RB-Seeker: Auto-detection of Redirection Botnets

February 10, 2009

Xin Hu, Matthew Knysz, Kang G. Shin
{huxin, mknysz, kgshin}@eecs.umich.edu

Computer Science & Engineering, University of Michigan, Ann Arbor

Outline

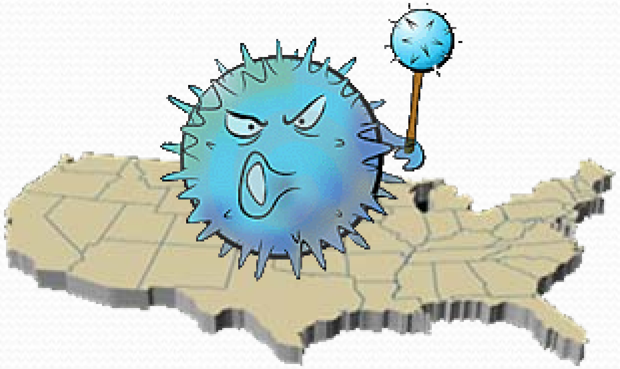
- Motivation of RB-Seeker
- System Architecture
- Overview of subsystems
- Evaluation of results
- Conclusion

Motivation: the botnet problem

- Financial Incentive
  - Underground market

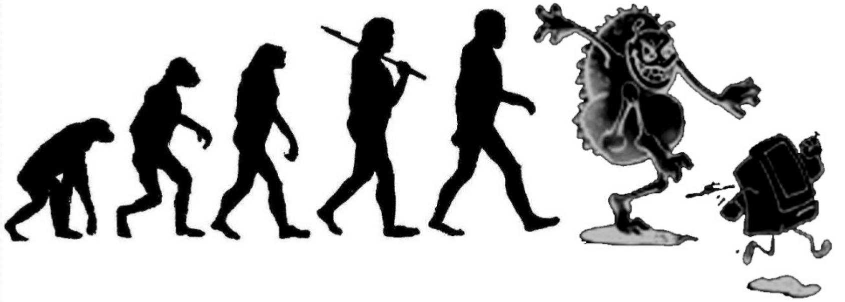
- Common uses of botnets:
  - Redirection/Proxy, Spam, ID theft, DDoS, phishing

- Can cause A LOT of damage
  - Can bring down entire systems or nations

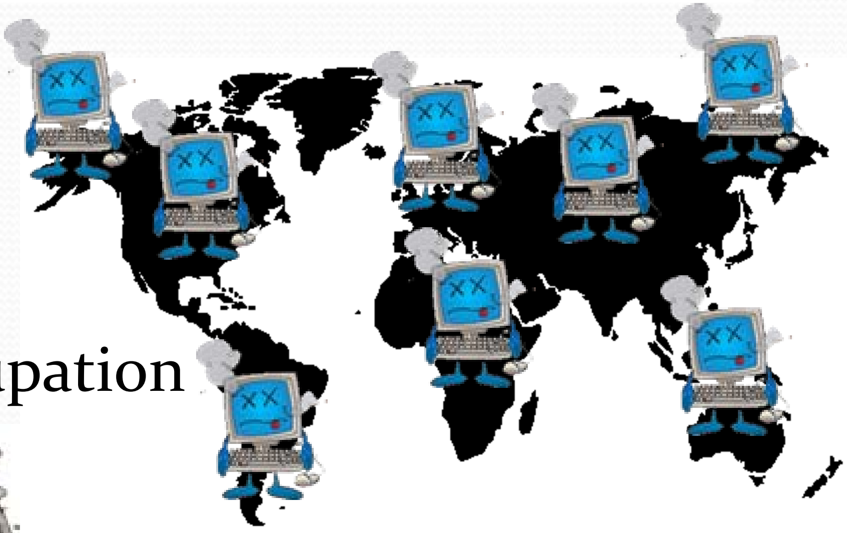


Motivation: botnet appeal

- Modular and Adaptable
  - Evolve to overcome defenses



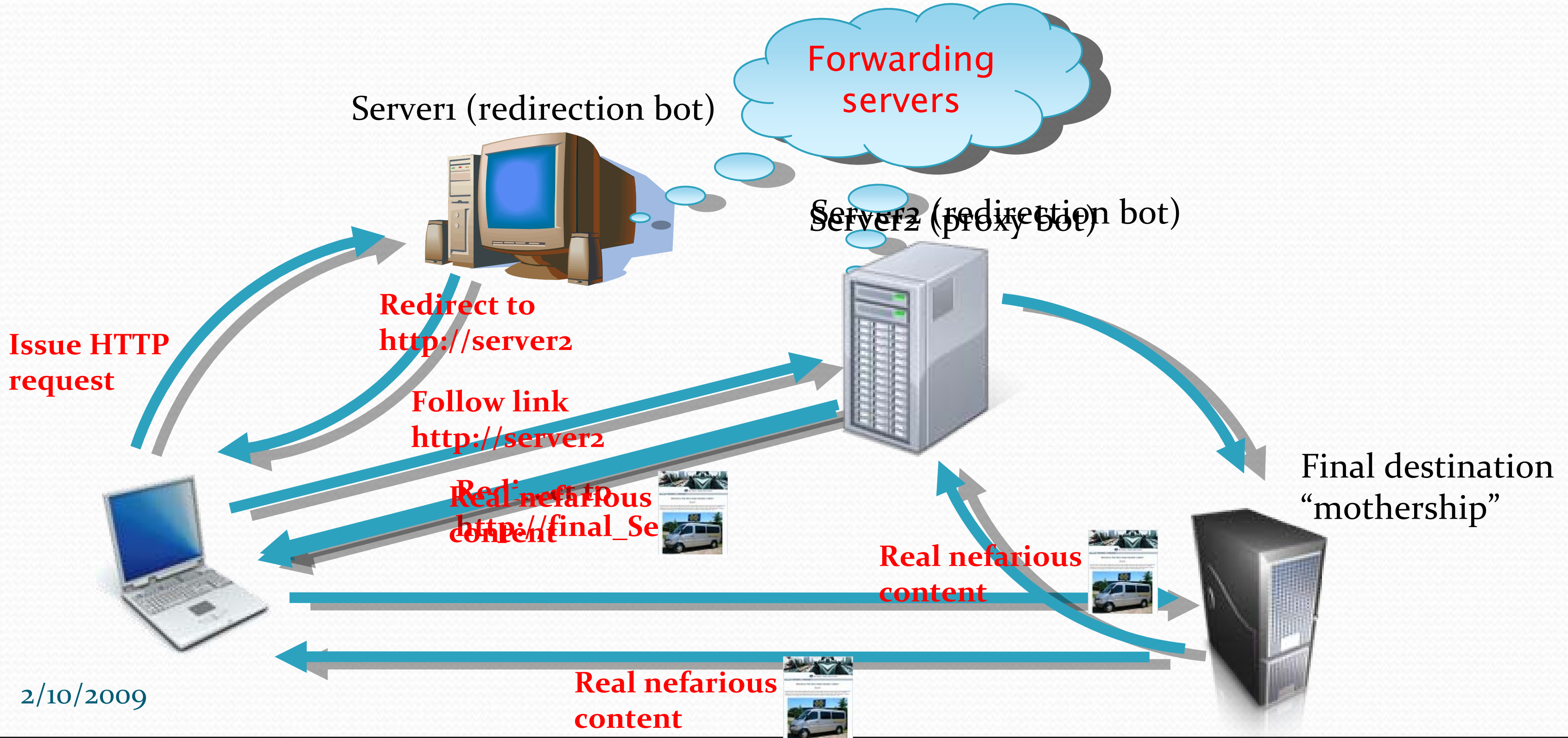
- Distributed nature
  - Difficult to find/stop botmaster



- Discreet
  - Propagation, infection, and occupation

Motivation: Redirection/Proxy Botnet

- Redirect users to malicious servers
  - Additional layer of misdirection
  - Protect mothership servers
  - Evade URL based detection or IP based black list


Forwarding servers
Server1 (redirection bot)
Server2 (redirection bot)
Server2 (proxy bot)
Issue HTTP request
Redirect to http://server2
Follow link http://server2
Redirect to http://final_Se
Real nefarious content
Final destination "mothership"
Real nefarious content
Real nefarious content

Motivation: RB-Seeker

- Botnet is an ideal source for redirection/proxy servers
- Botnets used for multiple purposes/scams
- Previous research: detection of C&C channel

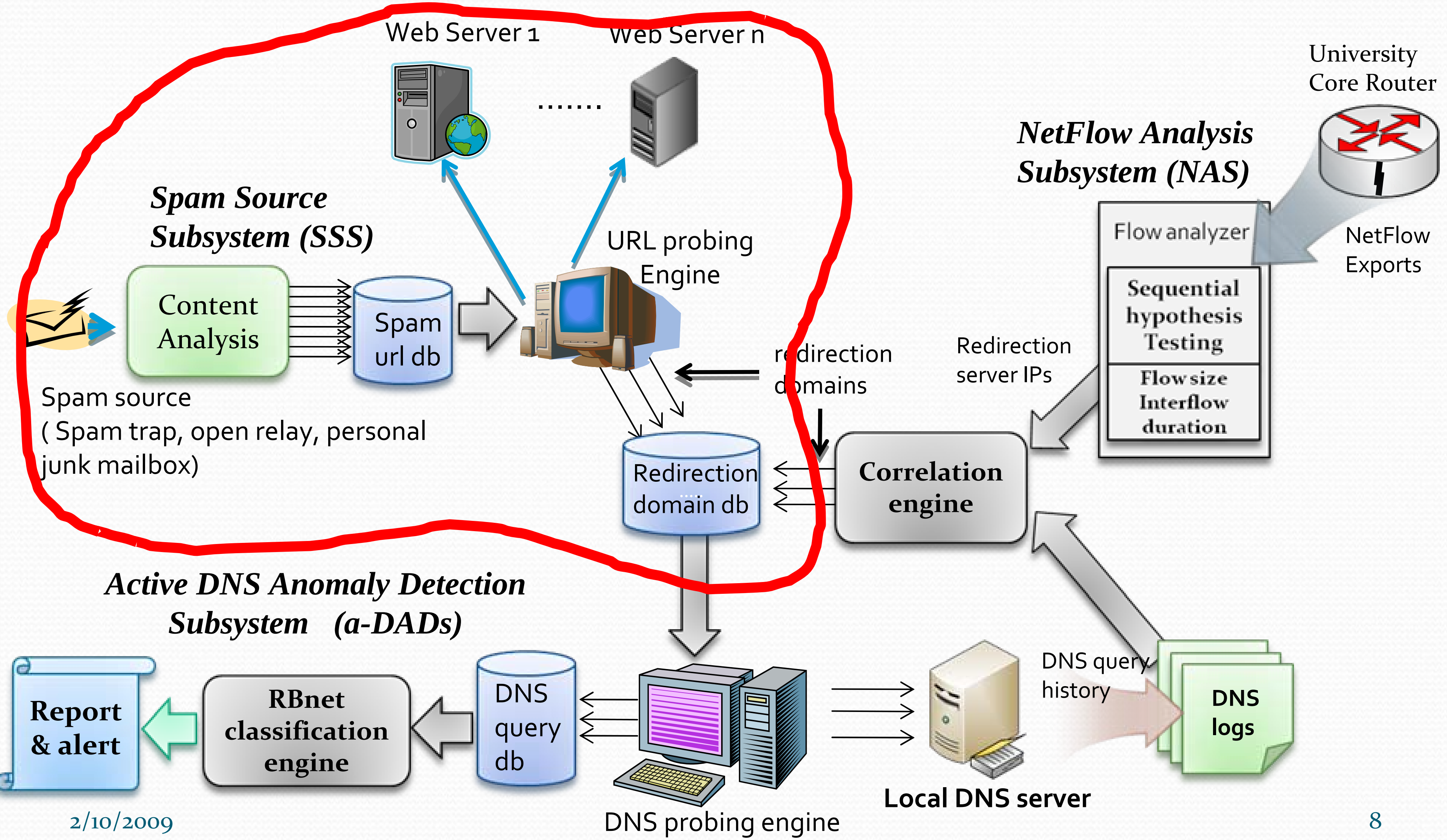
Overview: RB-Seeker

- Automatic detection of redirection/proxy botnets
- Utilizes 3 cooperating subsystems
- Behavior-based detection

- Quick identification of ***aggressive*** botnets (FP < 0.01%)
  - Advertise ***many*** IPs per query
  - Change IPs very often (short TTL)

- Accurate identification of ***stealthy*** botnets
  - Advertise ***few*** IPs per query
  - Change IPs more slowly (very small TTL, closely monitored)

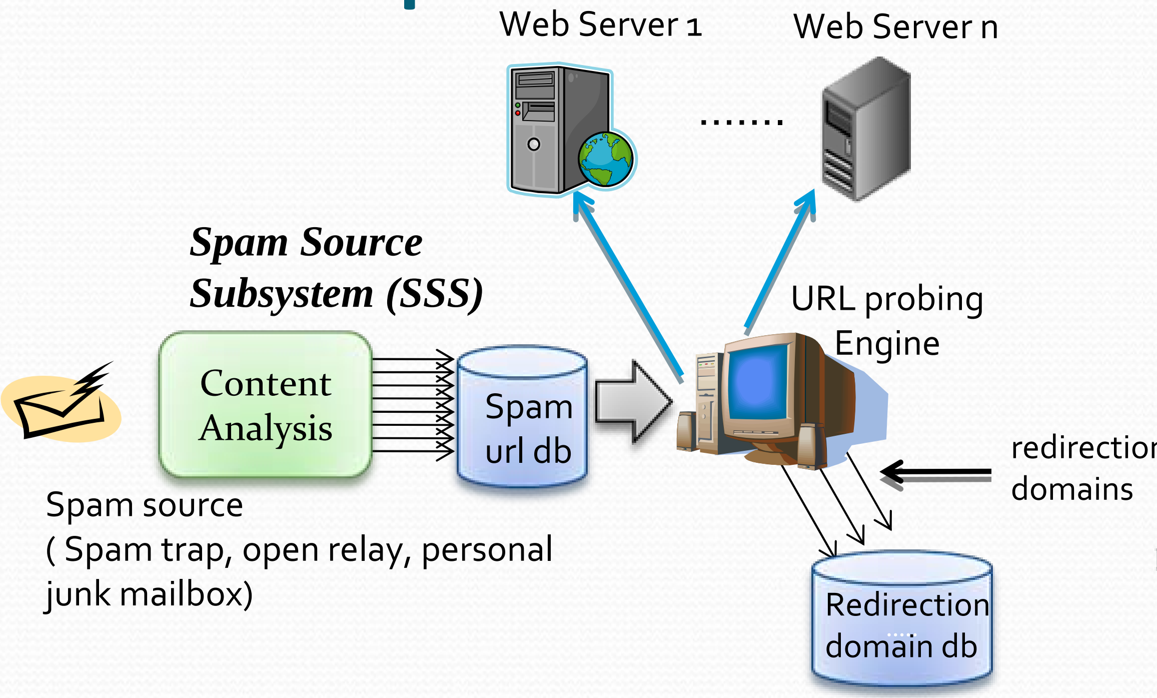
System Architecture

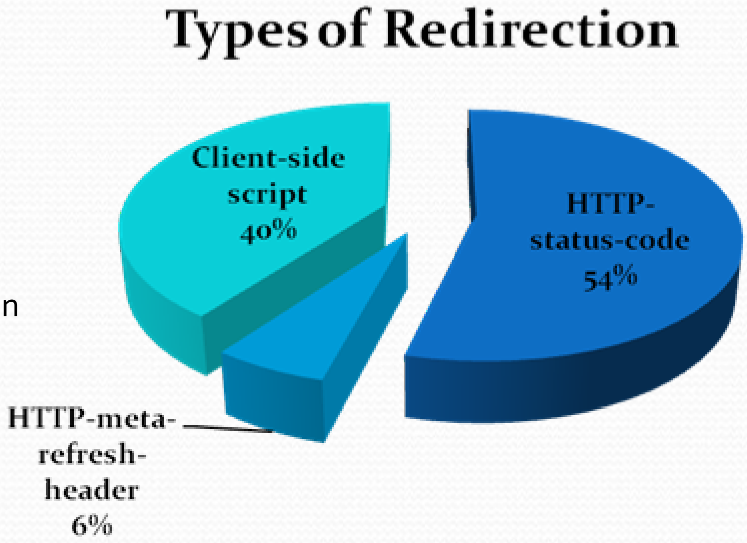


SSS: Spam Source Subsystem

- Redirection/proxy botnets are commonly used by spam/phishing campaigns
- SSS exploits this close relationship
  - Real time collection of spam emails: > 50,000 monthly

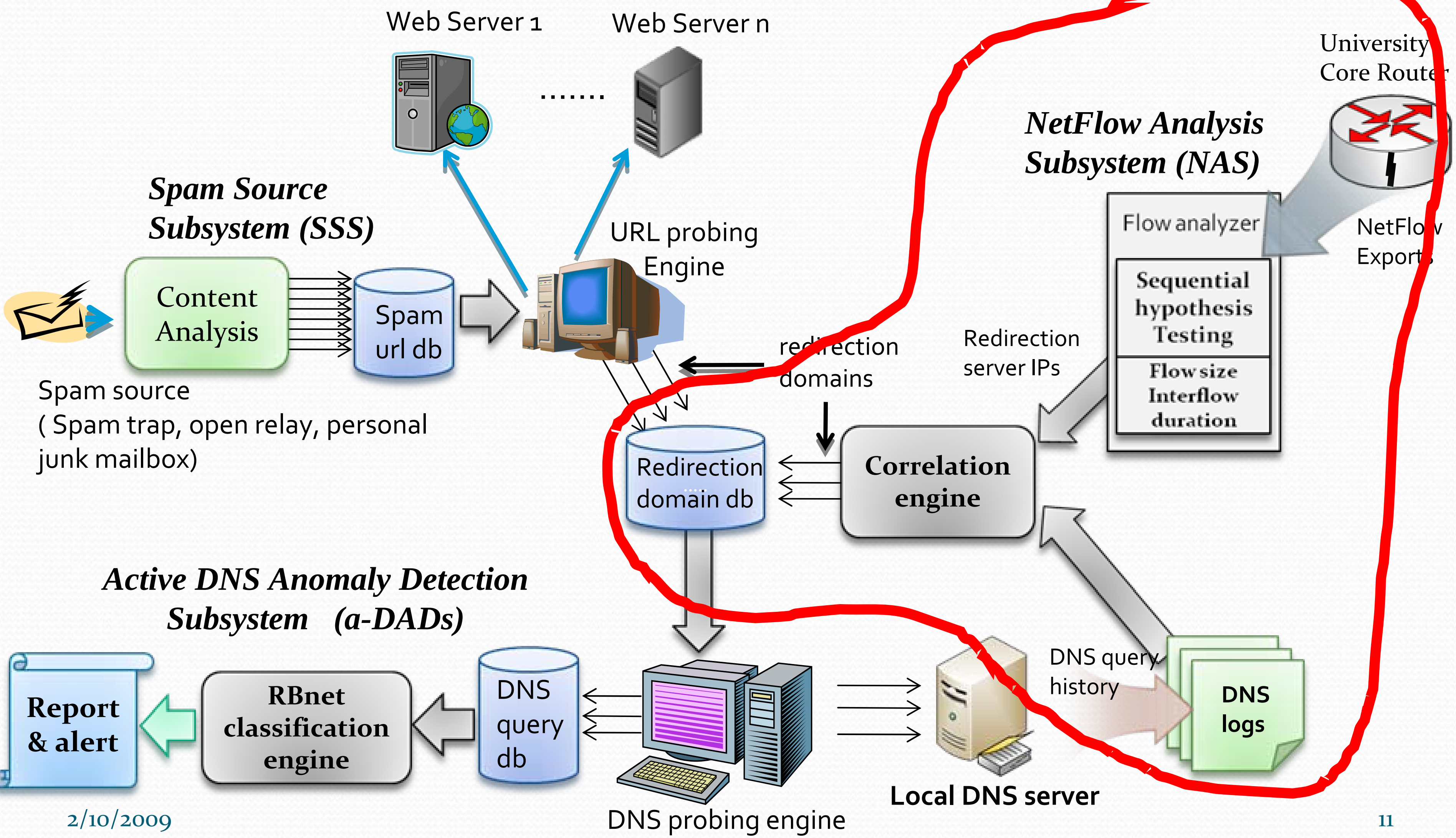
SSS: Spam Source Subsystem





1. Extract embedded URLs from message bodies
2. Probe extracted URLs to identify redirection URL links
3. Domains added to redirection domain database

System Architecture



NAS: NetFlow Analysis Subsystem

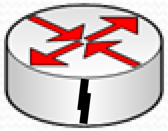
- Use NetFlow because:
  - Inspecting packet contents incurs too much overhead
  - Privacy concerns

- Spammers send image- or PDF-based emails
  - Evade content-based filtering
- User redirected to RBnet by clicking on malicious webpage
- Inspecting each email not always possible
  - Privacy concerns/laws

NAS: NetFlow Analysis Subsystem

- NetFlow: core router on campus
- Looks for suspicious redirection attempts
  - Without analyzing packet contents

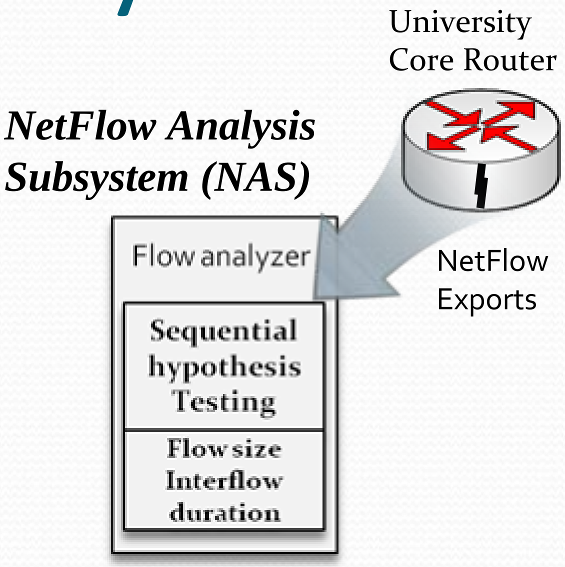
University
Core Router



NetFlow Analysis Subsystem (NAS)

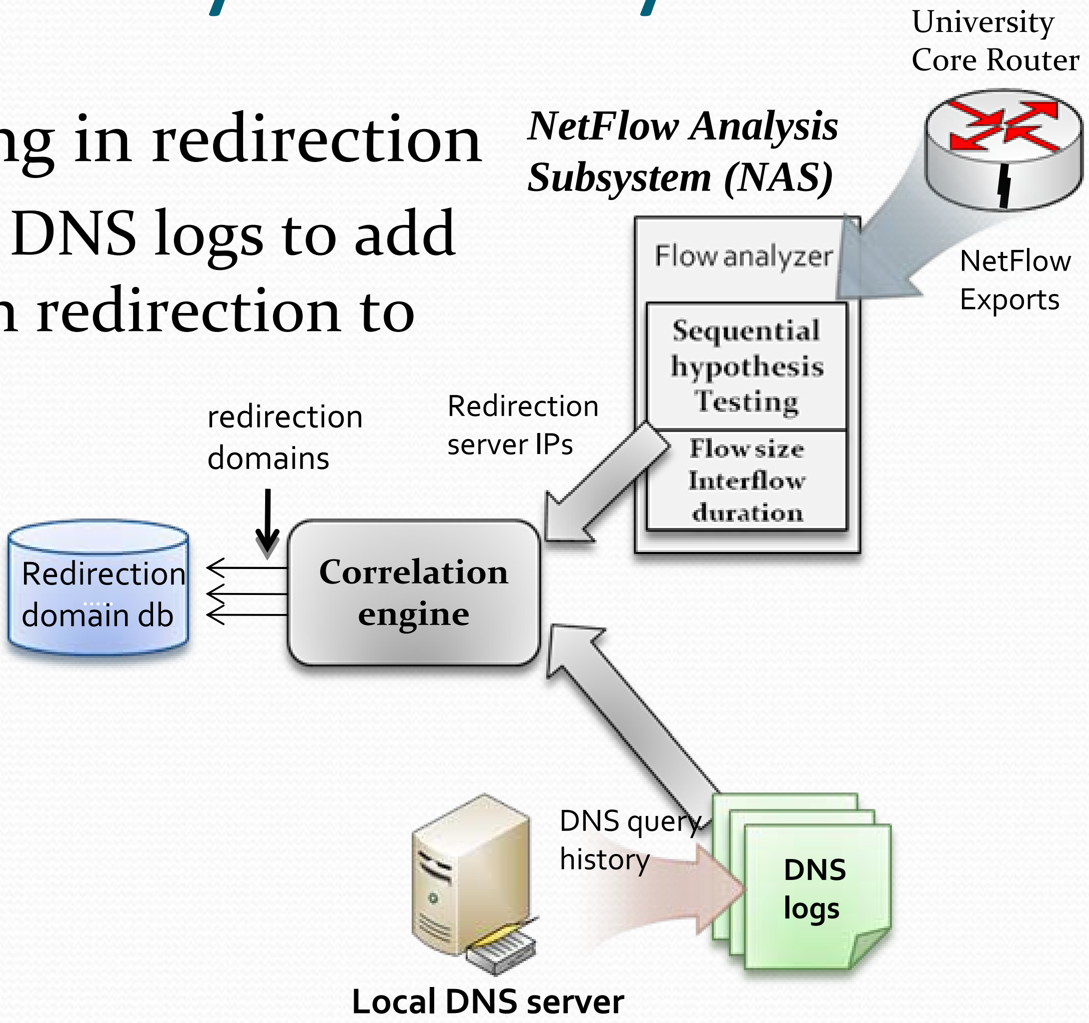
NAS: NetFlow Analysis Subsystem

- Sequential Hypothesis testing on:
  - Flow size, inter-flow duration, and flow duration


University Core Router
NetFlow Analysis Subsystem (NAS)
NetFlow Exports
Flow analyzer
Sequential hypothesis Testing
Flow size Interflow duration

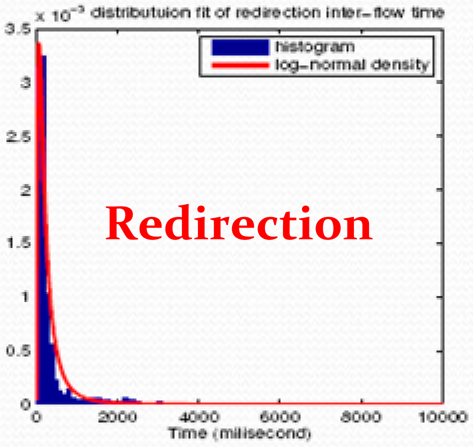
NAS: NetFlow Analysis Subsystem

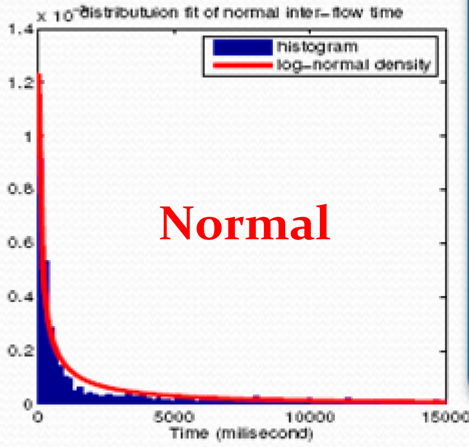
- Identifies IPs participating in redirection
  - Correlation engine uses DNS logs to add domains participating in redirection to redirection domain db

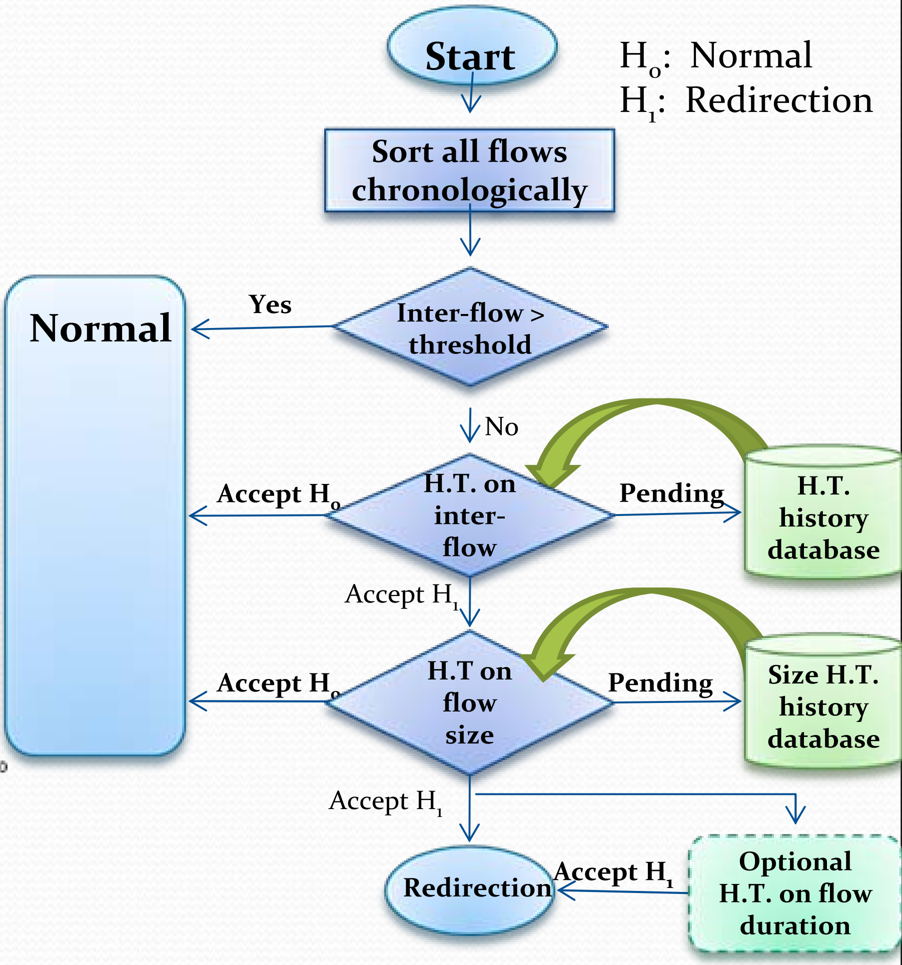


NAS: NetFlow Analysis Subsystem

		Mean	Median	Std dev
Flow duration (ms)	redirection	305.5	128.6	2159.2
	normal	33042.3	10028.8	91912.5
Inter-flow duration (ms)	redirection	392.7	154.4	872.4
	normal	40132.9	1345.5	87281.0
Flow size (bytes)	redirection	2401	629	44530
	normal	51495	4852	192431





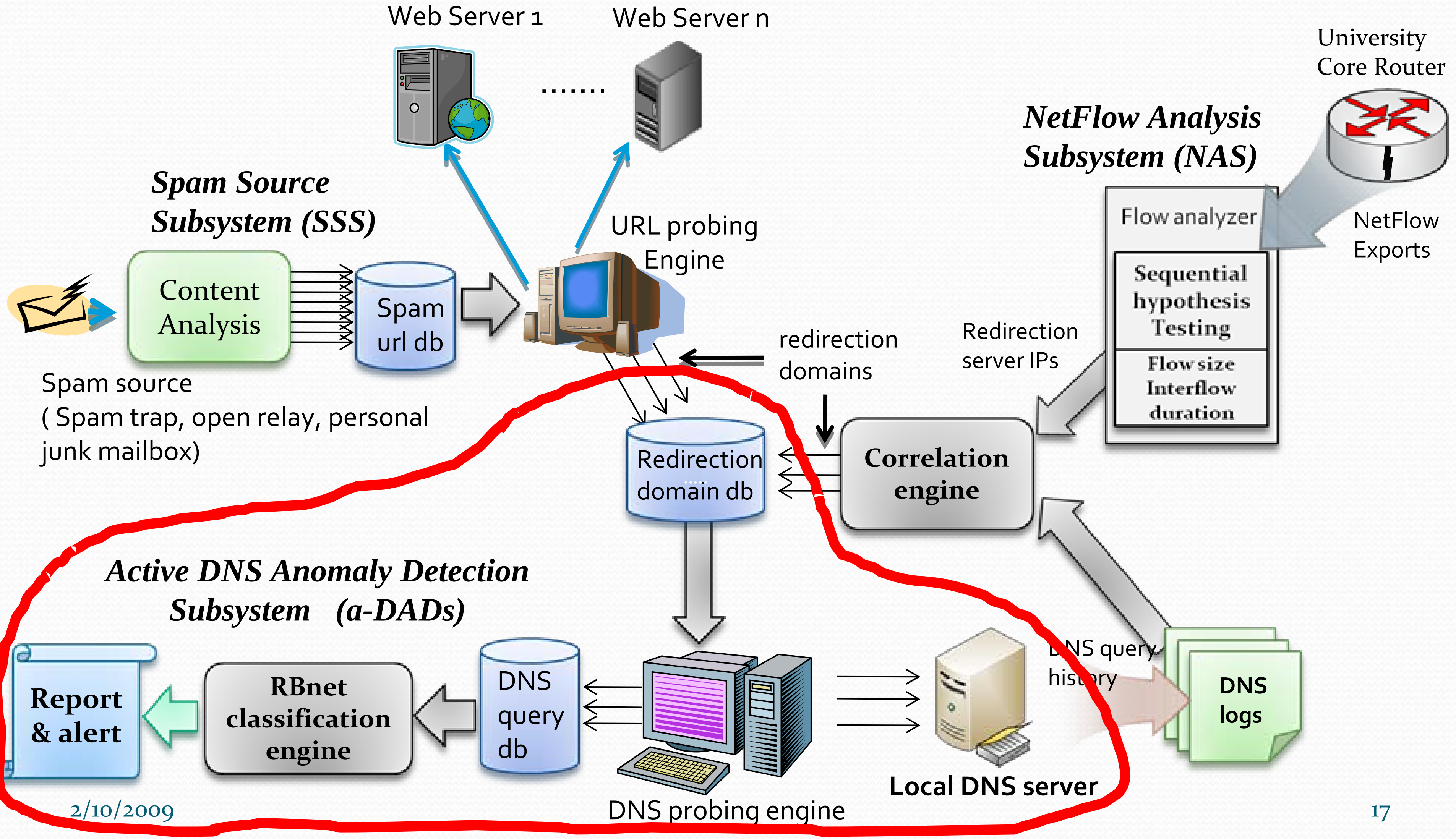


H_o: Normal
H_1: Redirection

Redirection:
obtained from SSS, servers identified as redirection

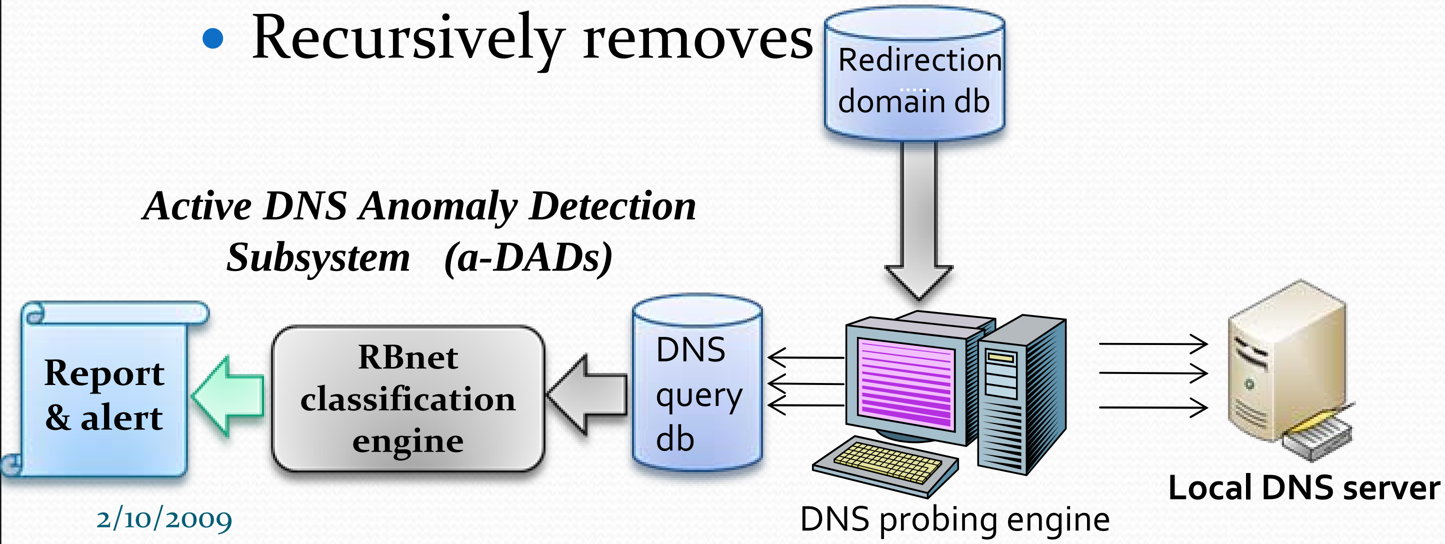
Normal:
normal web browsing over 2 days (removing redirection)

System Architecture



a-DADS: active DNS Anomaly Detection Subsystem

- Actively performs DNS queries on domains in redirection domain db
- Uses CDN Filter to remove Content Delivery Networks
  - CDNs behave similarly to redirection/proxy botnets
  - Recursively removes



a-DADS: active DNS Anomaly Detection Subsystem

- IP Usage:
  - RBnets will accrue more unique IPs over time
  - RBnets will have more unique IPs per valid query
- Reverse DNS names with “bad words”
  - e.g., broadband, cable, comcast, charter, etc…
- AS count
  - Number of different ASes the IPs belong to
  - RBnets consist of home computers scattered geographically

a-DADS: active DNS Anomaly Detection Subsystem

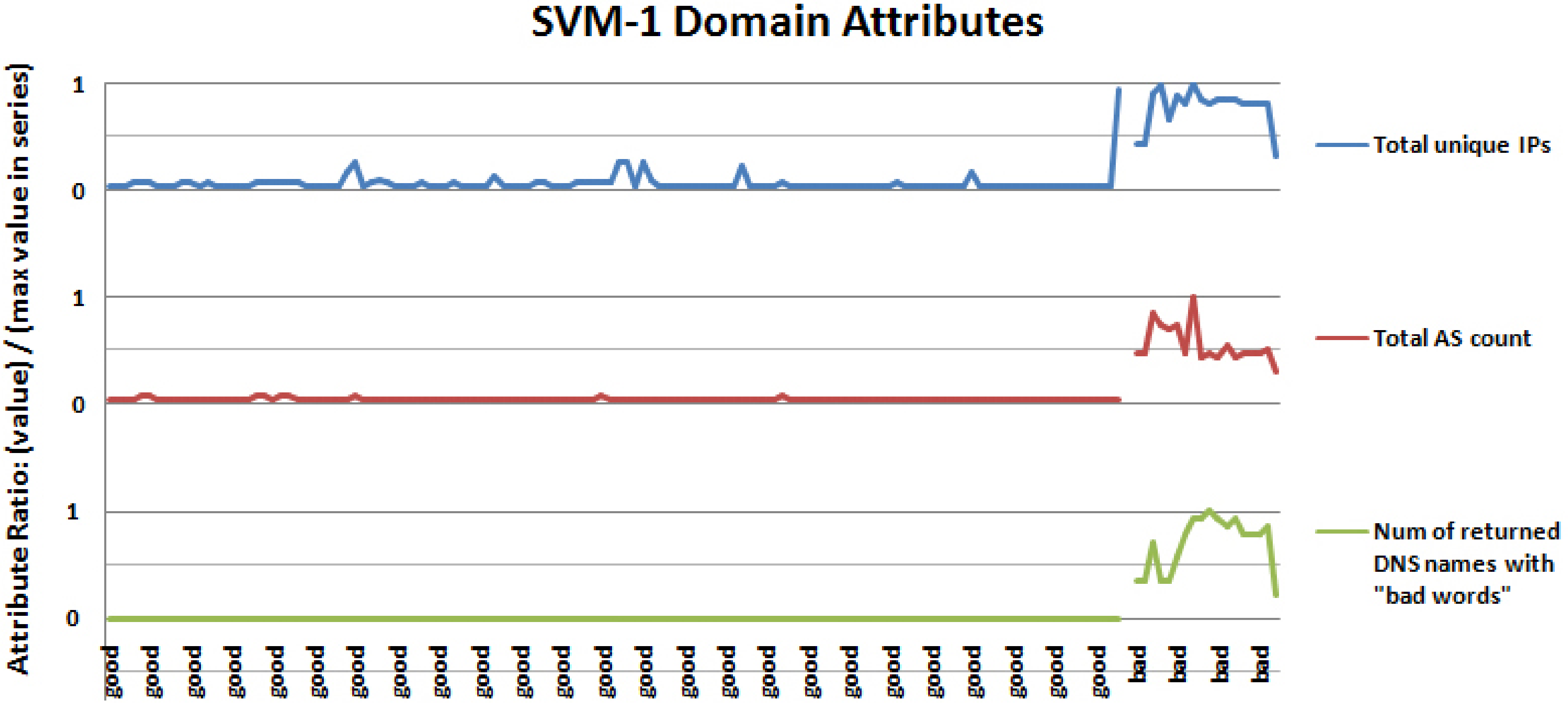
- Applies 2-tier linear SVM on remaining domains
  - Trained: 124 valid, 18 aggressive, 10 stealth
  - 10-fold cross validation on multiple classifiers
    - knn, decision tree, naïve Bayesian, various SVMs and kernel functions

$$F(x) = \begin{cases} w^T x - b > 0\,, & \textbf{if valid domain} \\ w^T x - b < 0\,, & \textbf{if RBnet domain} \end{cases}$$

a-DADS: active DNS Anomaly Detection Subsystem

- SVM-1:
  - detects ***Aggressive RBnets*** based on 2 valid queries
  - unique IPs, num ASes, DNS “bad words”

a-DADs: SVM-1 Aggressive RBnets

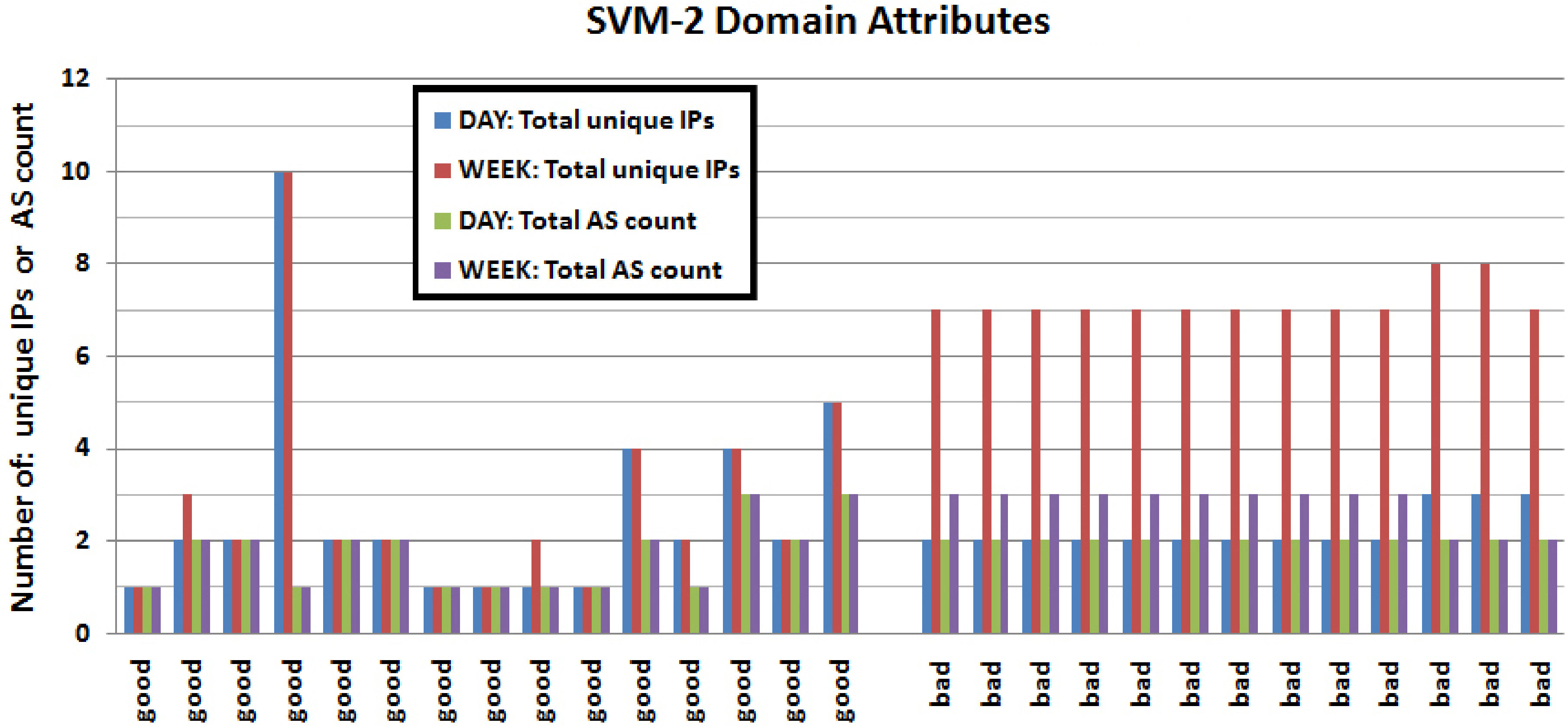


$$
\begin{aligned}
f(x) &= w^T x - b \\
&= -1.257 * N_{unique\_IPs} - 26.401 * N_{ASes} \\
&\quad -13.024 * N_{DNS\_bad\_words} + 162.851
\end{aligned}
$$

a-DADS: active DNS Anomaly Detection Subsystem

- SVM-2:
  - detects ***Stealth RBnets*** using a week of DNS queries
  - unique IPs, num ASes

a-DADs: SVM-2 Stealth RBnets



$$
\begin{aligned}
f(x) &= w^T x - b \\
&= 52.497 * N_{DAY\_unique\_IPs} - 63.109 * N_{WEEK\_unique\_IPs} \\
&\quad - 10.924 * (N_{DAY\_ASes} + N_{WEEK\_ASes}) + 227.985
\end{aligned}
$$

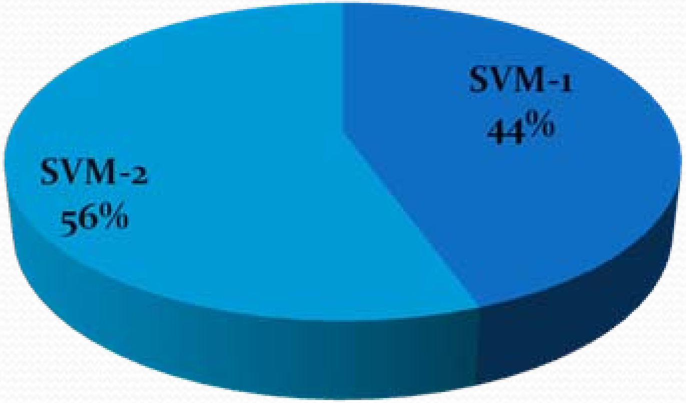
Evaluation of Results

- SSS and NAS identified 91,600+ suspicious domains over 2 month period
- a-DADS CDN Filter
  - Removed 5,005 CDN domains
  - Recursion 16.8% increase in identified CDN domains (13.1% in IPs)
  - Similar technique for valid domains reduced this to 35,000+ domains to be monitored

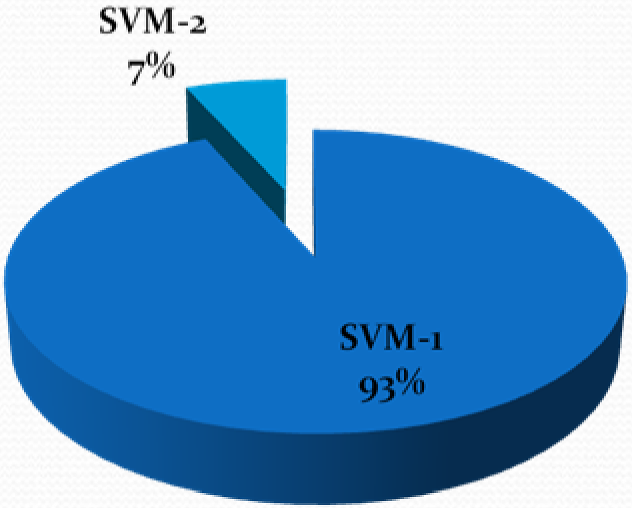
Evaluation of Results

	RBnet Domains	RBnet IPs	Valid Queries Used
SVM-1	125	3,541	2 queries
SVM-2	156	249	1 week
RB-Seeker	281	3,790	2 queries/1 week

SVM-1: Experienced 1 FP (< 0.008%)



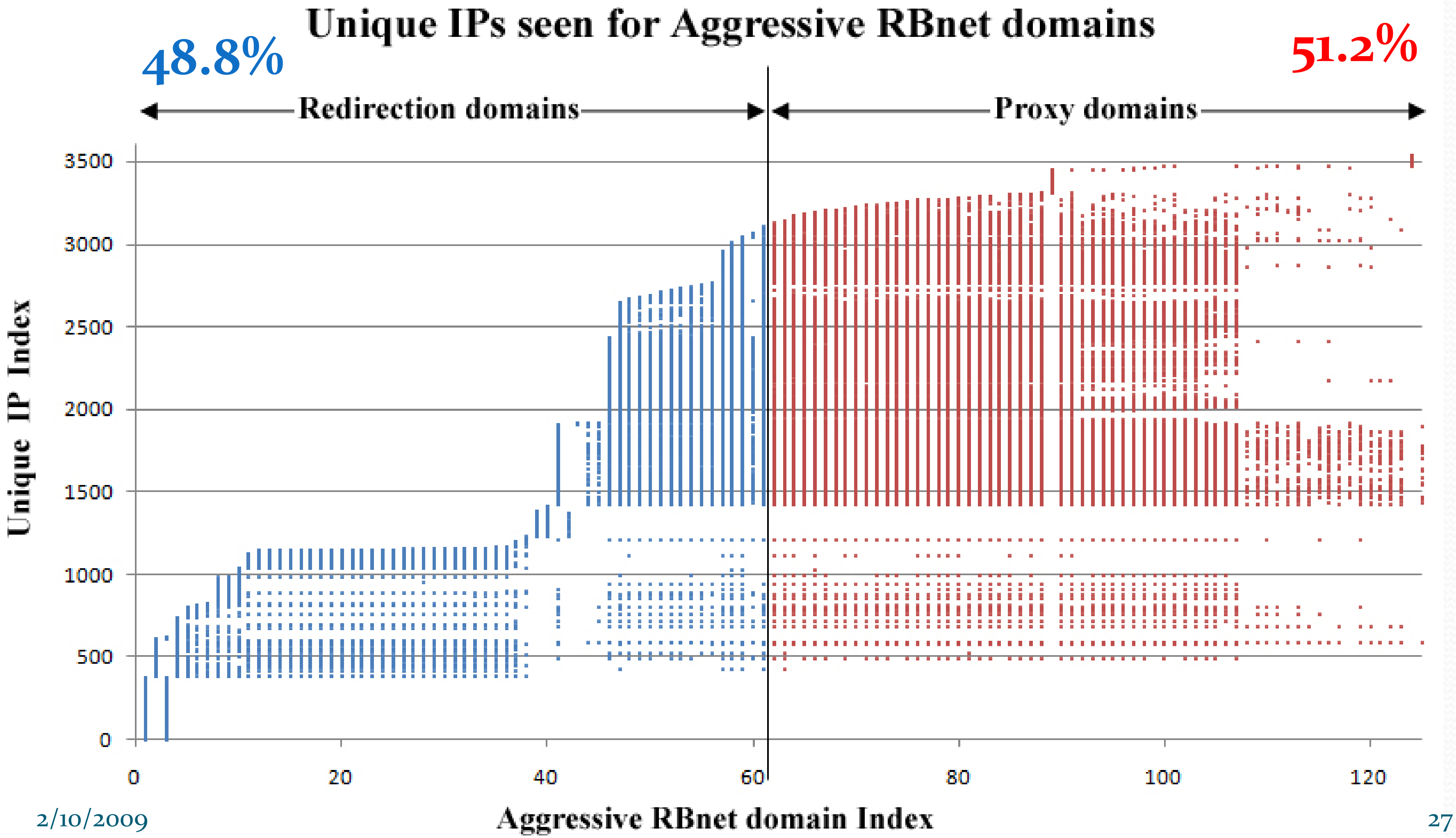
RBnet Domains



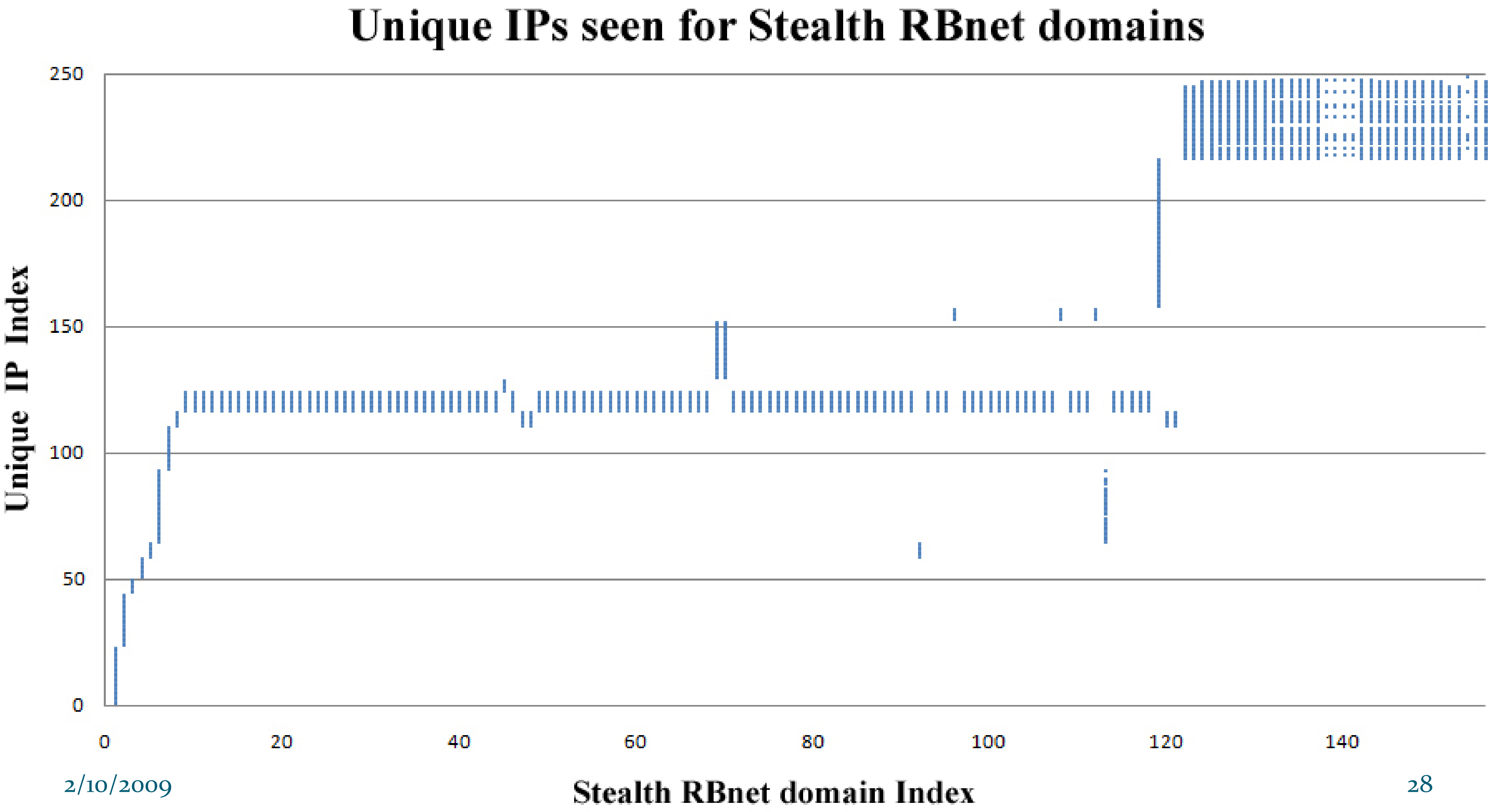
RBnet IPs

Aggressive RBnets:

Redirection vs. Proxy Botnets



Stealth RBnets



Evaluation of Results

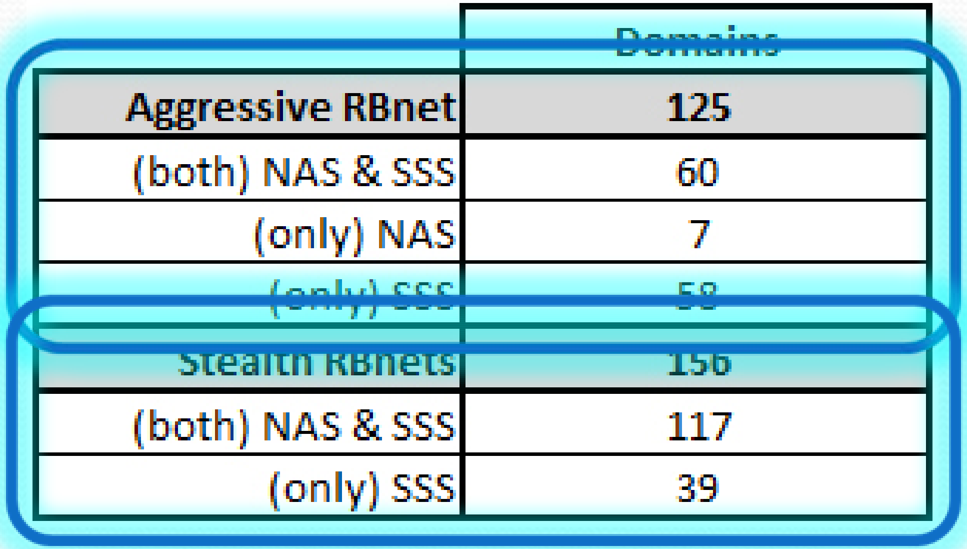
- FFSN detector:
  - Detected 124 of the 125 Aggressive RBnets
  - 1 FP: same as ours (mozilla.org)
  - Missed all the Stealth RBnets

Conclusion

- Designed and implemented system for detecting redirection/proxy botnets
- Uses network detection techniques
  - multiple data sources readily available to enterprise network environments
- Behavior-based detection works despite use of C&C protocol or structure
- Capable of detecting Aggressive and Stealthy RBnets
- Automatic detection with low false positives (< 0.01%)

Questions?

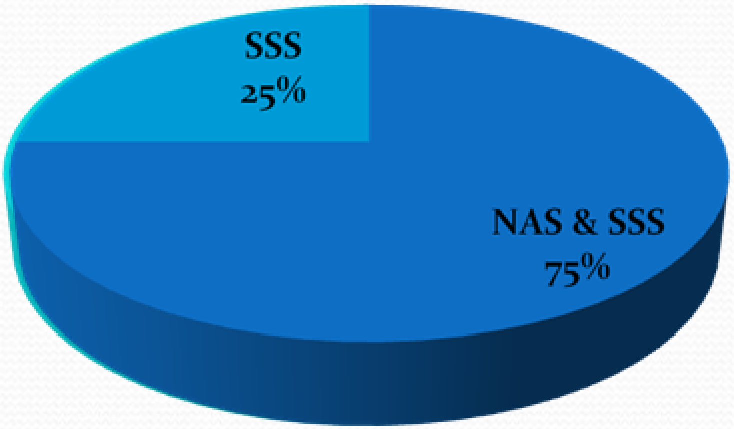
Evaluation of Results

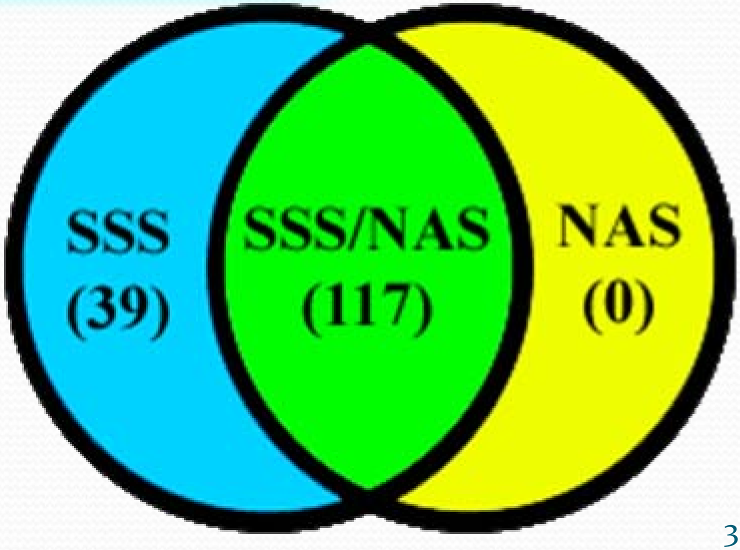


	Domains
Aggressive RBnet	125
(both) NAS & SSS	60
(only) NAS	7
(only) SSS	58
Stealth RBnets	156
(both) NAS & SSS	117
(only) SSS	39

Domains Aggressive RBnets

Domains Stealth RBnets


SSS
NAS & SSS


SSS
(39)
SSS/NAS
(117)
NAS
(0)