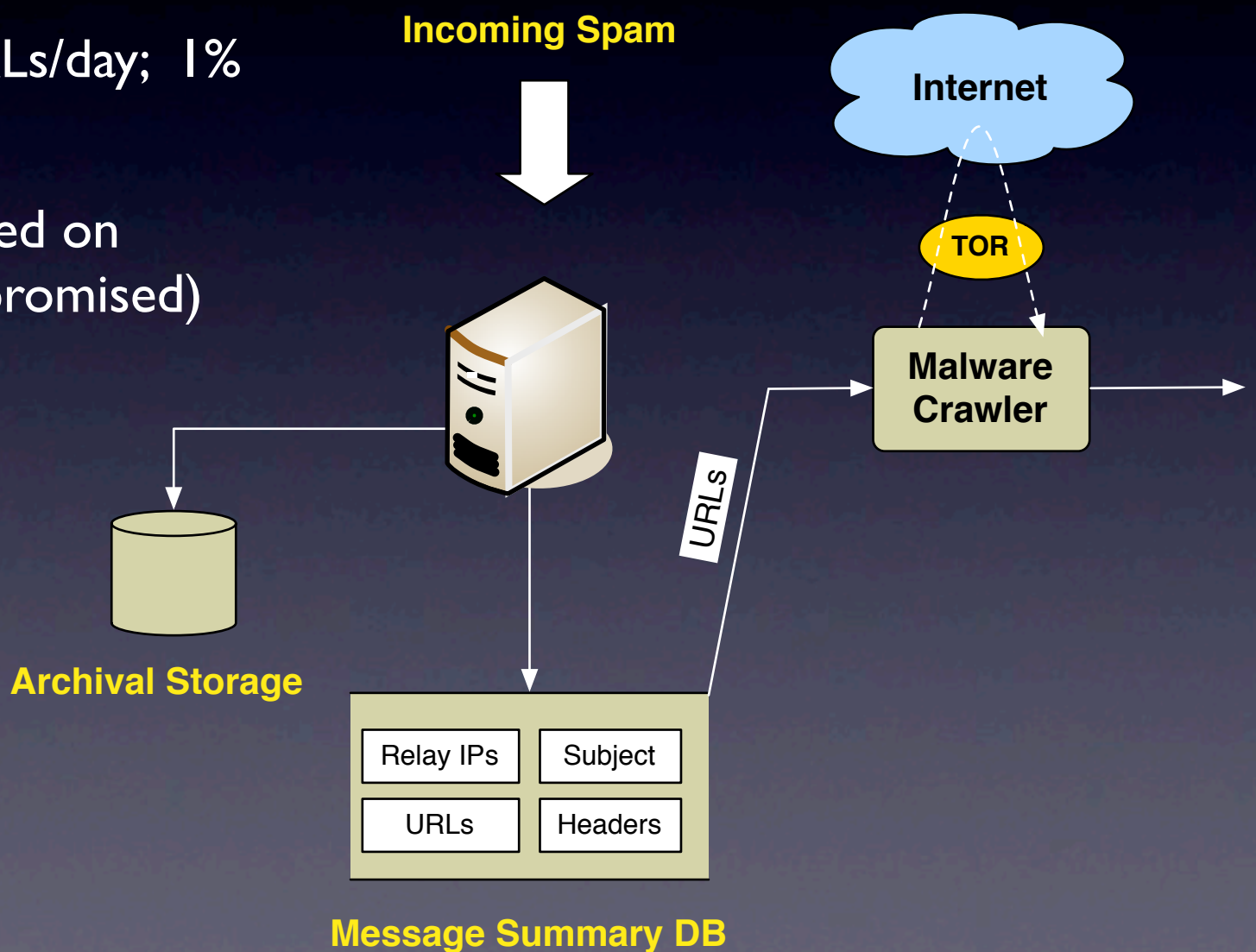
1. Malware Collection

- Augment honeypots with active crawling of spam URLs
- 100K unique URLs/day; 1% malicious

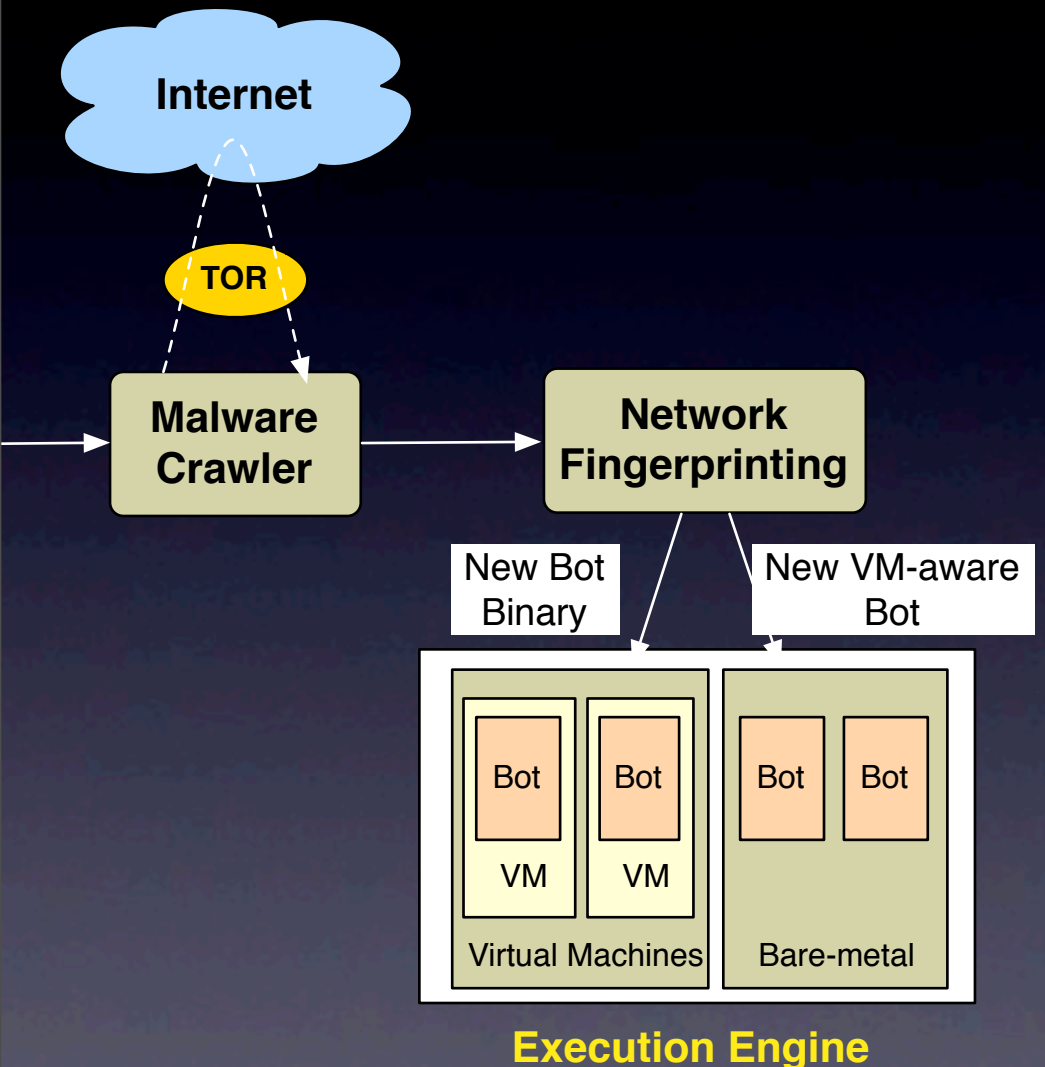


1. Malware Collection

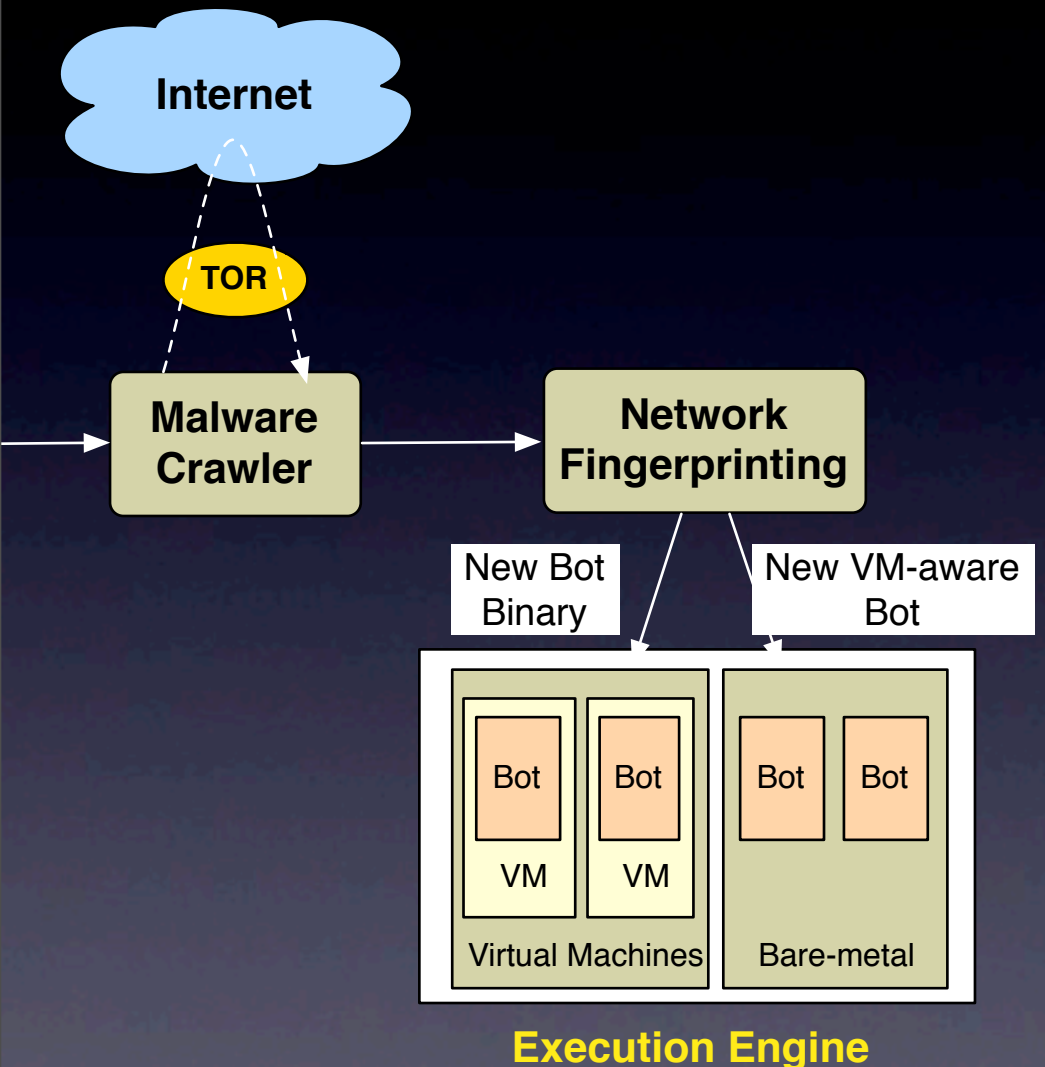
- Augment honeypots with active crawling of spam URLs
- 100K unique URLs/day; 1% malicious
- Most URLs hosted on legitimate (compromised) webservers



2. Network Fingerprinting

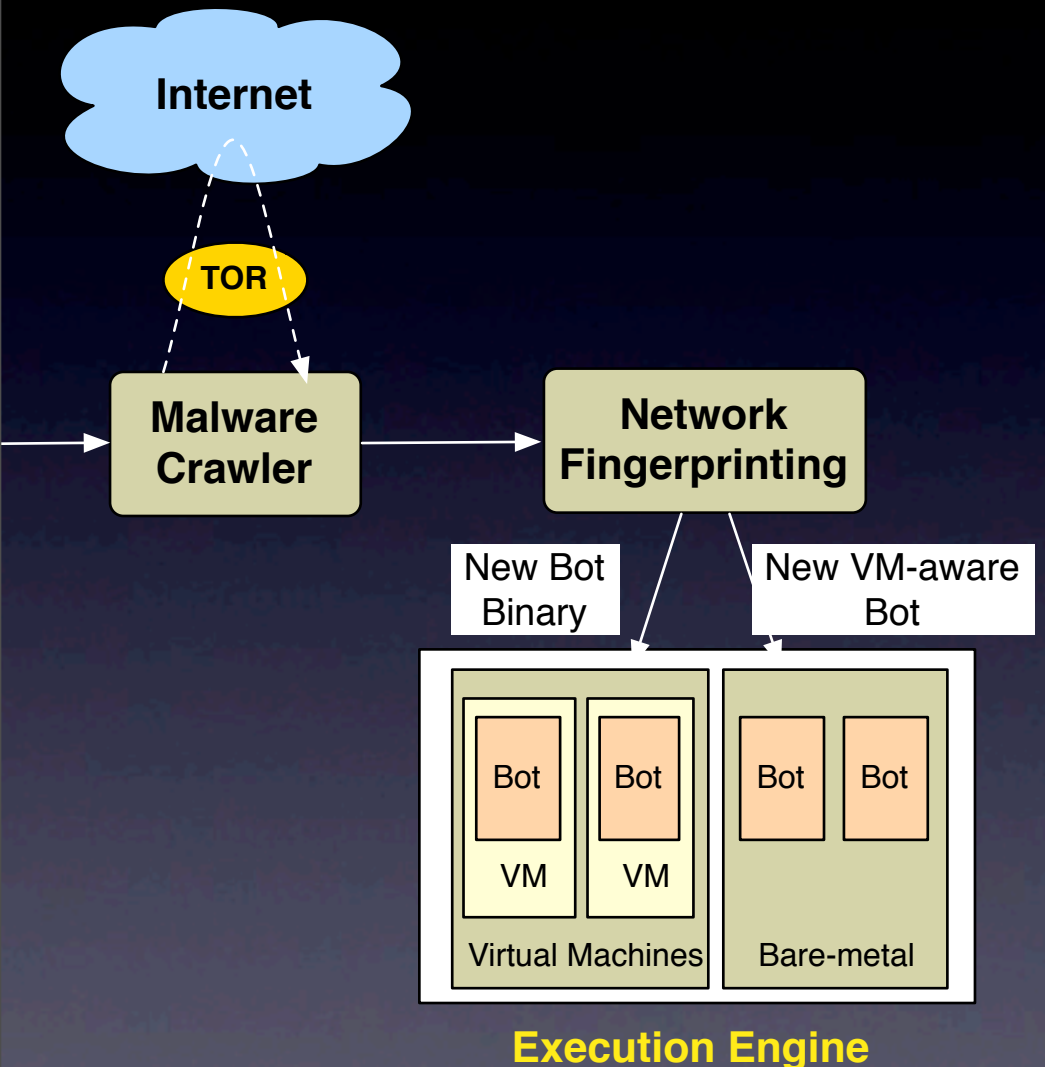


2. Network Fingerprinting



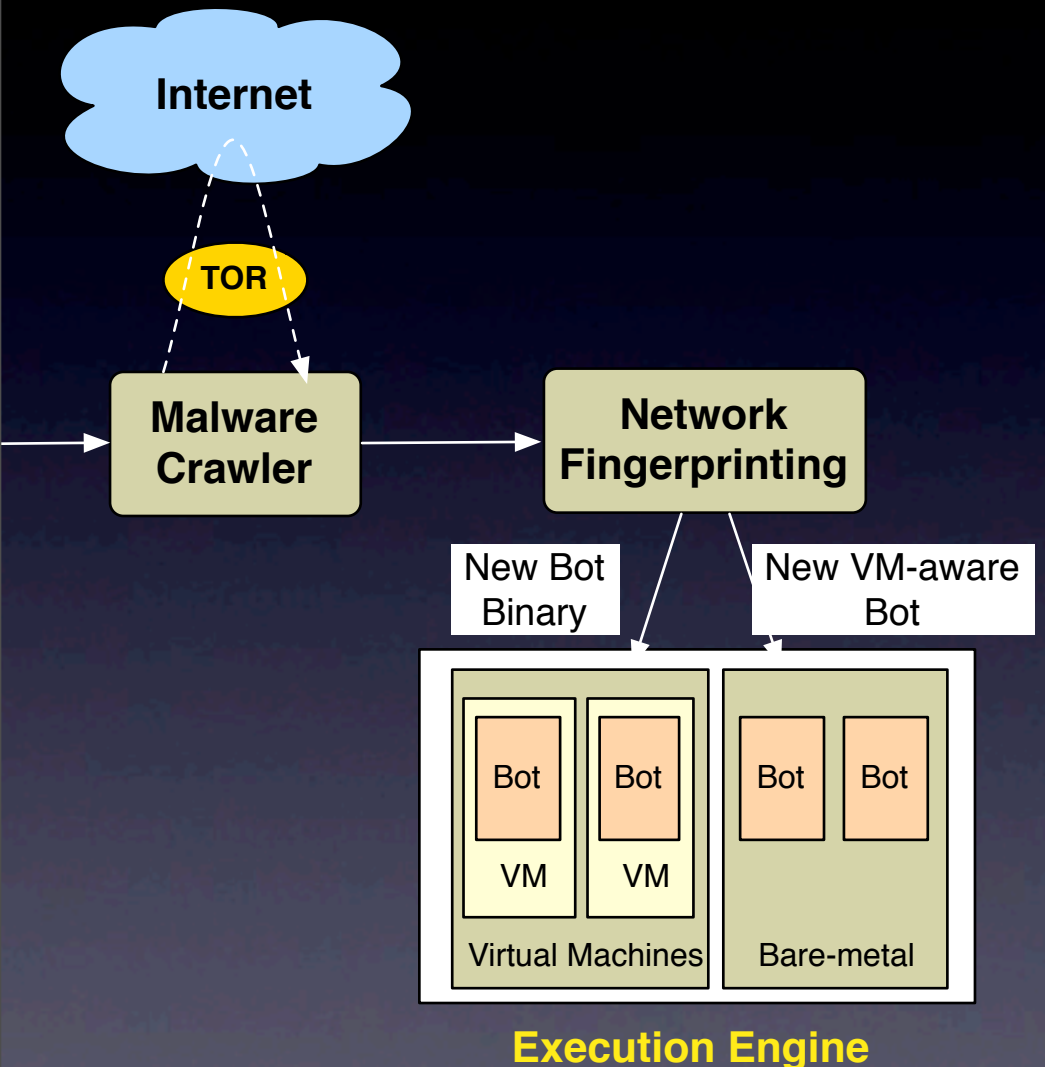
- Goal: find new bots while discarding duplicates

2. Network Fingerprinting



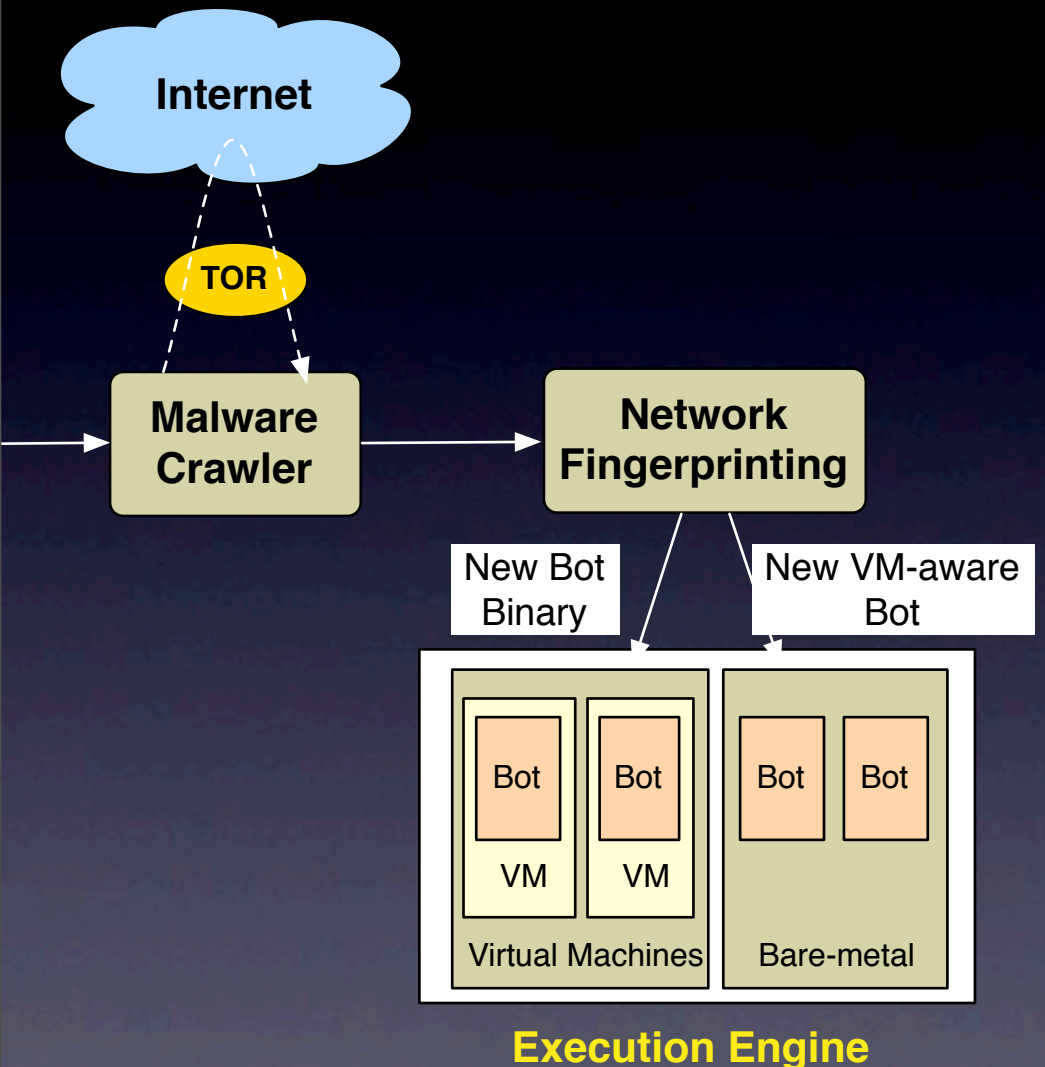
- Goal: find new bots while discarding duplicates
- *Simple hash* is insufficient

2. Network Fingerprinting



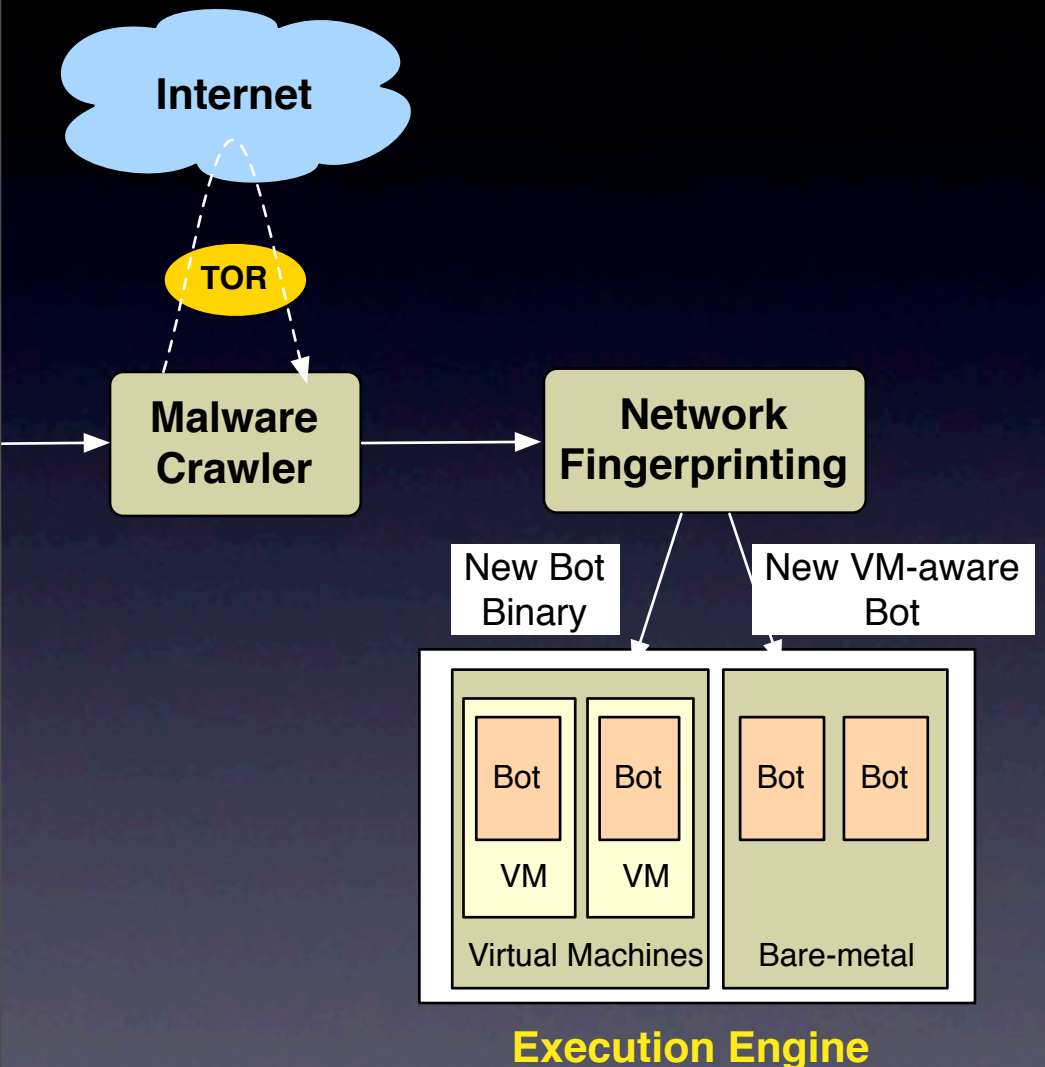
- Goal: find new bots while discarding duplicates
- *Simple hash* is insufficient
- Execute binaries and generate a *fingerprint*, which is a sequence of *flow records*

2. Network Fingerprinting



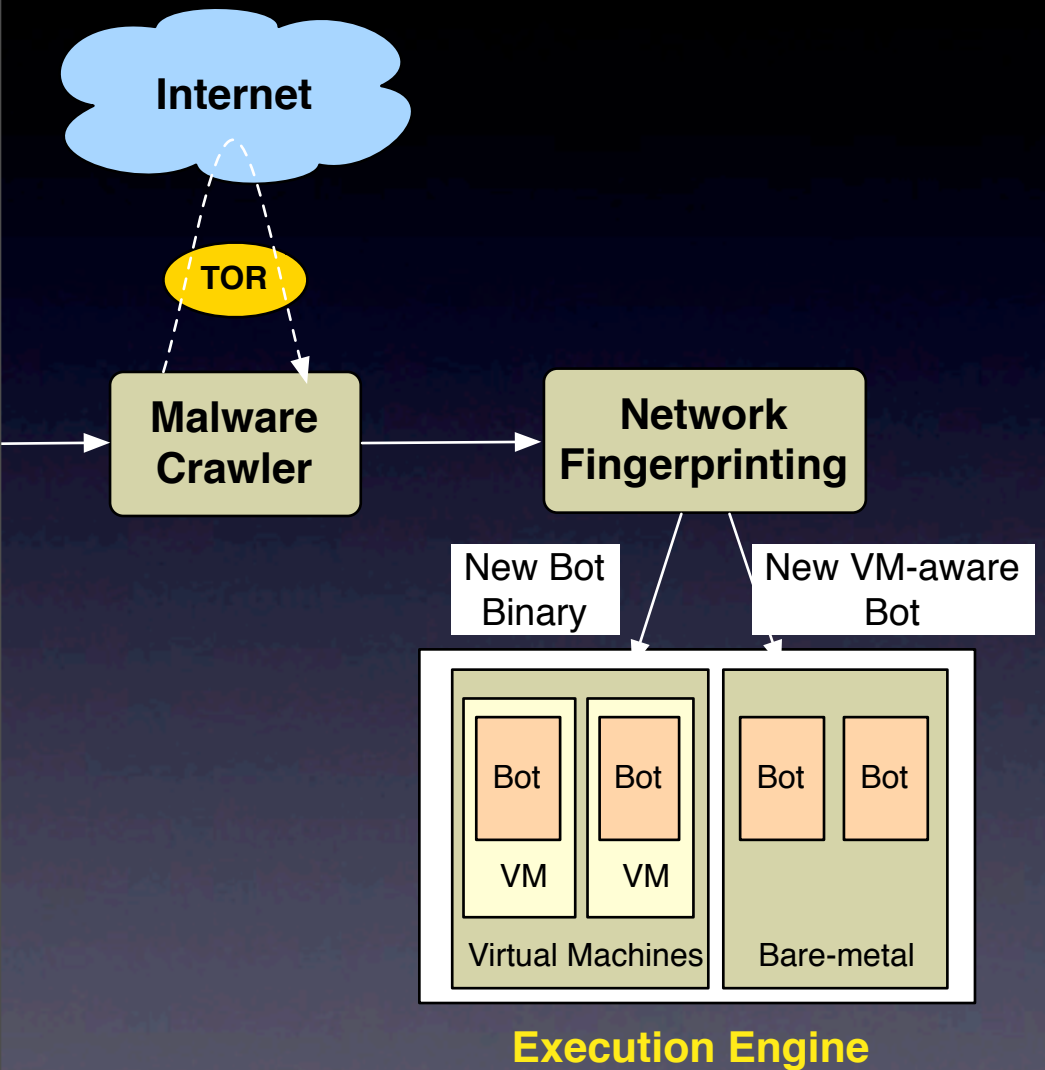
- Goal: find new bots while discarding duplicates
- *Simple hash* is insufficient
- Execute binaries and generate a *fingerprint*, which is a sequence of *flow records*
- Each *flow record* defined by (DNS, IP, TCP/UDP)

2. Network Fingerprinting



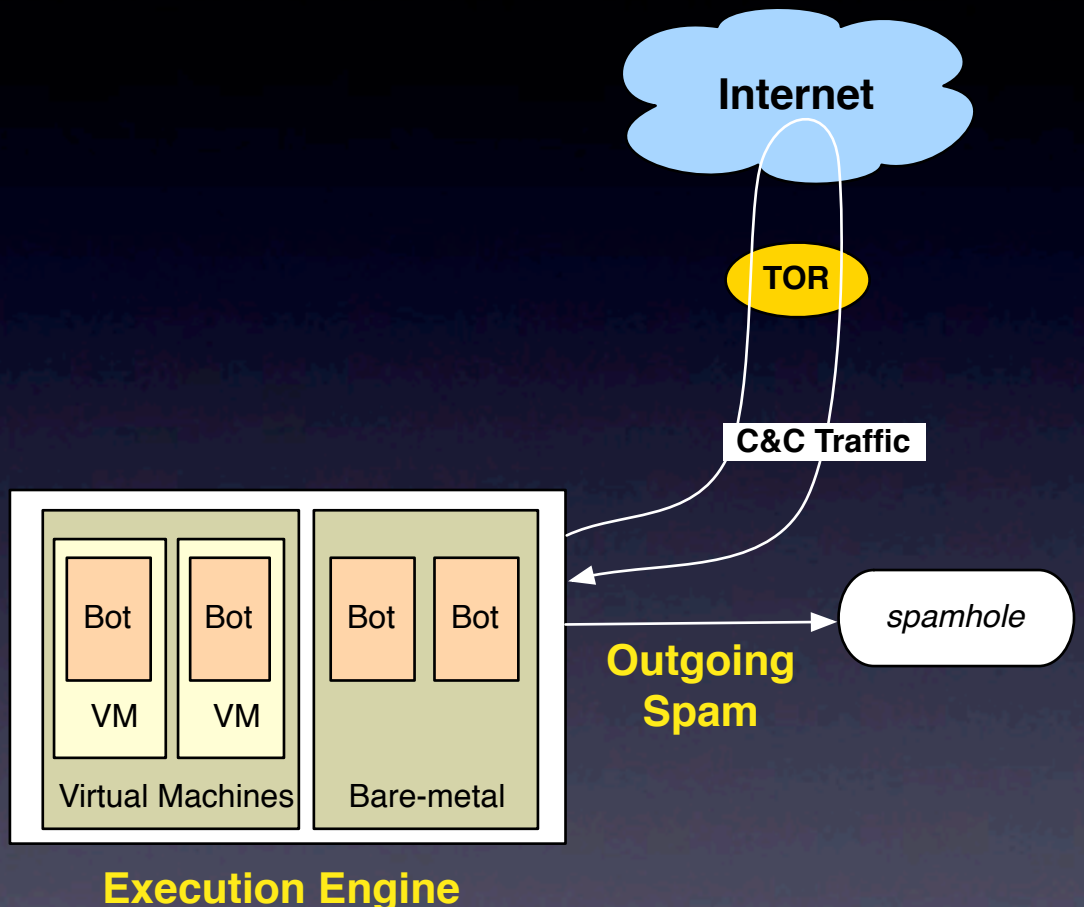
- Goal: find new bots while discarding duplicates
- *Simple hash* is insufficient
- Execute binaries and generate a *fingerprint*, which is a sequence of *flow records*
- Each *flow record* defined by (DNS, IP, TCP/UDP)
- Execute both inside and outside of VM to check for *VM detection*

2. Network Fingerprinting



- Goal: find new bots while discarding duplicates
- *Simple hash* is insufficient
- Execute binaries and generate a *fingerprint*, which is a sequence of *flow records*
- Each *flow record* defined by (DNS, IP, TCP/UDP)
- Execute both inside and outside of VM to check for *VM detection*
- Execute multiple times as some bots issue *random flows* (e.g., Google searches)

3. Monitor Running Bots

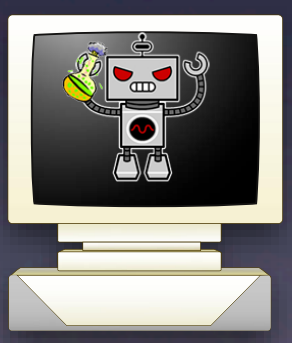


- Execute bots and trap all spam they send
- But need to *manually tweak* bots to get them to run

Manual Adjustments

- SMTP verification
  - One bot sent email to special server, which is verified later by the C&C server

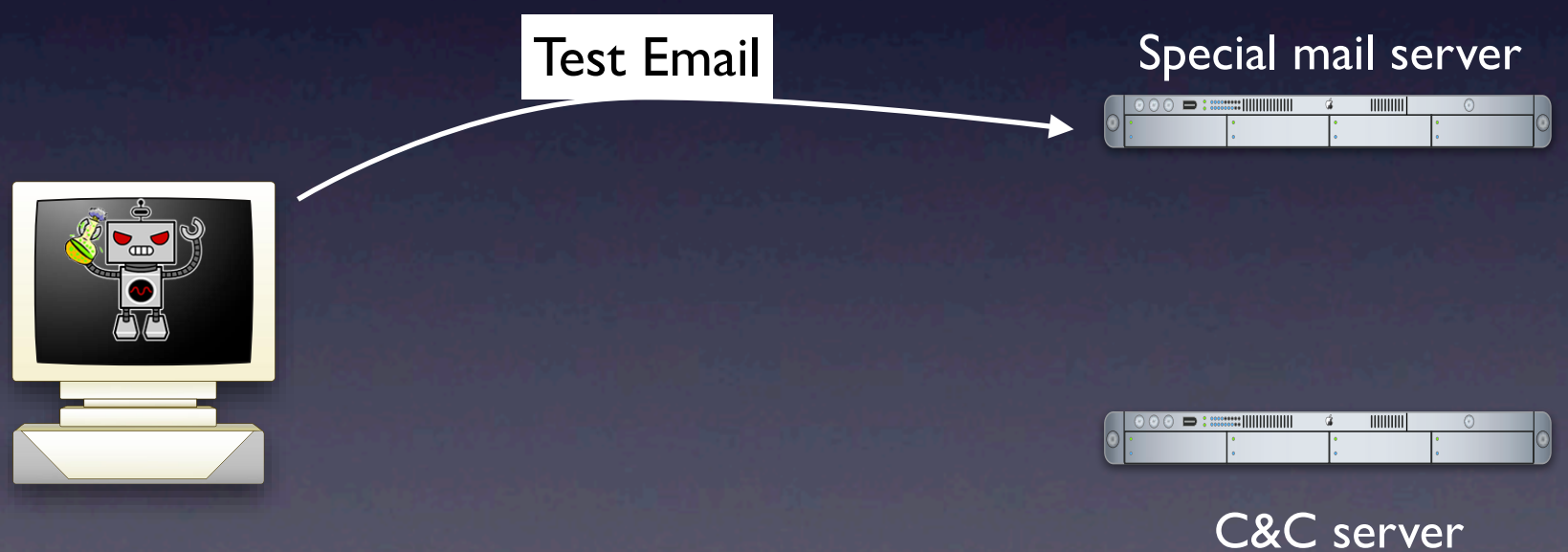
Special mail server



C&C server

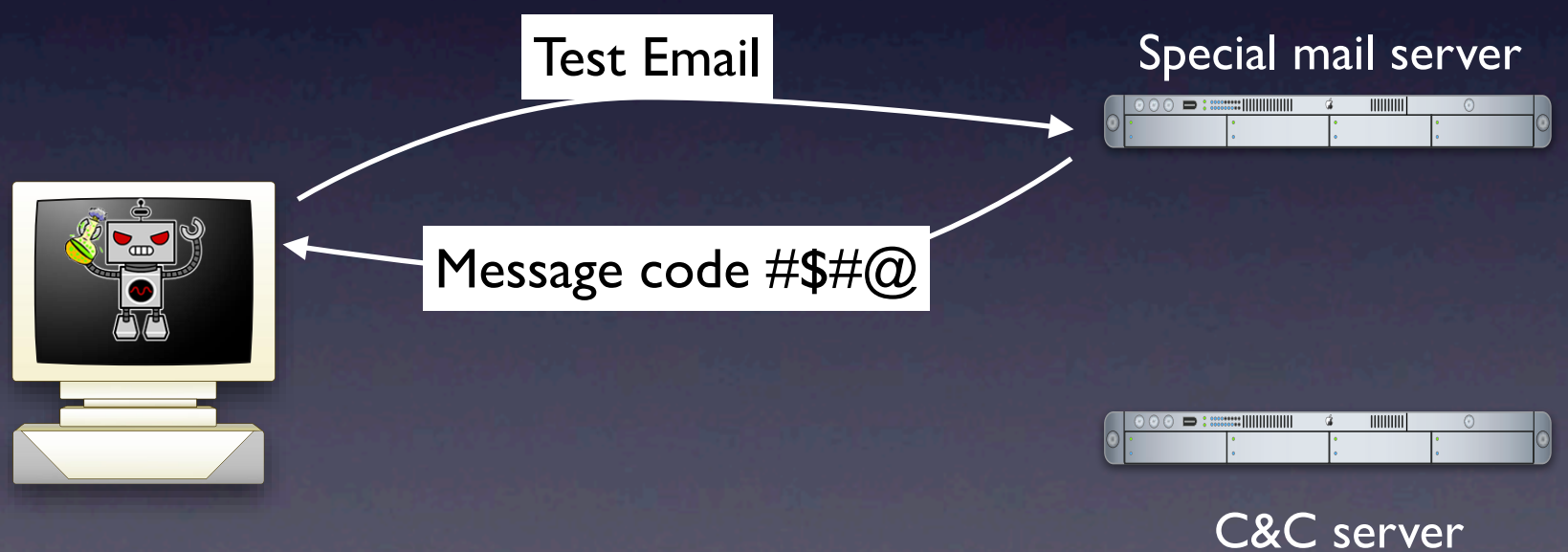
Manual Adjustments

- SMTP verification
  - One bot sent email to special server, which is verified later by the C&C server



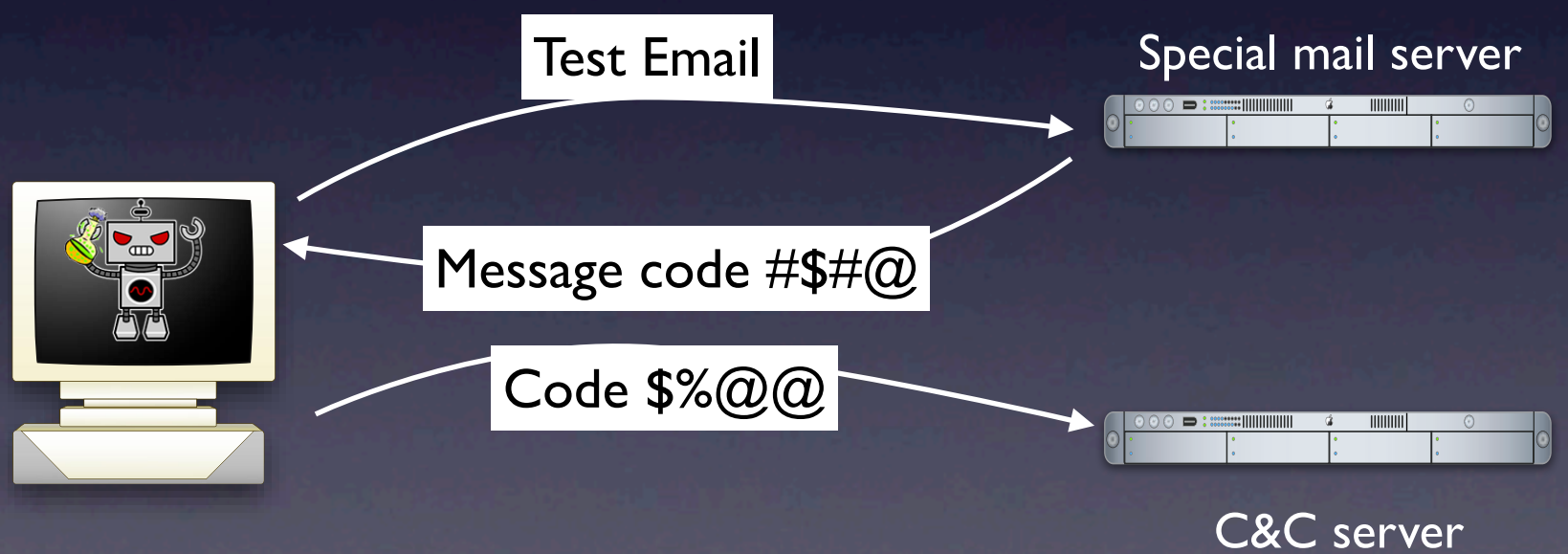
Manual Adjustments

- SMTP verification
  - One bot sent email to special server, which is verified later by the C&C server



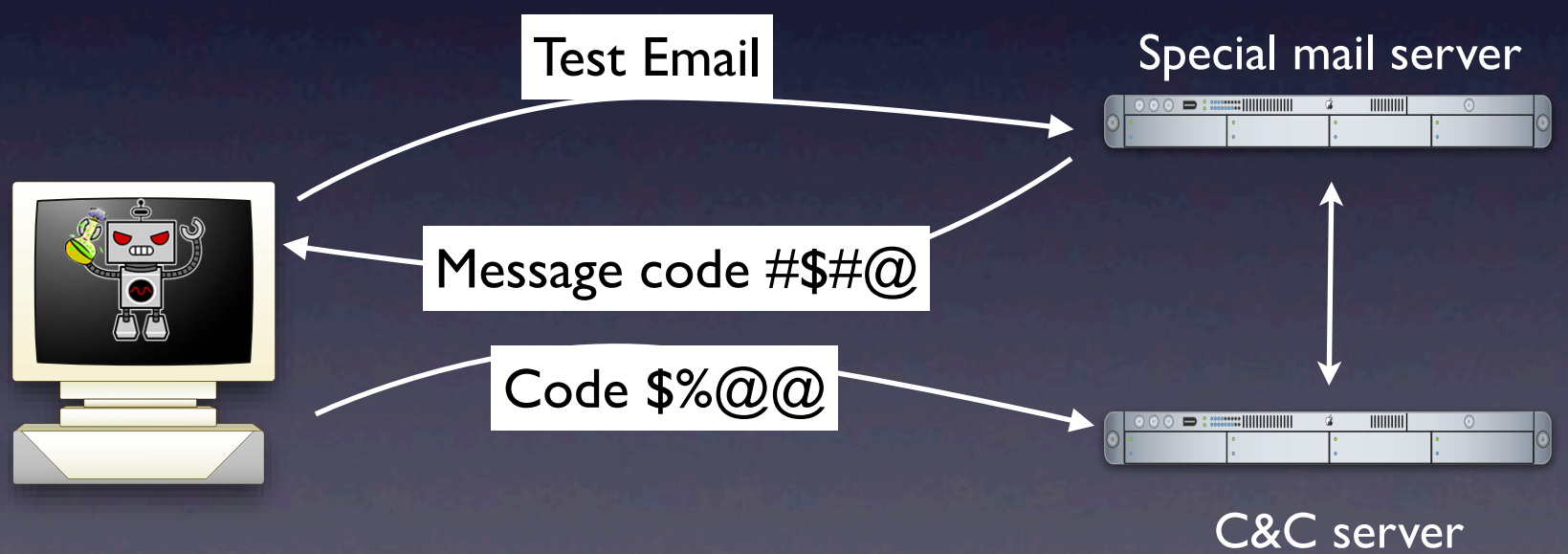
Manual Adjustments

- SMTP verification
  - One bot sent email to special server, which is verified later by the C&C server



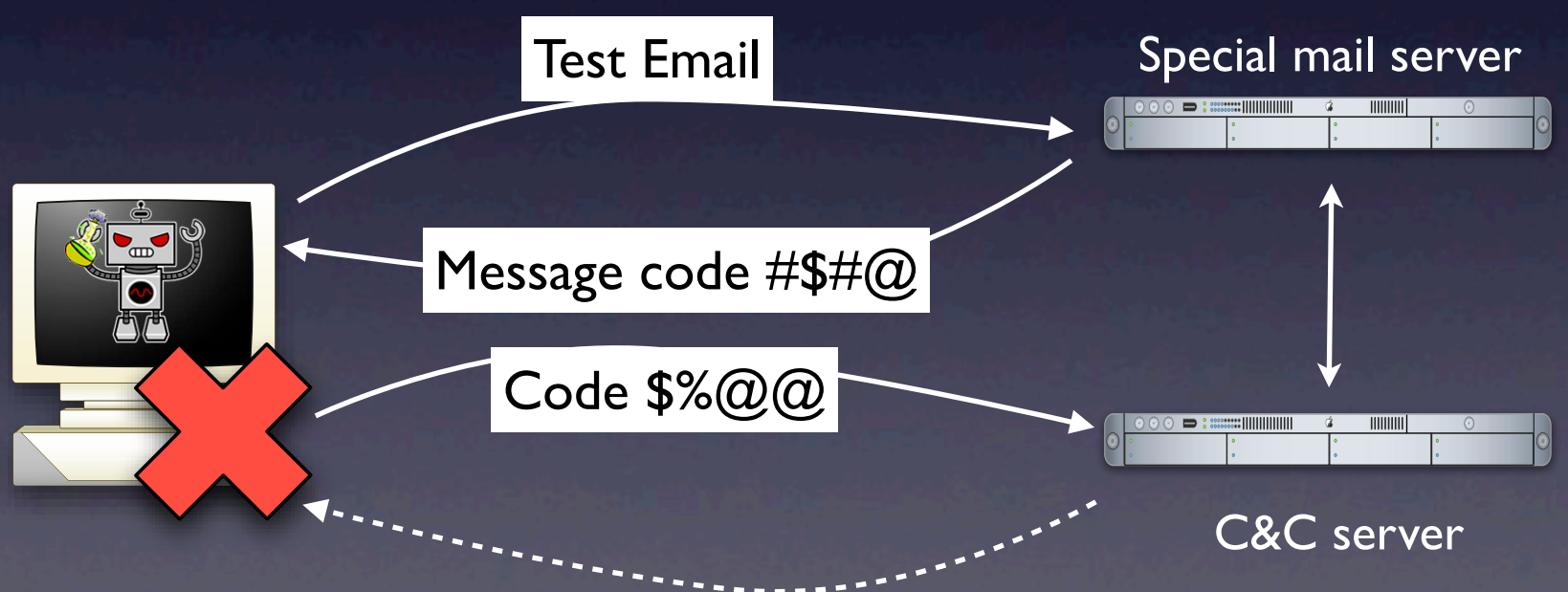
Manual Adjustments

- SMTP verification
  - One bot sent email to special server, which is verified later by the C&C server

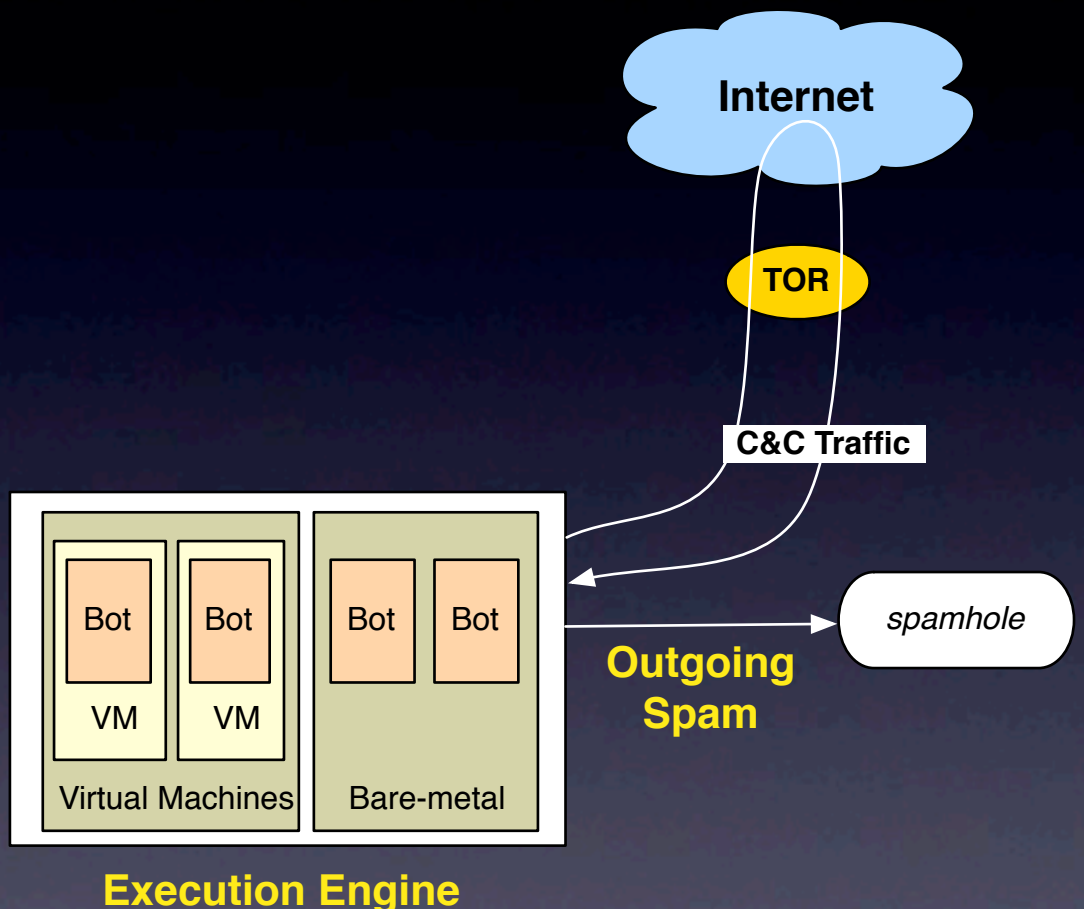


Manual Adjustments

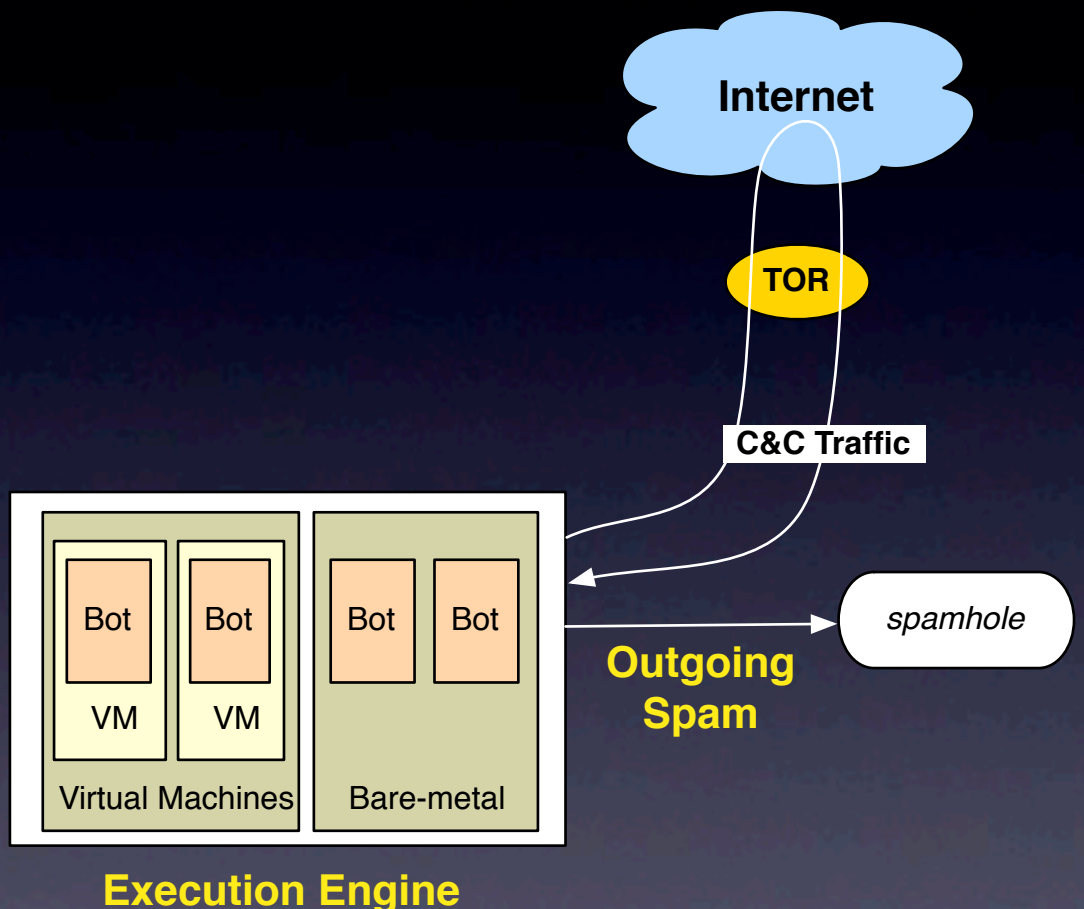
- SMTP verification
  - One bot sent email to special server, which is verified later by the C&C server



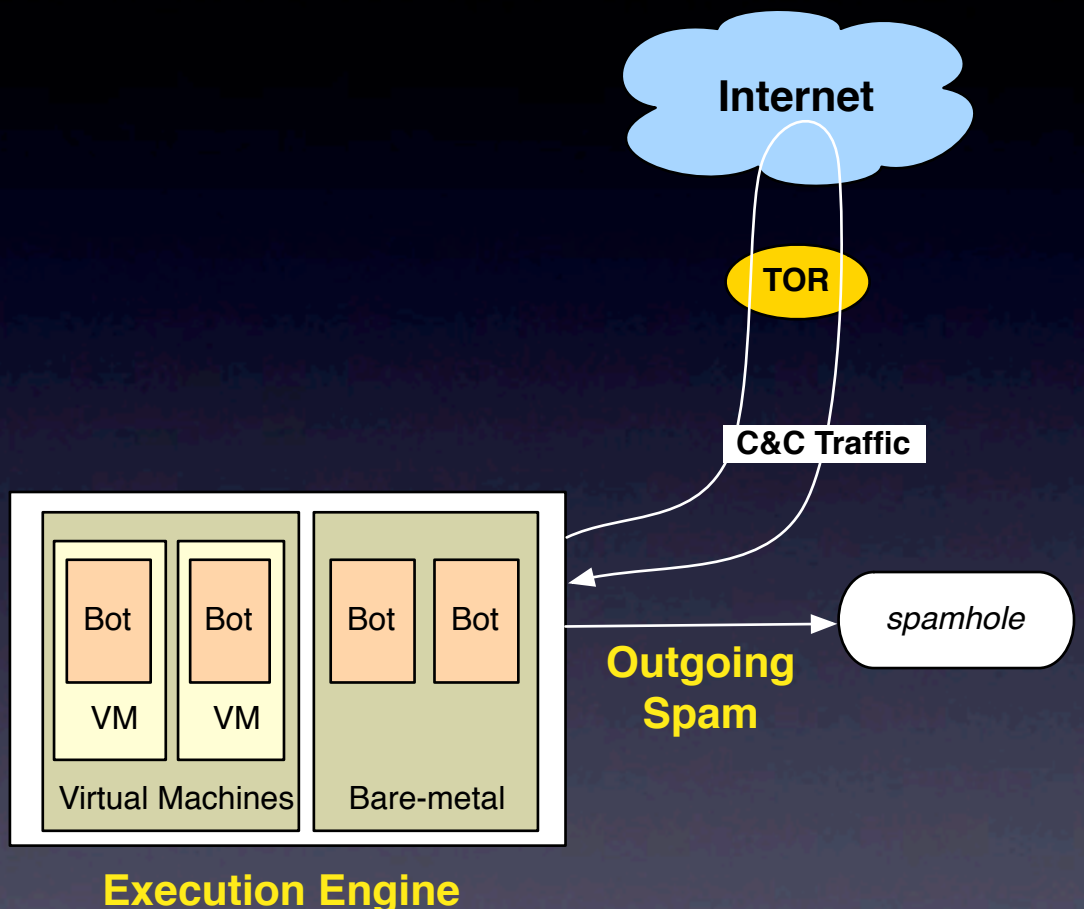
Coaxing Bots to Run



Coaxing Bots to Run

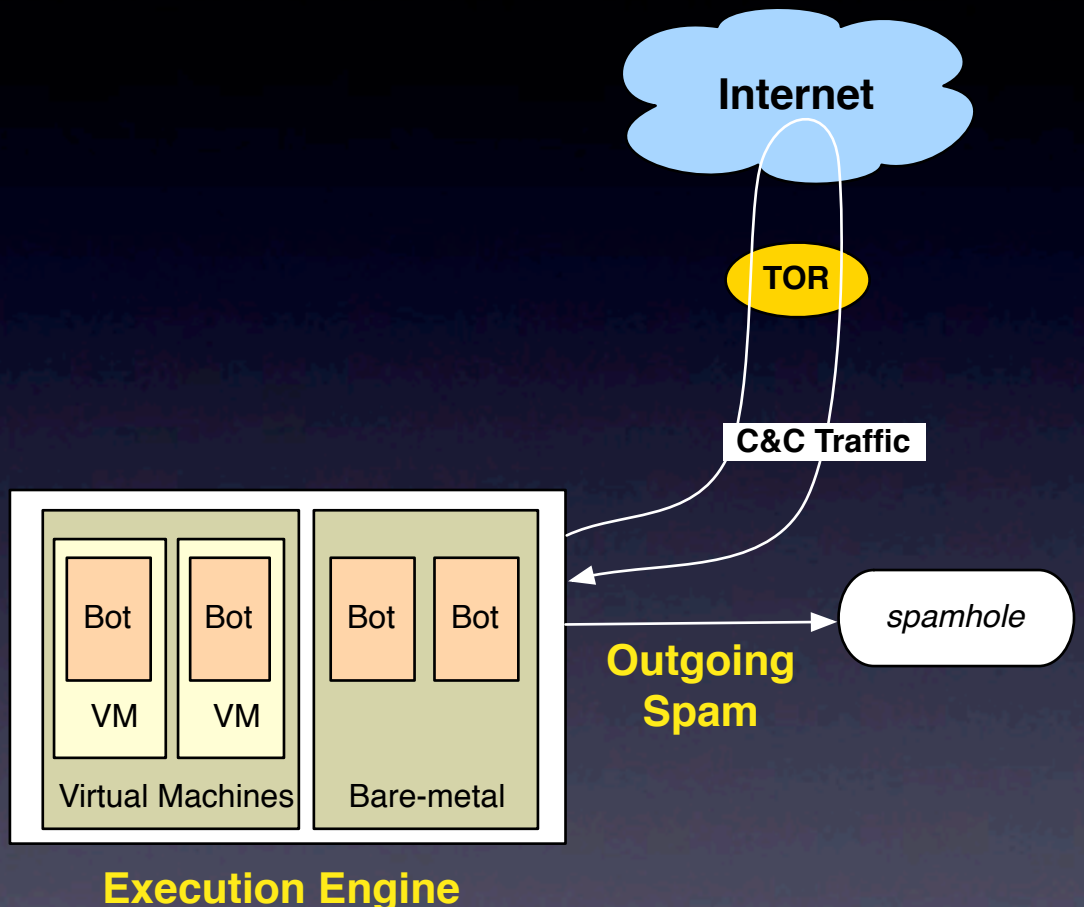


Coaxing Bots to Run



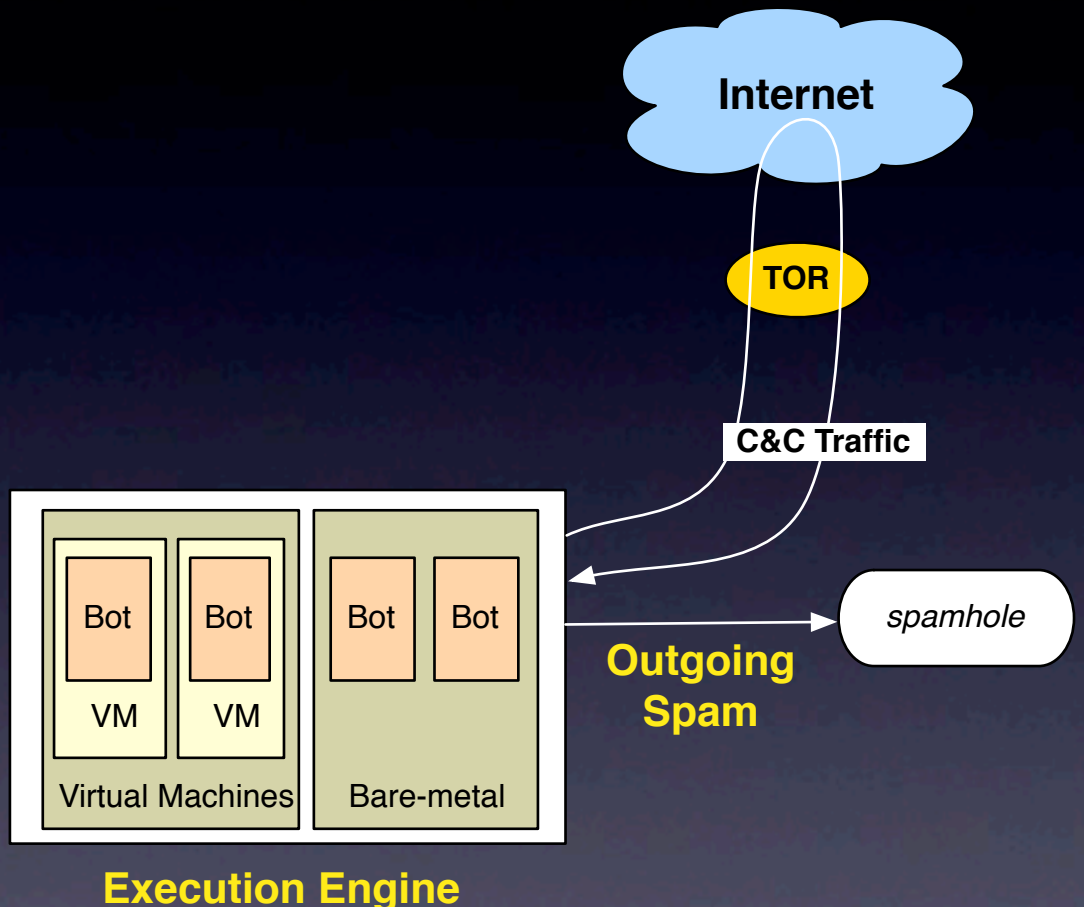
- Some bots send spam using webservices (such as HotMail)

Coaxing Bots to Run



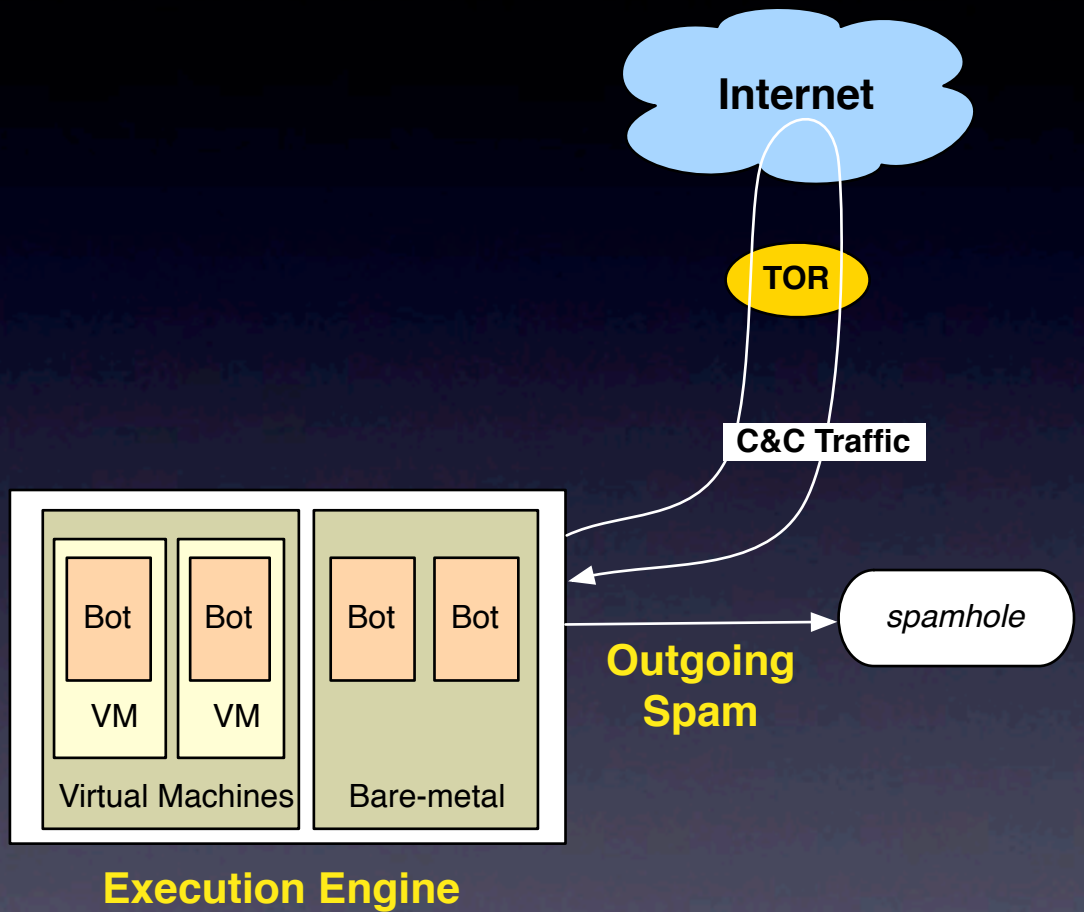
- Some bots send spam using webservices (such as HotMail)
- C&C servers are setup to blacklist suspicious IP ranges

Coaxing Bots to Run



- Some bots send spam using webservices (such as HotMail)
- C&C servers are setup to blacklist suspicious IP ranges
- Bots with 100% email delivery rate are considered suspicious

Coaxing Bots to Run



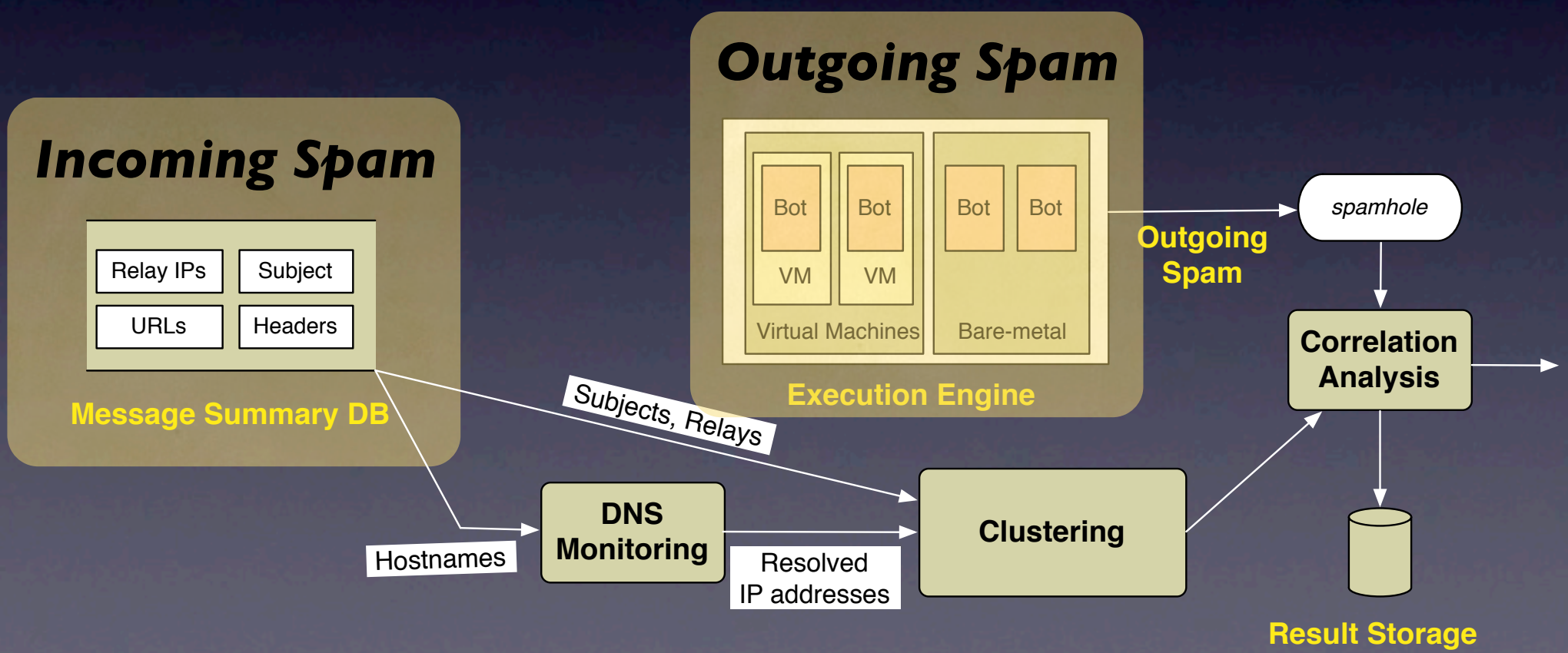
- Some bots send spam using webservices (such as HotMail)
- C&C servers are setup to blacklist suspicious IP ranges
- Bots with 100% email delivery rate are considered suspicious
- Fortunately only O(10) botnets; so manual tweaking possible

4. Clustering/Correlation Analysis

- Two sources of information:
  - Spam sent by bots running in BotLab (*Outgoing Spam*)
  - Spam received by UW (*Incoming Spam*)

4. Clustering/Correlation Analysis

- Two sources of information:
  - Spam sent by bots running in BotLab (*Outgoing Spam*)
  - Spam received by UW (*Incoming Spam*)



Measurements

- Analysis of *outgoing spam* feed
- Analysis of *incoming spam* feed
- *Correlation* of outgoing and incoming spam feeds

Behavioral Characteristics

Botnet	C&C Discovery	C&C servers contacted over lifetime	C&C protocol	spam send rate (msgs/min)
Grum				
Kraken				
Pushdo				
Rustock				
MegaD				
Srizbi				
Storm				

Behavioral Characteristics

Botnet	C&C Discovery	C&C servers contacted over lifetime	C&C protocol	spam send rate (msgs/min)
Grum	static IP	1		
Kraken	algorithmic DNS	41		
Pushdo	set of static IPs	96		
Rustock	static IP	1		
MegaD	static DNS name	21		
Srizbi	set of static IPs	20		
Storm	p2p (Overnet)	N/A		

Behavioral Characteristics

Botnet	C&C Discovery	C&C servers contacted over lifetime	C&C protocol	spam send rate (msgs/min)
Grum	static IP	1	encrypted HTTP	
Kraken	algorithmic DNS	41	encrypted HTTP	
Pushdo	set of static IPs	96	encrypted HTTP	
Rustock	static IP	1	encrypted HTTP	
MegaD	static DNS name	21	encrypted custom protocol (port 80)	
Srizbi	set of static IPs	20	unencrypted HTTP	
Storm	p2p (Overnet)	N/A	encrypted custom	

Behavioral Characteristics

Botnet	C&C Discovery	C&C servers contacted over lifetime	C&C protocol	spam send rate (msgs/min)
Grum	static IP	1	encrypted HTTP	344
Kraken	algorithmic DNS	41	encrypted HTTP	331
Pushdo	set of static IPs	96	encrypted HTTP	289
Rustock	static IP	1	encrypted HTTP	33
MegaD	static DNS name	21	encrypted custom protocol (port 80)	1638
Srizbi	set of static IPs	20	unencrypted HTTP	1848
Storm	p2p (Overnet)	N/A	encrypted custom	20

Botnet Mailing Lists

- Random fetch model allows us to estimate botnet mailing list sizes
  - As we see more of the spam feed, there will be more duplicates in recipient email addresses
- If mailing list size is *N* and if bot obtains *C* addresses for each C&C query, then probability that an email address will appear again in the next *K* emails is

$$1 - (1 - C/N)^{K/C}$$

- Some mailing list sizes: MegaD's is 850 million, Rustock's is 1.2 billion, Kraken's is 350 million
- Overlap between mailing lists is small (less than 28%)

Outgoing Spam Characteristics

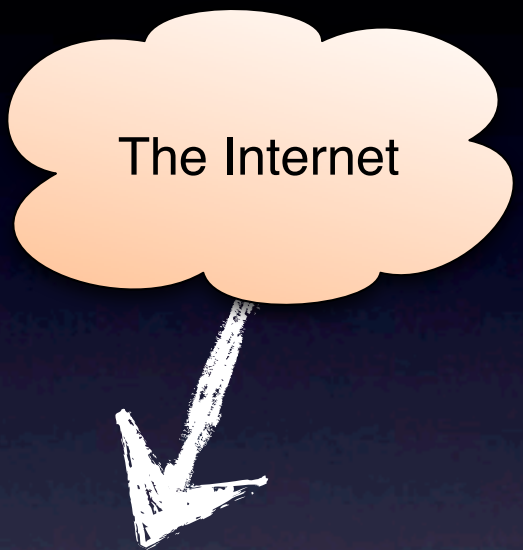
- Bots are *stateless*
  - List of recipients downloaded from C&C server is randomly chosen
  - Bots can be periodically restarted to quickly obtain information on ongoing spam campaigns
  - Some bots are buggy
- C&C servers change *infrequently*
- Some botnets are *partitioned*

Correlation Analysis

- Combine our sources of data:
  - *Outgoing spam* from BotLab
  - *Incoming spam* at UW

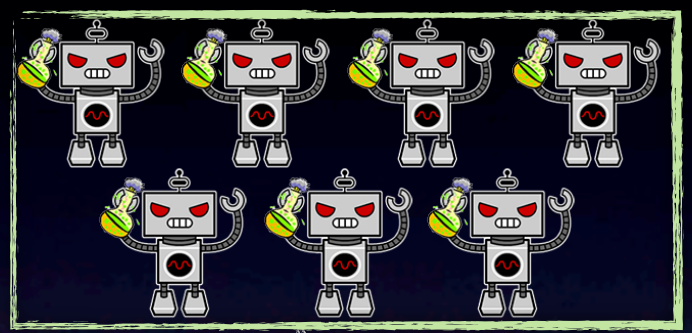
Combining our spam sources

- Incoming spam provides a different perspective
- Spam is received from almost every bot out in the world
- Local view of spam produced
- Global view of spam producers



2.5 million emails per day

Combining our spam sources



BotLab

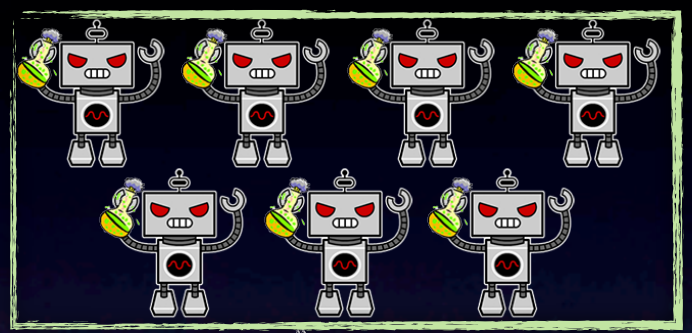
The Internet

Global view of spam produced

+

Global view of spam producers

Combining our spam sources



BotLab

The Internet

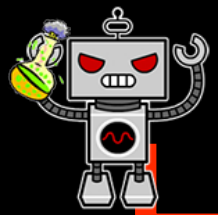
Global view of spam produced + Global view of spam producers

Challenge: create mapping between incoming spam and bot generated spam

Combining our spam sources

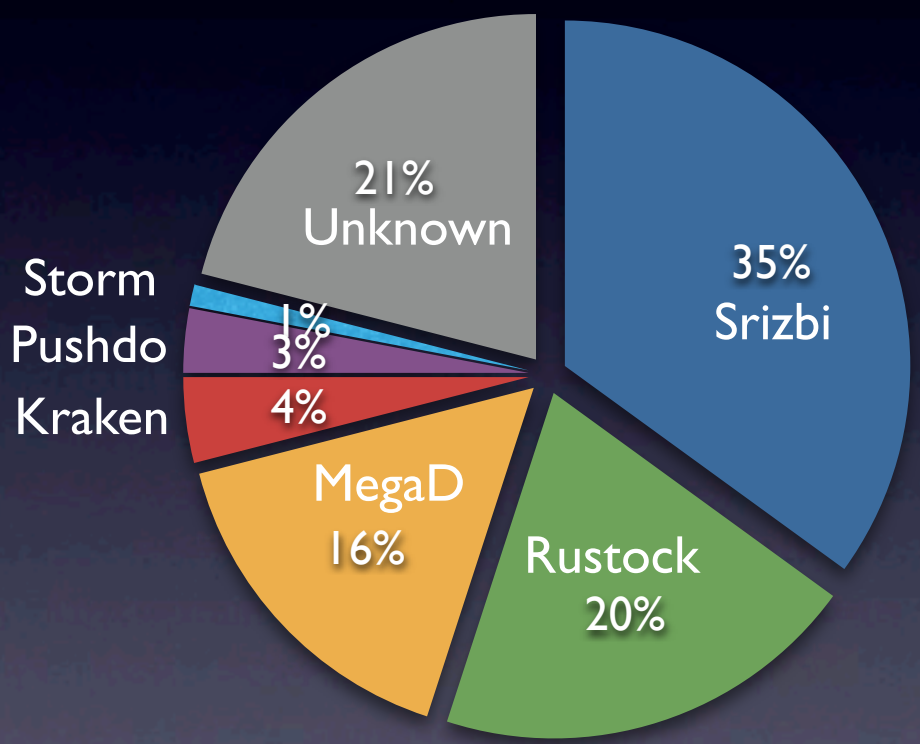
Combining our spam sources

- Observation:
  - Spam subjects are carefully chosen
  - NO overlap in subjects sent by different botnets (489 subjects/day per botnet)
- Solution: Use subjects to attribute spam to particular botnets

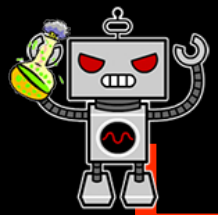


Who is sending all the spam?

The Internet

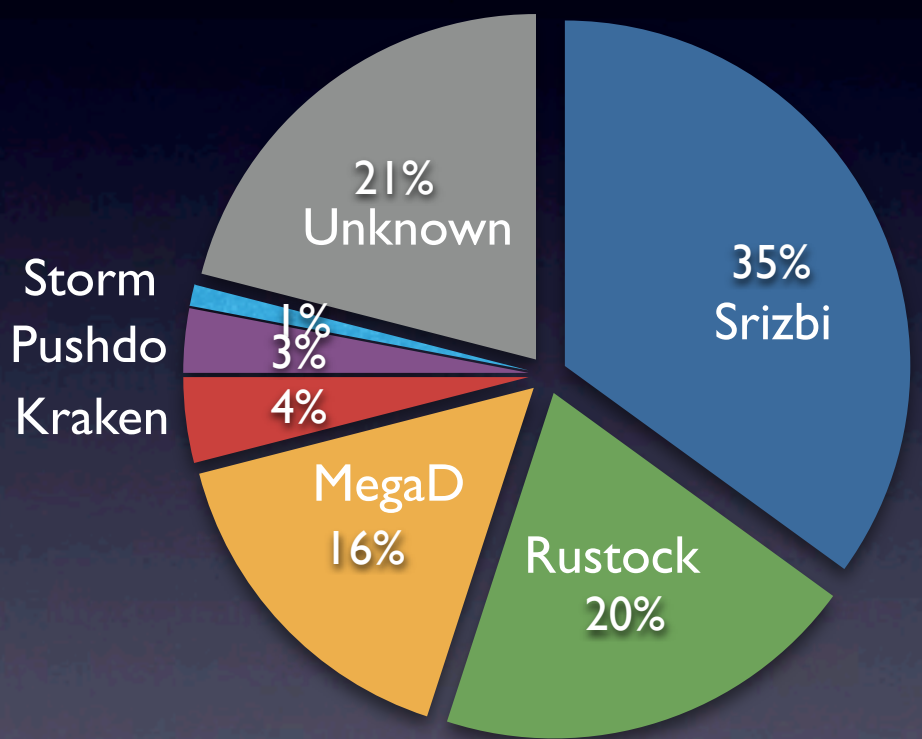


Average over 50 days



Who is sending all the spam?

The Internet



79% of the spam came from just 6 botnets!

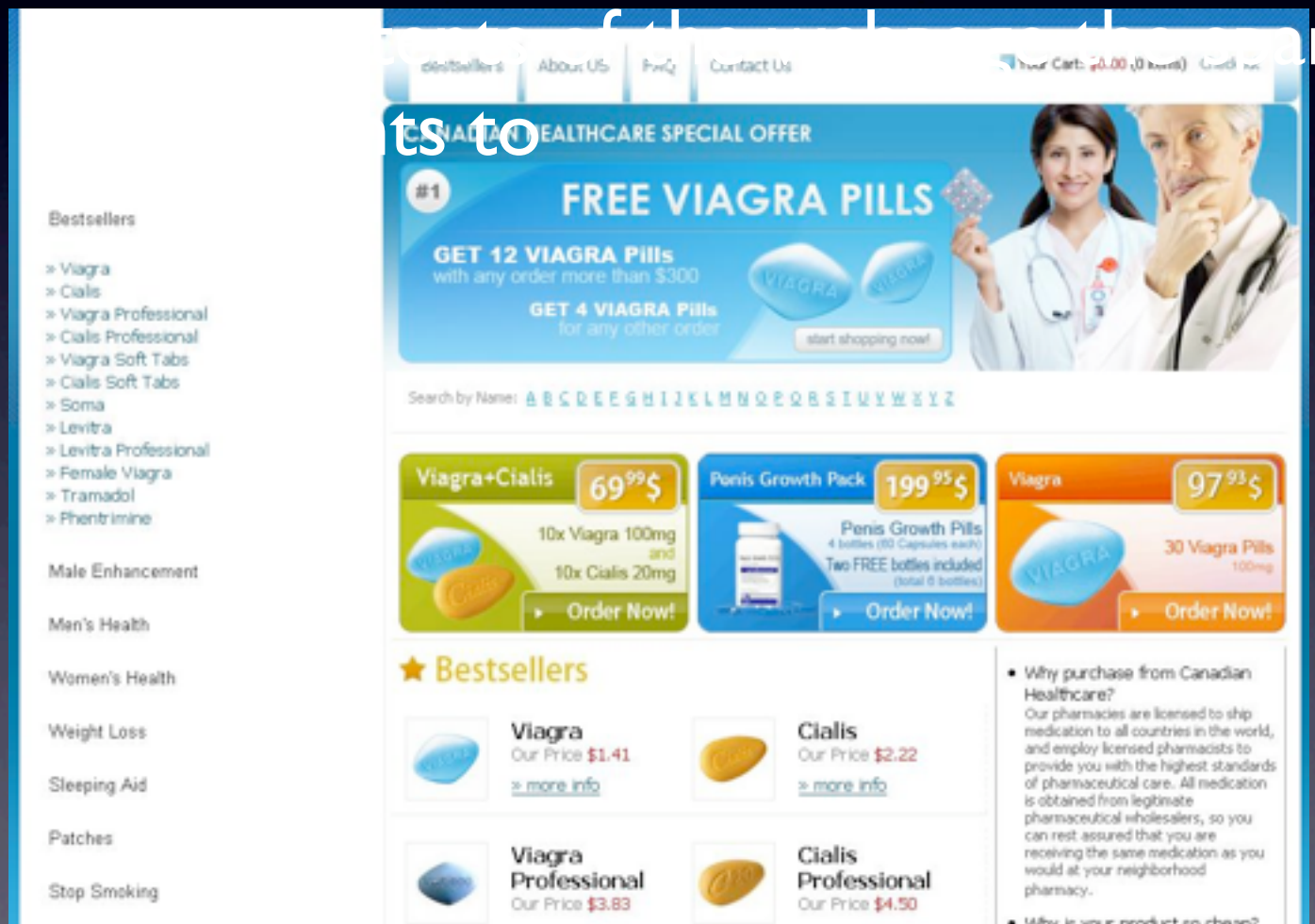
Average over 50 days

Botnets and spam campaigns

- We define a spam campaign by the contents of the webpage the spam URL points to

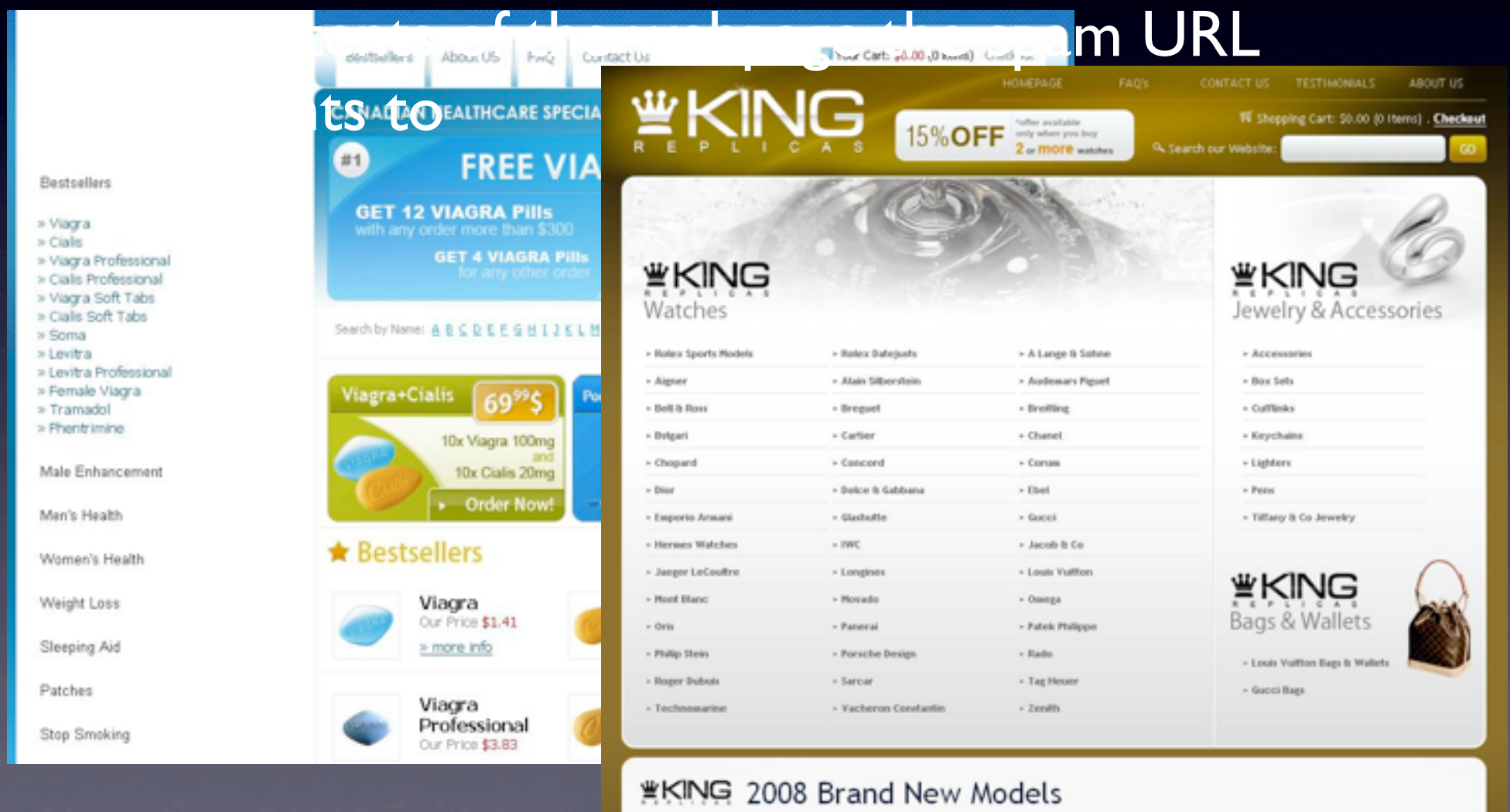
Botnets and spam campaigns

- We define a spam campaign by the [illegible] URL [illegible]ts to



Botnets and spam campaigns

- We define a spam campaign by the [illegible] URL [illegible] to

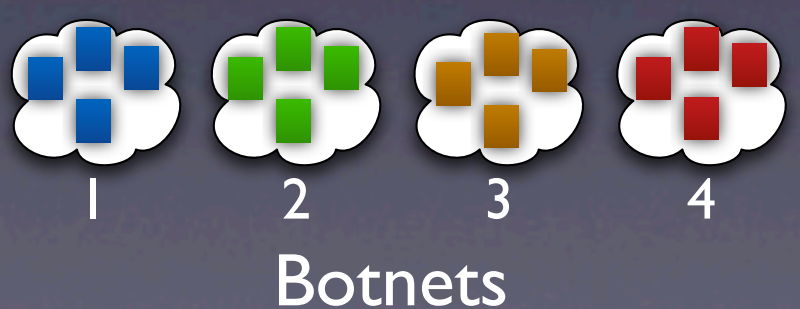


Botnets and spam campaigns

- We define a spam campaign by the contents of the webpage the spam URL points to
- We found the mapping between botnets and spam campaigns to be many-to-many

Where are campaigns hosted?

- How does the Web hosting infrastructure relate to the botnets?



Where are campaigns hosted?

- How does the Web hosting infrastructure relate to the botnets?

Where are campaigns hosted?

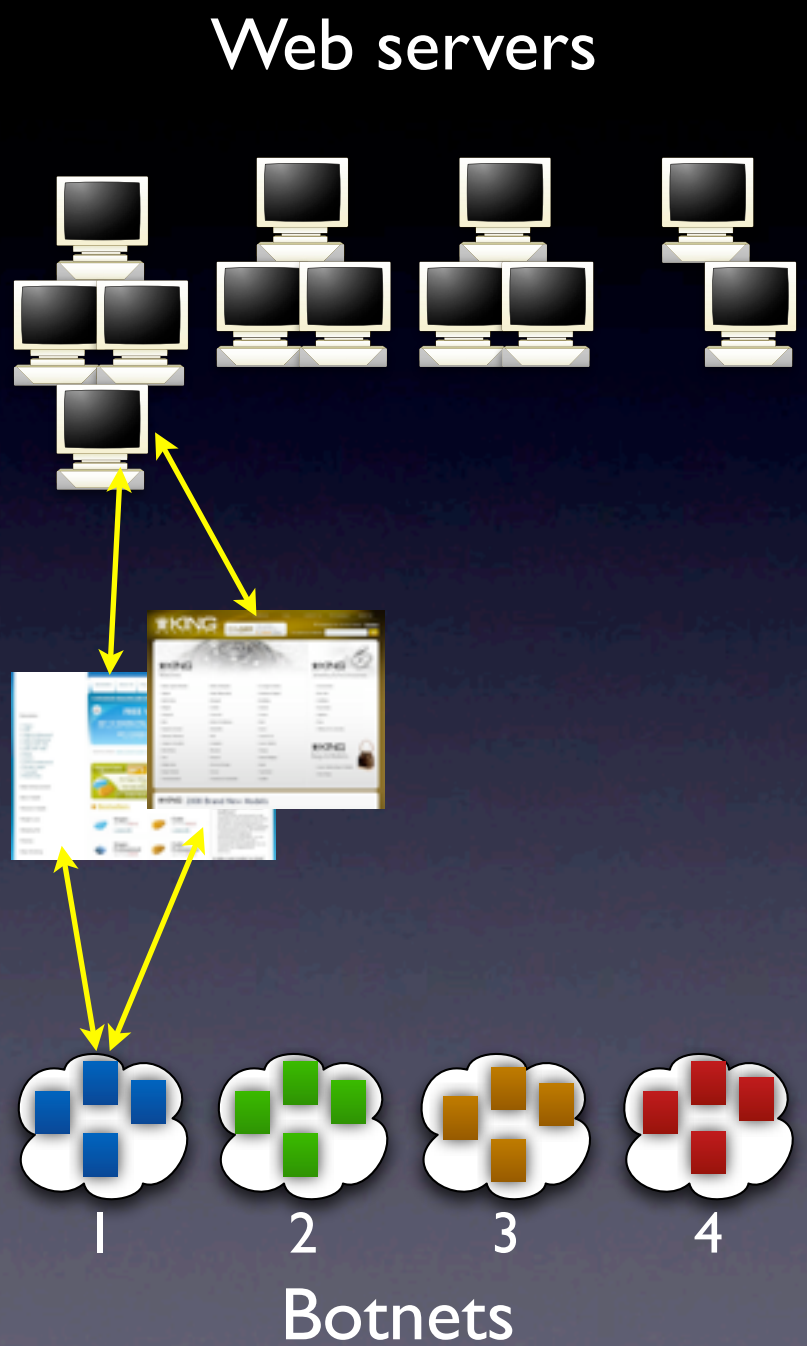
- How does the Web hosting infrastructure relate to the botnets?

Where are campaigns hosted?

- How does the Web hosting infrastructure relate to the botnets?
- Does all spam sent from one botnet point to a single set of web servers?

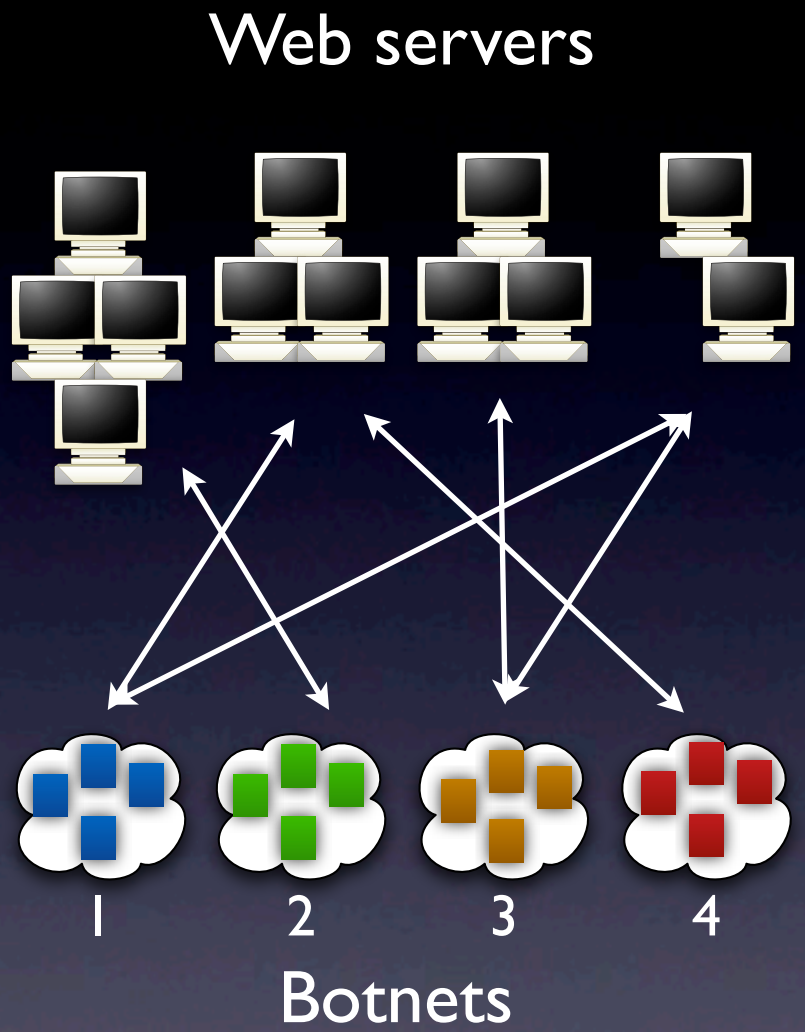
Where are campaigns hosted?

- How does the Web hosting infrastructure relate to the botnets?
- Does all spam sent from one botnet point to a single set of web servers?



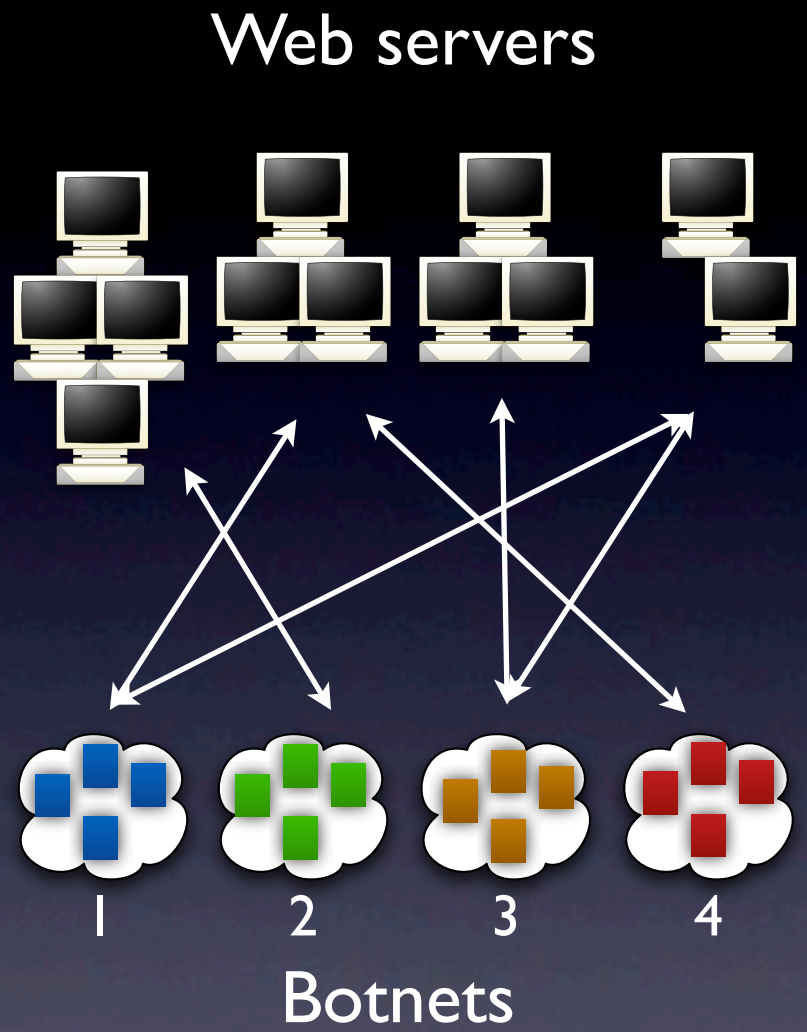
Where are campaigns hosted?

- How does the Web hosting infrastructure relate to the botnets?
- Our data shows a many-to-many mapping
- Suggests *hosting spam campaigns is a 3rd party service* and not tied to botnets



Where are campaigns hosted?

- How does the Web hosting infrastructure relate to the botnets?
- Our data shows a many-to-many mapping
- Suggests *hosting spam campaigns is a 3rd party service* and not tied to botnets



- 80% of spam points to just 57 Web server IPs

Botnet Membership

- What fraction of the botnet members can we identify in a single day at a given location?
- Again use probabilistic analysis based on the random recipient address model
  - Let *P* is the probability that a given spam message is sent to an UW email address
  - Let *N* be the number of email messages sent by a bot over a given period
  - Then probability of UW receiving a spam message:

$$1 - e^{-N*P}$$

Botnet Membership

- Even the most gentle bots send *N = 48K* messages per day
- UW receives 2.4M messages of a total world-wide estimate of 110B messages; *$P = 2.2*10^{-5}$*
- Over a 24-hour uptime, probability of identifying a botnet participant is *0.65*

Conclusions

- BotLab is an *engineering exercise* that pulls together many of the ideas proposed earlier
- Key components: *active crawling, executing captive bots, network fingerprinting, correlation*
- Enables a rich set of measurements. Results include:
  - Small number of botnets generate most of the spam
  - Complex (not one-to-one) relationships between botnets, spam campaigns, and hosting infrastructures
- BotLab also promises better defenses (safe browsing, spam filtering, bot detection, etc.)

Conclusion

- Botnets pose serious security challenges
- Requires greater understanding
- BotLab is an *engineering exercise* that pulls together many of the ideas proposed earlier
- Key components: *active crawling, executing captive bots, network fingerprinting, correlation*
- Potentially enables better defenses (safe browsing, spam filtering, bot detection, etc.)

- More questions? Just toss me an email (arvind@cs) or stop by my office (CSE 544).