Spamming Botnets: Signatures and Characteristics

Xie et al.

Presented by

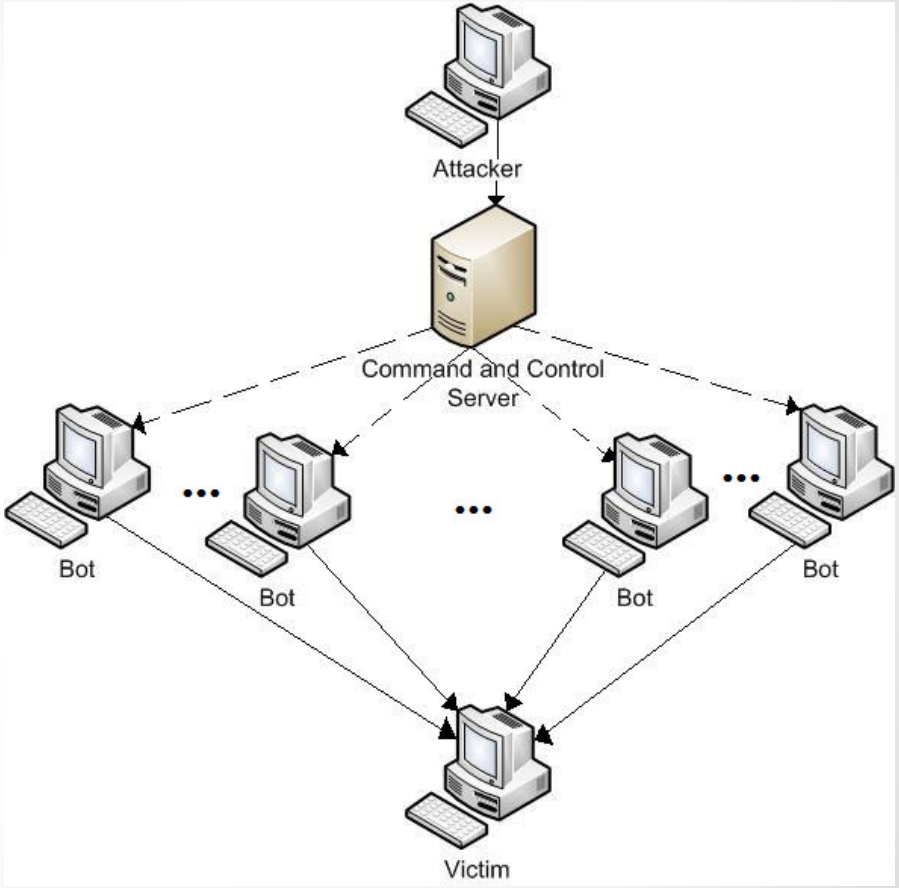
Kyle Martin

<kyle.martin@knights.ucf.edu>

UNIVERSITY OF CENTRAL FLORIDA

The Problems

- Botnets



The Problems

- Botnets
- Spam

The Problems

- Botnets
- Spam
- Botnets + Spam

AutoRE

- Generates botnet spam signatures
- No labeled data or external sources required
- Regular Expressions for embedded URL
- Organizes spam into groups by campaign
- Identify characteristics of spam botnets

URLs and Spam

Time	URLs	Source ASes	URLs
2006-11-02	66	38	http://www.lympos.com/n/?167&carthagebolets http://www.lympos.com/n/?167&brokenacclaim http://www.lympos.com/n/?167&acceptoraudience
2006-11-15	72	39	http://shgeep.info/tota/indexx.html?jhjb.cvqxjby,hvx http://shgeep.info/tota/indexx.html?ikjija.cvqxjby,hvx http://shgeep.info/tota/indexx.html?ivvx_ceh.cvqxjby,hvx

- Spam messages can contain multiple URLs
- Generic URLs usually included
- Polymorphic URLs

Email 1

http://www.shopping.com
http://www.w3.org/wai
http://www.psc.edu/networking/projects/tcp/
... ...
http://www.dvdfever.co.uk/co1118.shtml
... ...

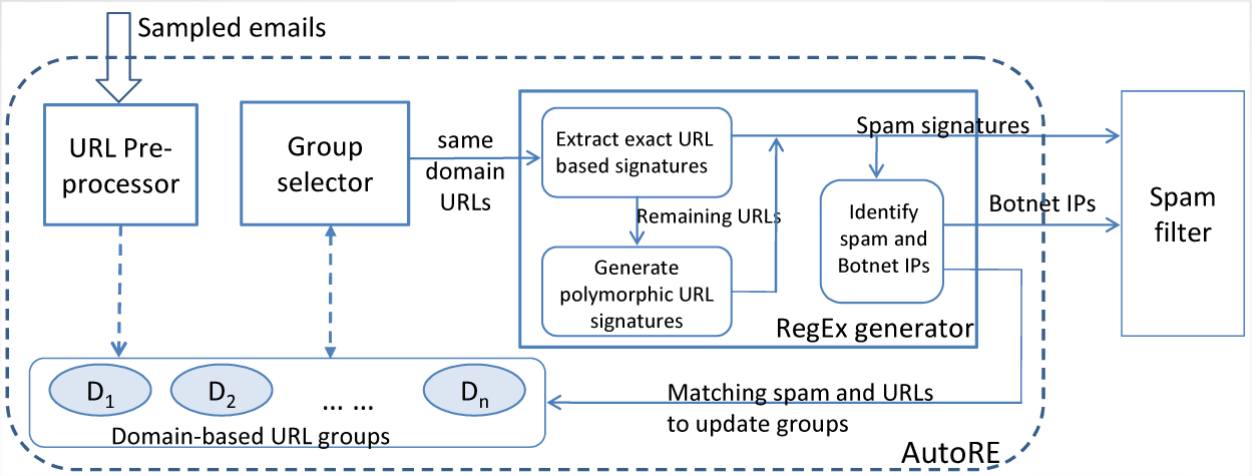
Email 2

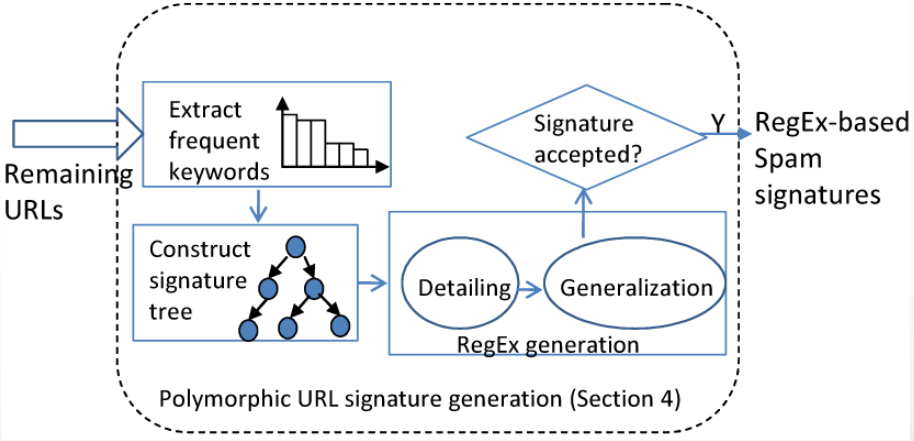
http://www.peacenvironment.net
http://www.w3.org/wai
http://www.bizrate.com
... ...
http://www.dvdfever.co.uk/co1118.shtml
... ...

Email 3

http://endosmosis.com/
http://www.talkway.com
http://www.bizrate.com
... ...
http://www.dvdfever.co.uk/co1118.shtml
... ...

Workflow


Sampled emails
URL Pre-processor
Group selector
same domain URLs
Extract exact URL based signatures
Spam signatures
Remaining URLs
Identify spam and Botnet IPs
Botnet IPs
Spam filter
Generate polymorphic URL signatures
RegEx generator
D1
D2
... ...
Dn
Domain-based URL groups
Matching spam and URLs to update groups
AutoRE

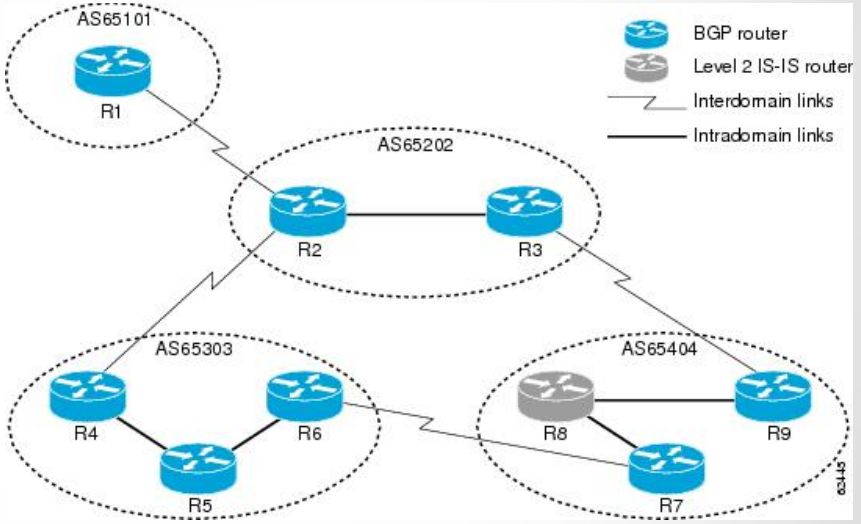

Remaining URLs
Extract frequent keywords
Signature accepted?
Y
RegEx-based Spam signatures
Construct signature tree
Detailing
Generalization
RegEx generation
Polymorphic URL signature generation (Section 4)

Grouping URLs

- URLs grouped by domain
- Select groups characterizing a campaign
  - Temporal correlation
  - Distinct IPs active in a span of time
  - Sharp spikes indicate strong correlation

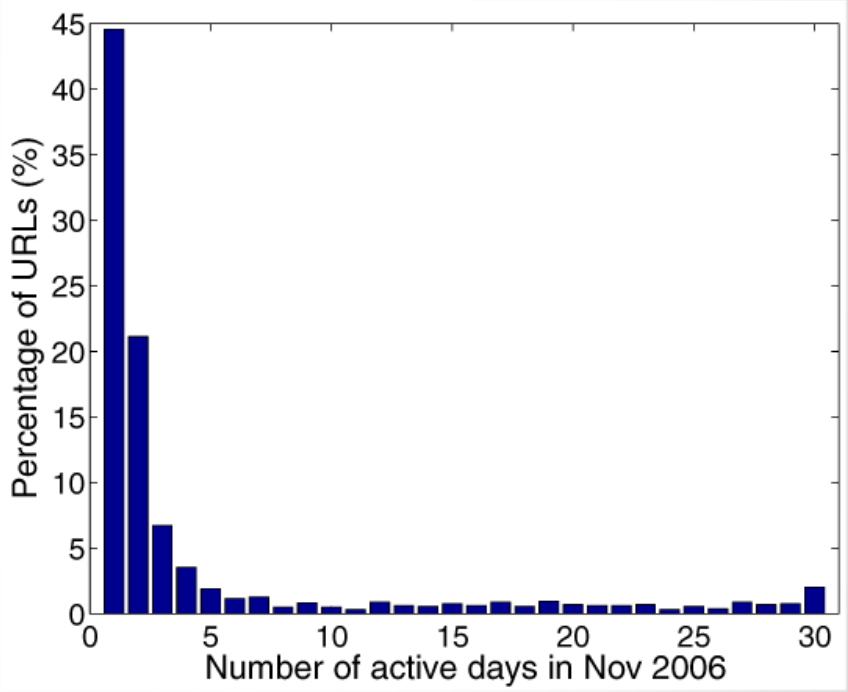
Generating Signatures

- Distributed
  - Number of Autonomous Systems



Generating Signatures

- Distributed
  - Number of Autonomous Systems
- Bursty
  - Long-term campaign durations (over days)



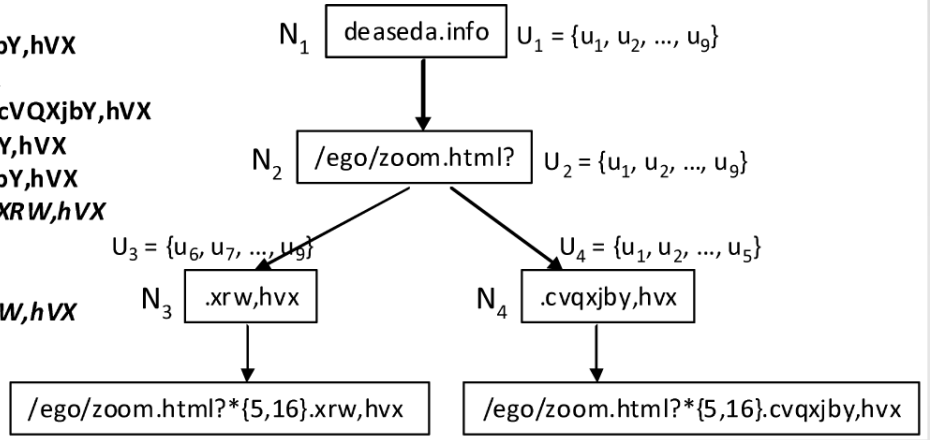
Generating Signatures

- Distributed
  - Number of Autonomous Systems
- Bursty
  - Long-term campaign durations (over days)
- Specificity
  - Probability of a random URL matching

$$P(e) = \frac{2^{B_e(u)}}{2^{B(u)}} = \frac{1}{2^{B(u)-B_e(u)}} = \frac{1}{2^{d(e)}}$$

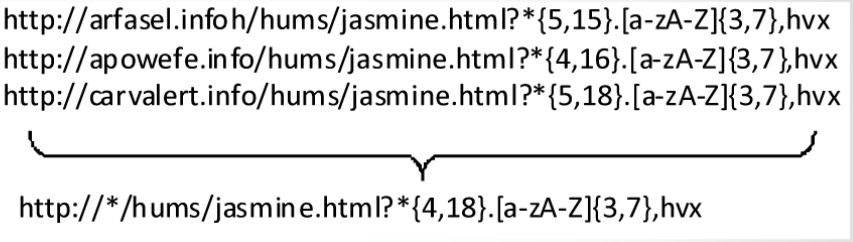
Regular Expression Generation

u_1: http://**deaseda.info**/ ego/zoom.html?QjQRP_xbZf**.cVQXjbY,hVX**
u_2: http:// **deaseda.info**/ ego/zoom.html? giAfS**.cVQXjbY,hVX**
u_3: http://**deaseda.info**/ ego/zoom.html? RQbWfeVYZfWifSd**.cVQXjbY,hVX**
u_4: http://**deaseda.info**/ ego/zoom.html? UbSjWcjHC**.cVQXjbY,hVX**
u_5: http://**deaseda.info**/ ego/zoom.html? VPS_eYVNfS**.cVQXjbY,hVX**
u_6: http://**deaseda.info**/ ego/zoom.html? QNVRcjgVNSbgfSR***.XRW,hVX***
u_7: http://**deaseda.info**/ ego/zoom.html? afRZXQ***.XRW,hVX***
u_8: http://**deaseda.info**/ ego/zoom.html? YcGGA***.XRW,hVX***
u_9: http://**deaseda.info**/ ego/zoom.html? aeSfLWVYgRIBH***.XRW,hVX***



- Signature Trees
- Detailing

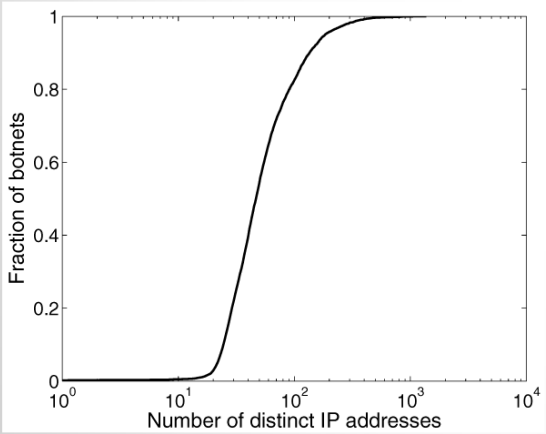
Regular Expression Generation

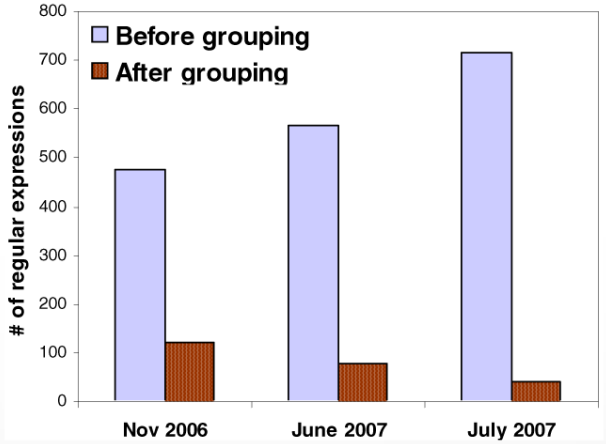


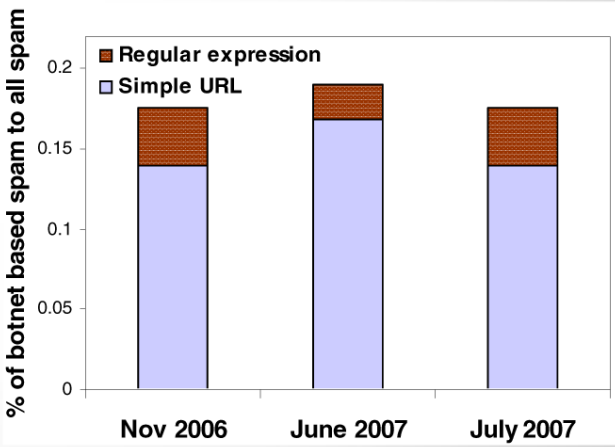
- Signature Trees
- Detailing
- Generalizing
- Quality Evaluation

Results

Month	Nov 2006		June 2007		July 2007		Total
	CU	RE	CU	RE	CU	RE	
Num. of spam campaigns	1,229	519	1835	591	2826	721	7,721
Num. of ASes	3,176	1,398	4,495	1,906	4,141	1,841	5,916
Num. of botnet IPs	88,243	23,316	113,794	19,798	85,036	29,463	340,050
Num. of spam emails	118,613	26,897	208,048	26,637	159,494	40,777	580,466
Total botnet IPs	100,293		131,234		113,294		340,050

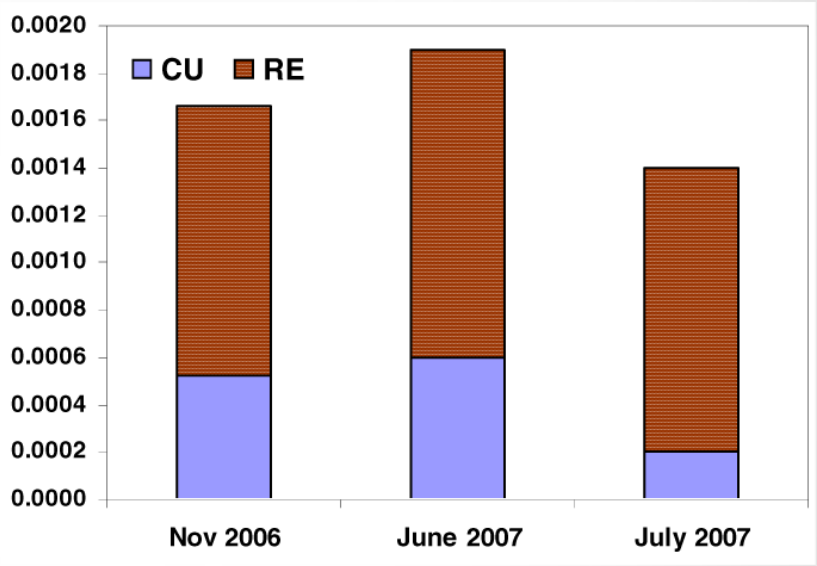






Validation

- Quality of Signatures
  - False positive rate



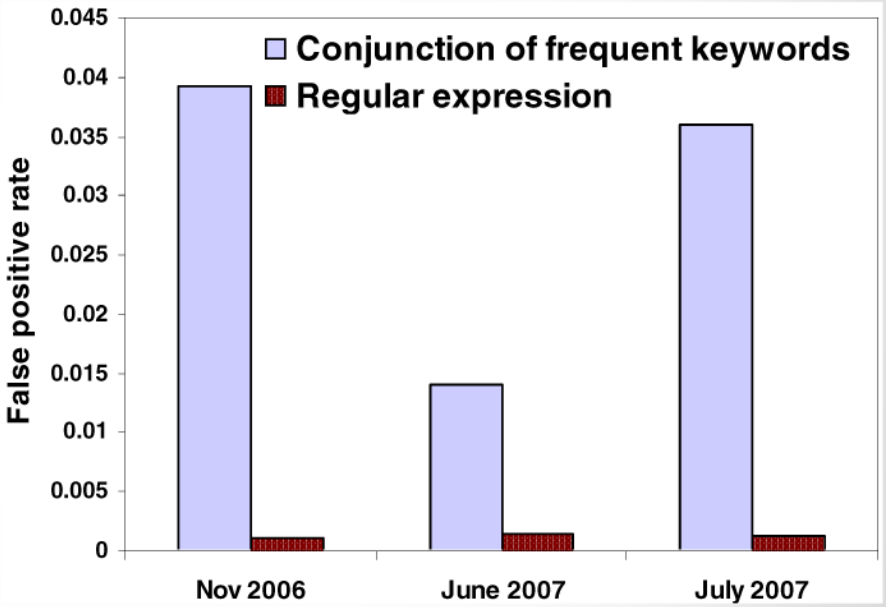
Validation

- Quality of Signatures
  - False positive rate
  - Over time

Month	Nov 2006			June 2007		
	CU	RE	Total	CU	RE	Total
# of spam emails	2	3	5	6,751	43,778	50529
# of non-spam emails	10	0	10	154	561	715

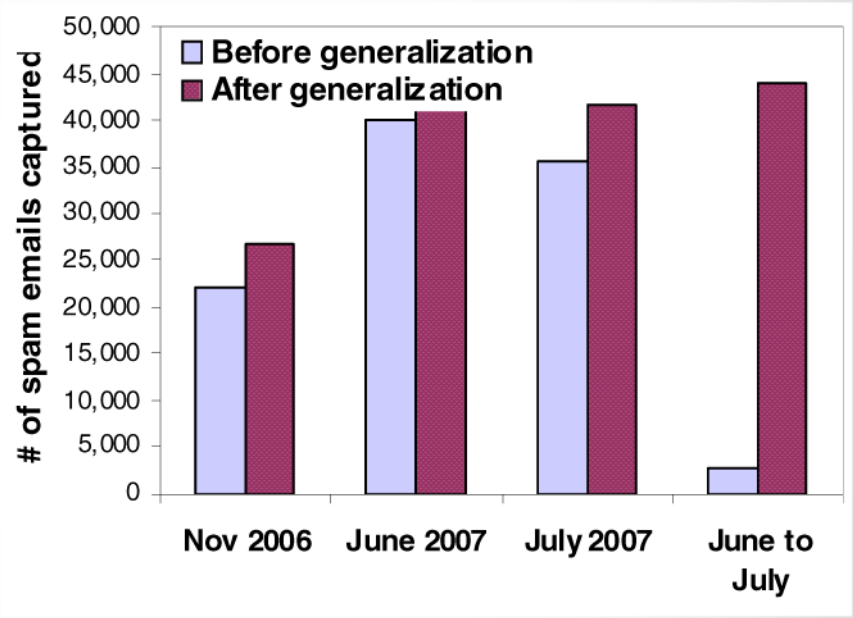
Validation

- Quality of Signatures
  - False positive rate
  - Over time
  - REs vs. Conjugation



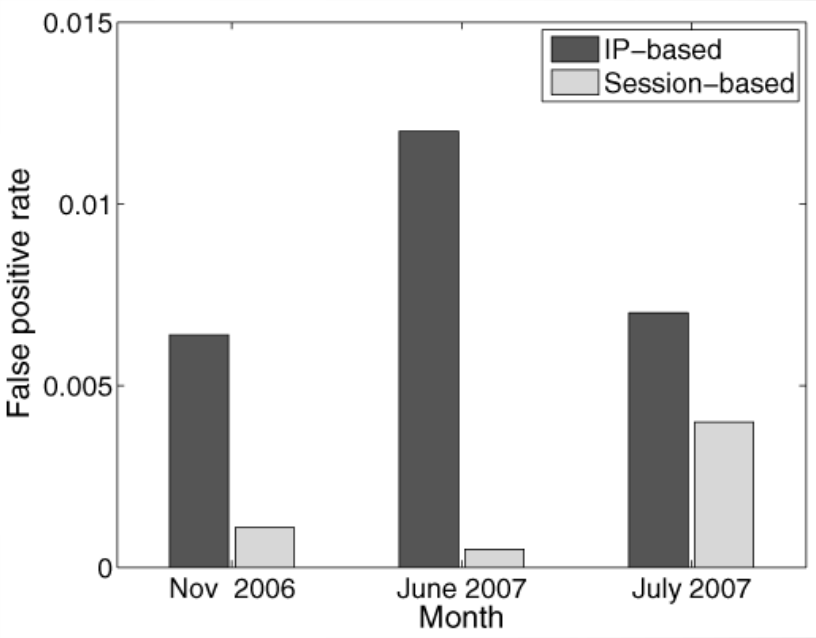
Validation

- Quality of Signatures
  - False positive rate
  - Over time
  - REs vs. Conjugation
  - Effect of generalization



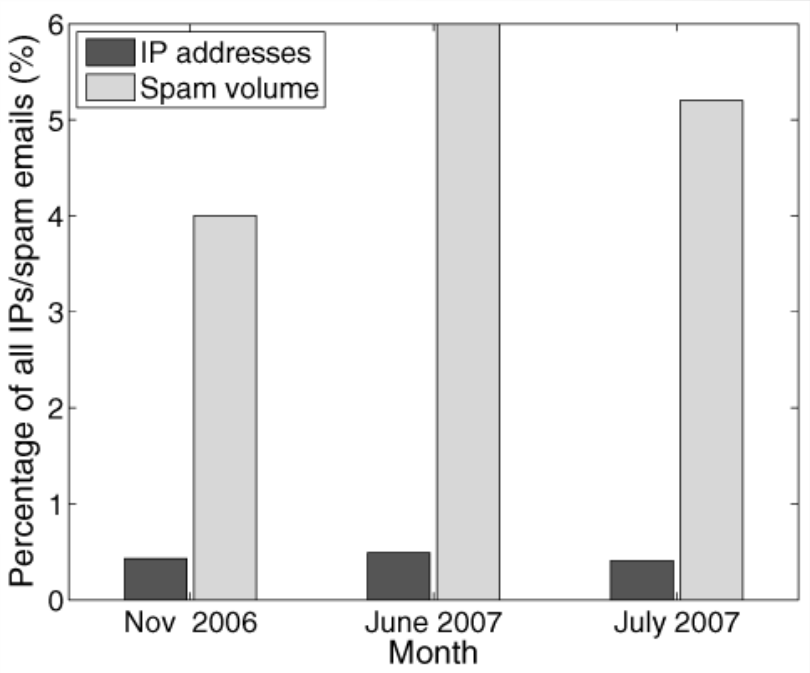
Validation

- Quality of Signatures
  - False positive rate
  - Over time
  - REs vs. Conjugation
  - Effect of generalization
- Host Identification
  - Based on long-term spamming history



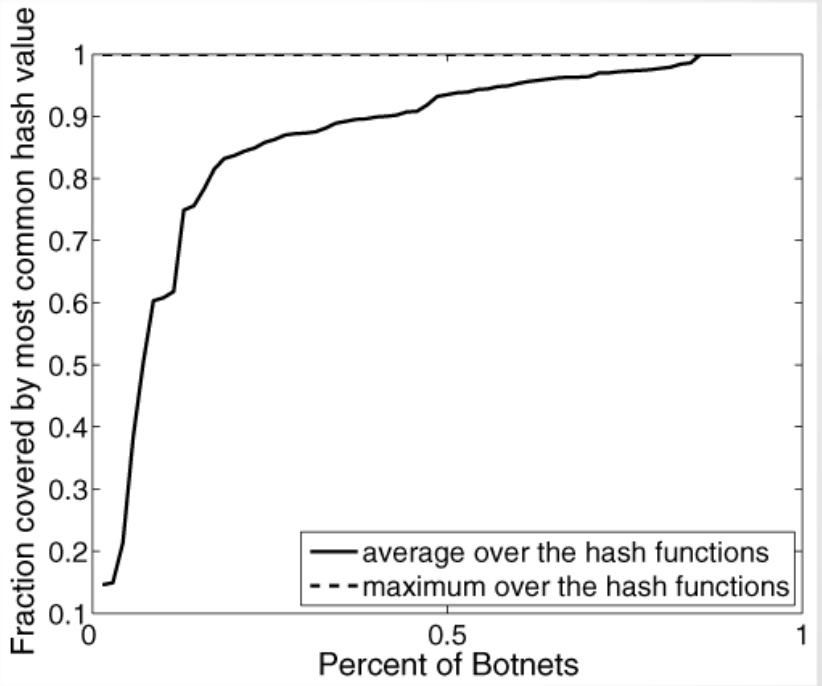
Validation

- Quality of Signatures
  - False positive rate
  - Over time
  - REs vs. Conjugation
  - Effect of generalization
- Host Identification
  - Based on long-term spamming history



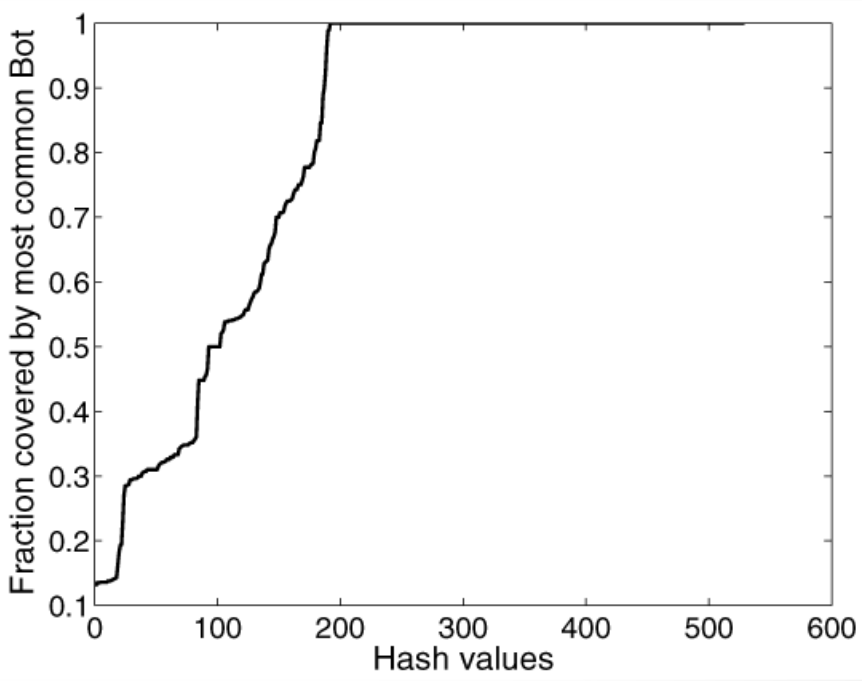
Validation

- Quality of Signatures
  - False positive rate
  - Over time
  - REs vs. Conjugation
  - Effect of generalization
- Host Identification
  - Based on long-term spamming history
- Campaign Identification
  - Based on similarity of URL destinations



Validation

- Quality of Signatures
  - False positive rate
  - Over time
  - REs vs. Conjugation
  - Effect of generalization
- Host Identification
  - Based on long-term spamming history
- Campaign Identification
  - Based on similarity of URL destinations



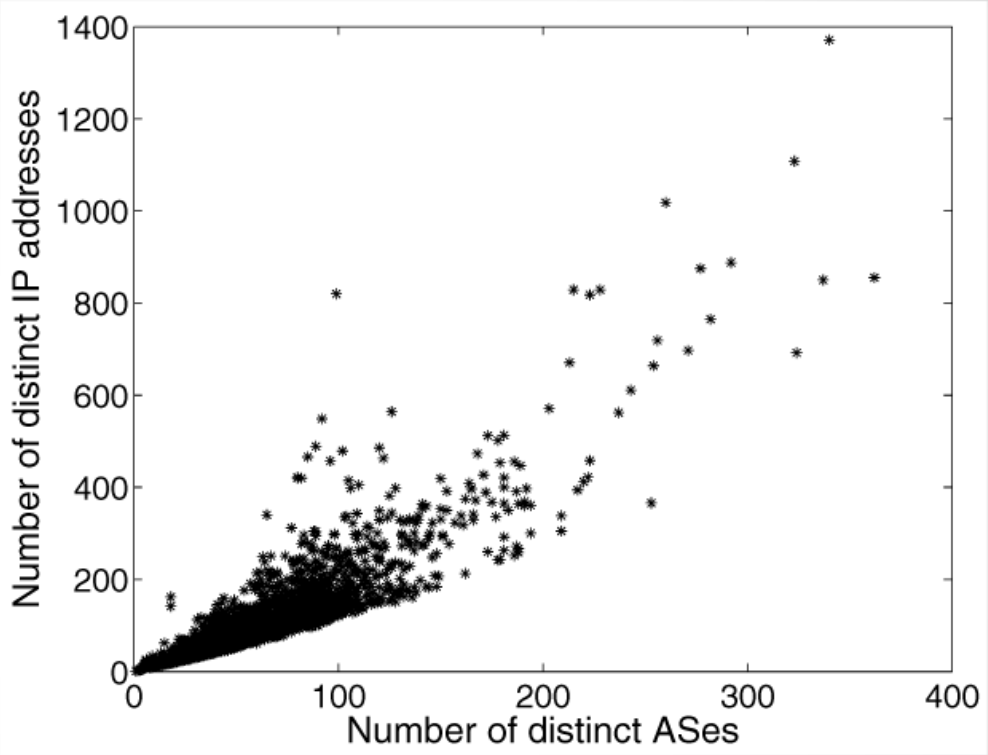
Campaign Characteristics

- IP Distribution

AS description	AS Number	Number of bot IPs
Korea Telecom	4766	15757
Verizon Internet service	19262	11426
France Telecom	3215	11303
China 169-backbone	4837	9960
Chinanet-backbone	4134	8113

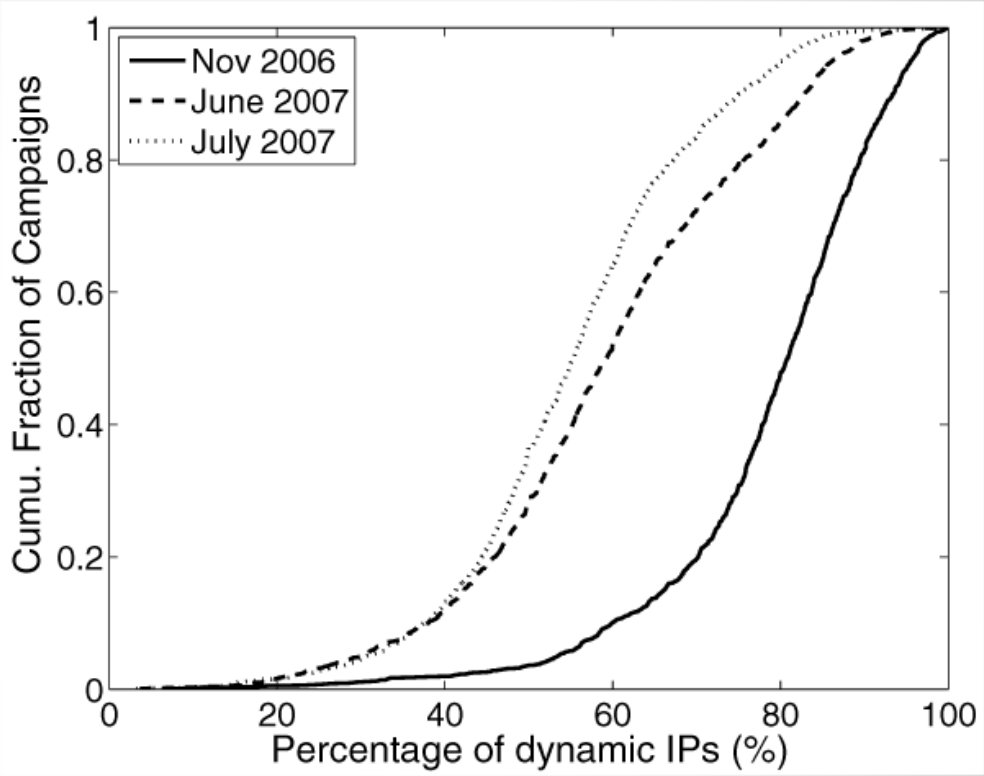
Campaign Characteristics

- IP Distribution



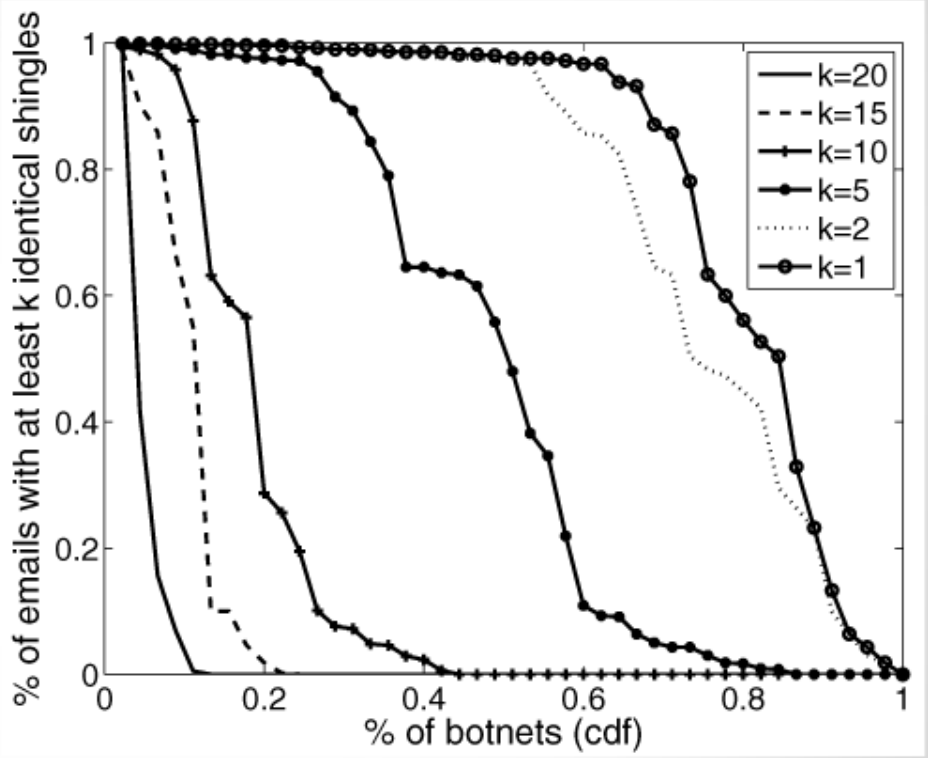
Campaign Characteristics

- IP Distribution



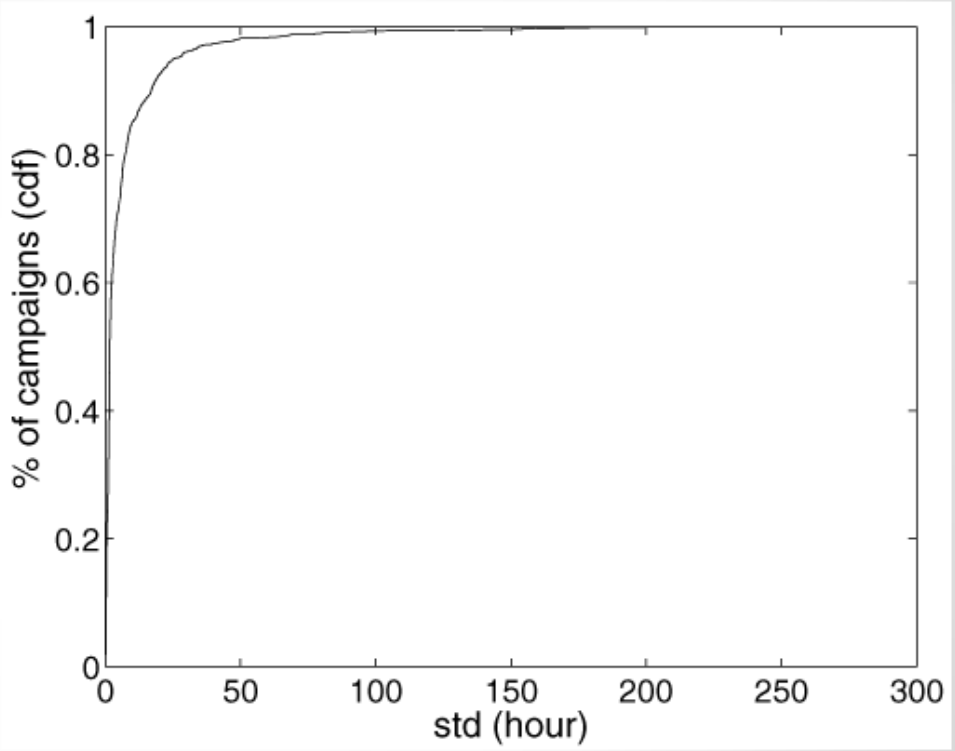
Campaign Characteristics

- IP Distribution
- Per Campaign
  - Similarity of content



Campaign Characteristics

- IP Distribution
- Per Campaign
  - Similarity of content
  - Similarity of time



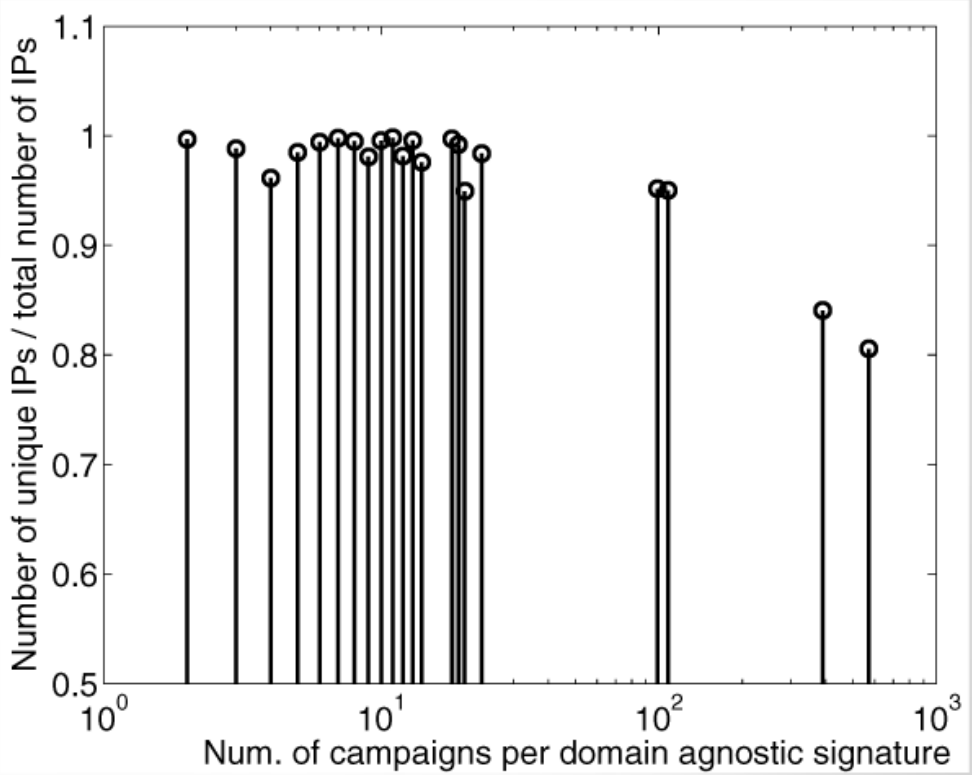
Campaign Characteristics

- IP Distribution
- Per Campaign
  - Similarity of content
  - Similarity of time
  - Similarity of behavior

% of outliers	$< 5\%$	$5 - 10\%$	$10 - 15\%$	$> 15\%$
Nov 2006, CU	59%	27%	8%	6%
Nov 2006, RE	69%	21%	6%	4%
June 2007, CU	74%	23%	2.5%	0.5%
June 2007, RE	44%	42%	9%	5%

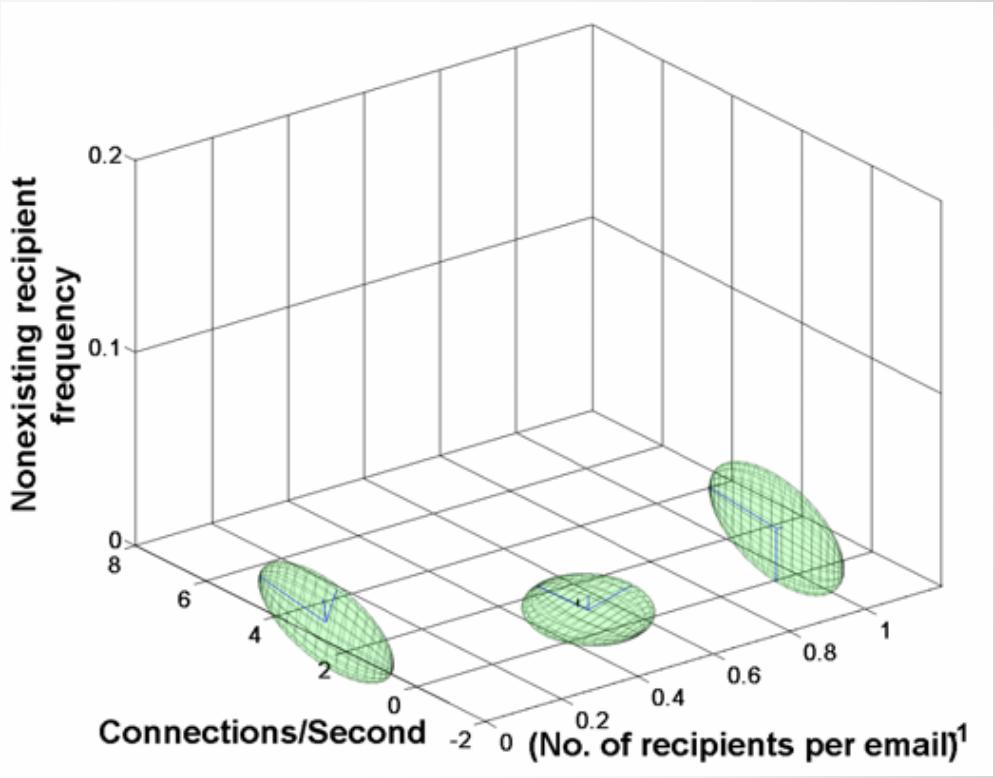
Campaign Characteristics

- IP Distribution
- Per Campaign
  - Similarity of content
  - Similarity of time
  - Similarity of behavior
- Different Campaigns
  - Botnets with similar signatures



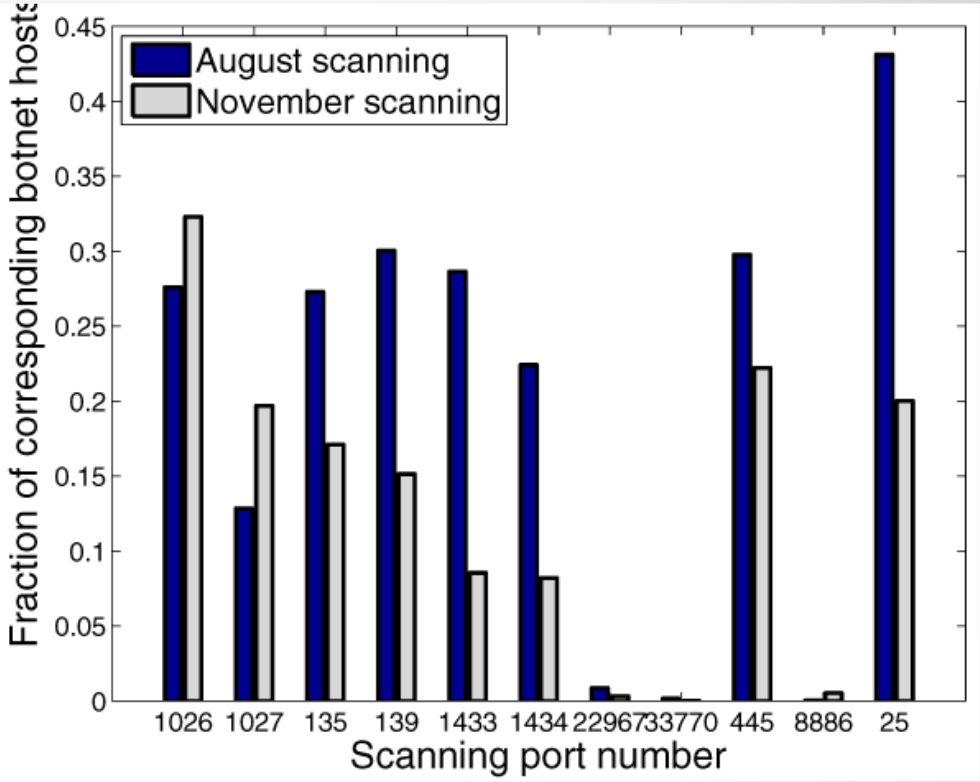
Campaign Characteristics

- IP Distribution
- Per Campaign
  - Similarity of content
  - Similarity of time
  - Similarity of behavior
- Different Campaigns
  - Botnets with similar signatures
  - Sending pattern clusters



Campaign Characteristics

- IP Distribution
- Per Campaign
  - Similarity of content
  - Similarity of time
  - Similarity of behavior
- Different Campaigns
  - Botnets with similar signatures
  - Sending pattern clusters
- Scanning traffic



Strengths

- Generates botnet signatures without labeling
- No external services required
- Signatures have a low false positive rate
- Signatures useful for characterizing botnet spamming behaviors

Weaknesses

- Only based on URLs, not effective against text, images, and other content in spam.
- Only considers spam observed by a single ISP, could be more effective with collaboration.
- Intra-message Polymorphic URLs

Extensions

- Non-botnet RE signatures
- A centralized/distributed signature repository (like Spamhaus)
- Forming RE signatures over non-URL content
- Composite signatures based on multiple content types (URLs, text, images, attachments, etc.)

Questions?

References

- [1]“62445.jpg (JPEG Image, 542 × 330 pixels).” [Online]. Available: http://www.cisco.com/en/US/i/000001-100000/60001-65000/62001-63000/62445.jpg. [Accessed: 16-Apr-2012].
- [2]“botnet1.jpg (JPEG Image, 589 × 584 pixels) - Scaled (94%).” [Online]. Available: http://ritcyberselfdefense.files.wordpress.com/2011/09/botnet1.jpg. [Accessed: 15-Apr-2012].
- [3]“spam-can-collection-2009-09-med.jpg (JPEG Image, 1582 × 1070 pixels) - Scaled (51%).” [Online]. Available: http://www.alaska.net/~royce/spam/spam-can-collection-2009-09-med.jpg. [Accessed: 15-Apr-2012].
- [4]“spam.jpg (JPEG Image, 990 × 660 pixels) - Scaled (83%).” [Online]. Available: http://owni.fr/files/2010/05/spam.jpg. [Accessed: 15-Apr-2012].
- [5] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, “Spamming botnets: signatures and characteristics,” SIGCOMM Comput. Commun. Rev., vol. 38, no. 4, pp. 171–182, Aug. 2008.