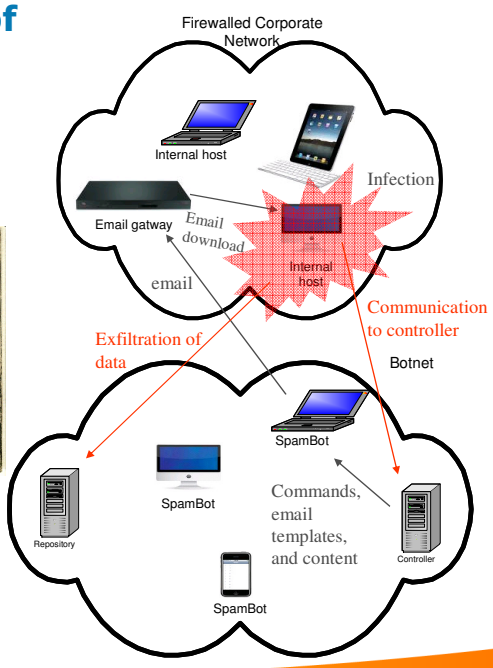
Detection of spam hosts and spam bots using network traffic modeling

Anestis Karasaridis
Willa K. Ehrlich, Danielle Liu, David Hoeflin

The evolution of malevolence

From ancient mythology to now

Heracles (Hercules), with the help of Iolaos, fights the **Lernaean Hydra** in one of his 12 Labours



Overview

- Unsolicited email (UE, a.k.a. "spam") is not only a nuisance but has become one of the main infection vectors to propagate malware.
- We observe innovative techniques to get UE through anti-spam filters, fighting the same techniques used to detect them.
- We have seen it being sent to targeted individuals ("spear phishing"), customized to take advantage of some knowledge of the recipient.
- Email can carry links or attachments for malware that are not suspect at a first glance taking advantage of weakness in many common applications such as browsers (IE, Firefox), Acrobat Reader, Javascript, AV software itself, etc.
- Once the end-user device is infected, it typically joins a botnet for remote control of its future activities (such as sending spam, launching DoS attacks, leaking private data etc.)
- We offer an end-to-end detection solution: spam hosts -> compromised spam hosts -> spam hosts that join a botnet (spam bots)

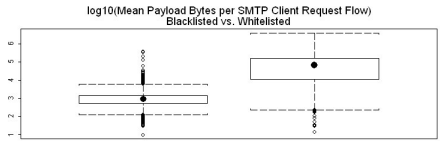
Spam host detection (data and main observations)

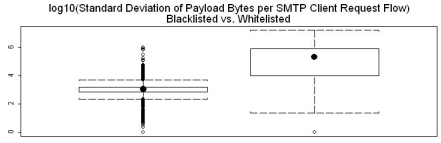
•Most current spam detection depends on full or partial text (including headers): Privacy intrusive, does not scale well and requires frequent retraining and rule updates.

•Our data consist of flow records from peering links: allow minimal privacy invasion and scalable for very large networks. Also gives us a broader view of the host activities

•One important observation: spam has markedly different statistical properties than ham (the opposite of spam)

- Mean and variance of bytes per flow for Simple Mail Transfer Protocol (SMTP) flows for spam are lower than ham





Multivariate traffic models of email spammers vs. legitimate SMTP clients

- Traffic model for a given SMTP client class is a bivariate Gaussian distribution with the two RVs being:
  - X1 : mean bytes per SMTP request flow (BPF) for a given host across its flows
  - X2: standard deviation of bytes per SMTP request flow for a given host across its flows

Traffic Model Parameters by Client Class

Parameter Notation	Parameter Interpretation
$\mu_{X_{1j}}, j = 1, 2$	$\mu_{X_{1j}} = E_j[\log Y_{1ij}]$, Y_{1ij}: mean (across flows) of BPF for client i in class j
$\mu_{X_{2j}}, j = 1, 2$	$\mu_{X_{2j}} = E_j[\log Y_{2ij}]$, Y_{2ij}: stddev (across flows) of BPF for client i in class j
$\mathrm{Var}(X_{1j})$	$\sigma^2_{X_{1j}} = E_j(\log Y_{1ij} - \mu_{X_{1j}})^2$
$\mathrm{Var}(X_{2j})$	$\sigma^2_{X_{2j}} = E_j(\log Y_{2ij} - \mu_{X_{2j}})^2$
$\mathrm{Cov}(X_{1j}X_{2j})$	$E_j(\log Y_{1ij} - \mu_{X_{1j}})(\log Y_{2ij} - \mu_{X_{2j}})$

Application of Bayesian theory to SMTP client classification

- Consider a traffic vector, **x**, for an (unknown) SMTP client *i* consisting of: $X1_i$=log10(mean BPF) and $X2_i$=log10(stddev BPF).
- Classes c_S and c_L for spam and legitimate SMTP hosts, respectively
- From Bayes Theorem

  where $P(C(x) = c_j) = P(c_j / x) = P(c_j) * P(x / c_j) / P(x)$

  $P(c_j)$ is the probability of class j independently of the observed data

  $P(\mathbf{x}/c_j)$ is the conditional probability of the traffic vector **x** given it is in class j based on the bivariate normal density function
- Since denominator does not depend on a category, classify an SMTP client as an spammer whenever

where T>0.5 $$P(C(x)=c_S)=\frac{P(c_S)*P(x/c_S)}{P(c_S)*P(x/c_S)+P(c_L)*P(x/c_L)}>T,$$

- If we assign equal probability to the two classes (i.e., $P(c_S)=P(c_L)$) then $$P(C(x)=c_S)=P(x/c_S)/((P(x/c_S)+P(x/c_L))>T$$
- By varying T, tradeoff
  - Detection probability (i.e., correctly classifying true Spammer as Spammer)
  - Probability of false positive (i.e., incorrectly classifying legitimate SMTP client as Spammer)

Accuracy of traffic models in classifying blacklisted vs. whitelisted SMTP clients

- Evaluate classification accuracy by using a two groups of known SMTP clients (blacklisted and whitelisted).
- Collected data for a period of 300 hours. For each hour, analyzed set of SMTP flows associated with approximately 2000 known SMTP clients and applied the spam detection algorithm to classify an SMTP client as spammer or legitimate
- Evaluated accuracy of classification
  - Detection: P(Classify spammer/Blacklisted SMTP client)
  - False Positive: P(Classify spammer/Whitelisted SMTP client)
- Results of model validation

T	P(Classify Spammer/ Blacklisted)	P(Classify Legitimate/ Blacklisted)	P(Classify Legitimate/ Whitelisted)	P(Classify Spammer/ Whitelisted)
0.8	0.887	0.031	0.864	0.05
0.85	0.862	0.027	0.852	0.042
0.9	0.817	0.023	0.836	0.032
0.95	0.708	0.018	0.808	0.018

Detection of compromised spam hosts using host traffic profiling

- Analyze remaining (other than SMTP) flows of detected spam hosts to find significant local and remote ports
- Significant is determined using relative uncertainty (normalized entropy) of a remote port given a remote host or a local port (local is our spam host)
- Establish the host traffic profile of a whitelisted (normal) SMTP host, which shows mostly significant 25/tcp (SMTP) and 53/udp (DNS) remote ports.
- Identify spam hosts that have abnormal host traffic profile (i.e., they exhibit significant ports other than remote 25/tcp and 53/udp), indicating compromise.

Example A: HTP of whitelisted SMTP client

Traffic Component	Entropy of Local Ports	Significant Local Ports	Entropy of Remote Hosts	Interpretation
Remote Port 25/tcp	0.972	N/A	0.622	e-mail client
Remote Port 53/udp	0	UDP-53	0.788	Host Interacting via local port 53/udp with remote port 53/udp

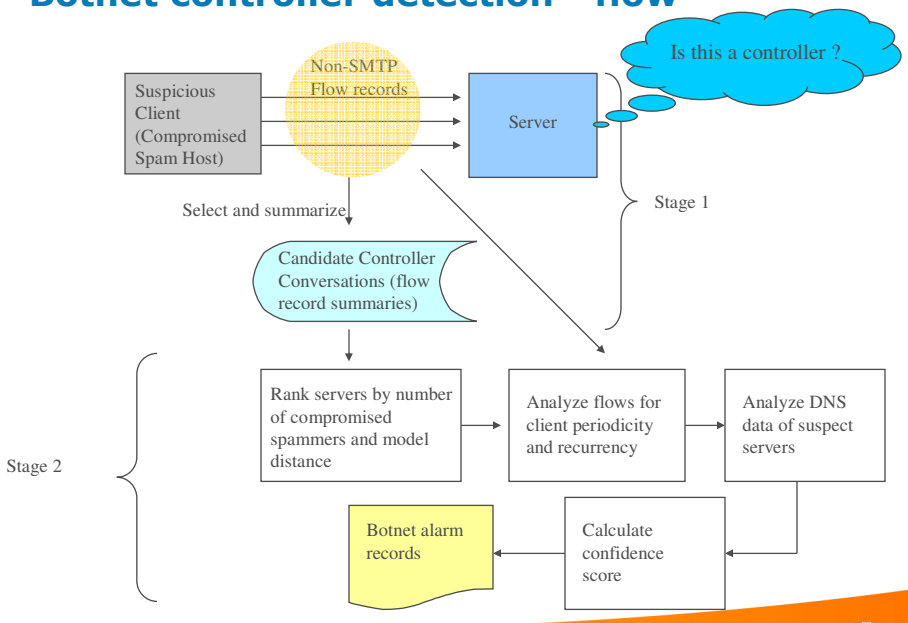
Example B: HTP of a likely compromised SMTP client

Traffic Component	Entropy of Local Ports	Significant Local Ports	Entropy of Remote Hosts	Significant Remote Hosts	Traffic Statistics (Outbound flows)	Interpretation
Remote port tcp/25	0.952	N/A	0.906	N/A	$N_{flows} = 149$ $\overline{bpf} = 986.2$ $Std(bpf) = 538.2$ $Cv(bpf) = 0.546$	SMTP Client
Remote port udp/11526	0	udp/63301	0	10.20.30.51	$N_{flows} = 325$ $\overline{bpf} = 53$ $Std(bpf) = 0$ $Cv(bpf) = 0$	Host performing P2P with 10.20.30.51 on remote port udp/11526 with local port udp/63301
Remote port udp/11543	0	udp/63301	0	10.20.30.52	$N_{flows} = 325$ $\overline{bpf} = 53$ $Std(bpf) = 0$ $Cv(bpf) = 0$	Host performing P2P with 10.20.30.52 on remote port udp/11543 with local port udp/63301

Botnet controller detection - description

- Two-stage approach:
  - Stage 1: Summarize flow records of likely compromised spam hosts, using 3 different approaches: a) connections to typical control ports (e.g., port 80, 8080 for http, 6667 for IRC etc.), b) connections to hubservers (servers/ports with high fan-in, c) connections where the number of flows per address, packets per flow, and bytes per packet are within 90%-th percentile range of the model. The records constitute the candidate controller conversations (CCC)
  - Stage 2: Aggregate CCC records, rank by number of clients to a server/port, and calculate distance to common control models and find clients with periodic or recurrent behavior. Use DNS passive replication database to identify servers with no DNS domain or with transient domains. Assign confidence score.

Botnet controller detection - flow



Transient domain detection (part of the DNS analysis of suspected controllers)

- In addition to flow records we have access to a DNS passive replication database which aggregates internet-wide DNS resolutions of domains to IP addresses.
- Database facilitates queries for historical data by domain or IP address
- We developed algorithms to detect transient and fast flux domains. Transient domains are domains that hop sequentially between addresses of diverse providers. Fast-flux domains are domains that map to frequently changing sets of addresses.
- This analysis helps us improve our confidence that a particular IP address is a controller.
- Direct access to a suspected server by its IP address or use of transient domains or domains that are used briefly and then discarded, increase the likelihood that a suspected server is compromised

Examples of automated detection of controllers of some of the largest spam botnets (1/2):

```
#Server_IP|Server_Port|Number_of_Suspicious_Clients|
 Port_Euclidean_Distance|Number_of_Suspicious_Periodic_
 Clients|0-EntropyClients|Number_of_domains|Number_of_
 recent_domains|Number_of_Transient_domains|Aggregate_Score
10.232.229.114|80|3|78.80|1|3|4|1|0|115
Date/Hr Timestamp: 2010012504
Recent domain used-selementusaks.org
Total number of domains that have mapped to this IP-4
Number of domains that have recently mapped to this IP-1
Examining domain selementusaks.org
Number of historical IPs-0
Domain has a small number of historical IPs (0 <2)
```

Alarm record for a controller of the Ozdok botnet. One suspicious client shows periodic and 3 clients show recurrent behavior

```
#Server_IP|Server_Port|Number_of_Suspicious_Clients|
 Port_Euclidean_Distance|Number_of_Suspicious_Periodic_
 Clients|0-EntropyClients|Number_of_domains|Number_of_
 recent_domains|Number_of_Transient_domains|Aggregate_Score
10.19.191.55|443|2|36.41|0|1|0|0|0|85
Date/Hr Timestamp: 2009120701
Total number of domains that have mapped to this IP-0
Number of domains that have recently mapped to this IP-0
There are no domains associated currently with this
 IP address
```

Alarm record for an controller of the Cutwail botnet. The controller is directly accessed by its IP address since there are no records of DNS domains pointing historically to this address

Examples of automated detection of controllers of some of the largest spam botnets (2/2):

```
#Server_IP|Server_Port|Number_of_Suspicious_Clients|
 Port_Euclidean_Distance|Number_of_Suspicious_Periodic_
 Clients|0-EntropyClients|Number_of_domains|Number_of_
 recent domains|Number of Transient domains|Aggregate Score
10.51.196.242|80|2|45.36|0|0|6064|63|2|110
Date/Hr Timestamp: 2010020304
Recent domain used-lambert.66ghz.com
Recent domain used-zal.te.ua,
...
Examining domain lambert.66ghz.com
Number of historical IPs-0
Domain has a small number of  historical IPs (0 <2)
...
Examining domain zal.te.ua
Number of historical IPs-2
Average Time Overlap between IPs-0.000
Average IP distance-1.000
Domain zal.te.ua appears to be transient
#Domain|IPaddress|NumResponses|NumDomainstoIP|
 StartTime|EndTime|Lifespan(Days)
10.51.196.242|4|6064|20091112@23:25:21|20100129@23:43:01|78
10.43.65.6|5|307|20090113@15:24:39|20090510@15:41:05|117
10.32.73.138|4|233|20081027@07:41:07|20081127@15:53:43|31.3
```

Alarm record for a controller of the Zeus botnet. Two DNS domains linked to the suspected controller appear to be Transient

Conclusions and Future work

- Developed a new end-to-end approach and tools to passively detect spam botnets and their respective controllers using flow data and DNS metadata.
  - Spam host detection is based on bytes per flow statistics of flow data. Classification using a Bayesian approach over a bivariate Gaussian traffic model.
  - Suspect spam bots are spam hosts that have a host traffic profile different than a whitelisted spam host
  - Botnet controller detection is based on a two stage algorithm that uses flow summaries and distances to common control protocol models. Heuristics such as periodicity, recurrency and DNS transiency provide an overall confidence score for a suspected controller.
- Continue to refine models and heuristics to avoid likely false positives mostly related to bot contacts to CDNs supported sites