

Squeeze

Detecting Malware's Failover C&C Strategies

Matthias Neugschwandtner
Paolo Milani Comparetti
Christian Platzer

International Secure Systems Lab
Vienna University of Technology



Motivation

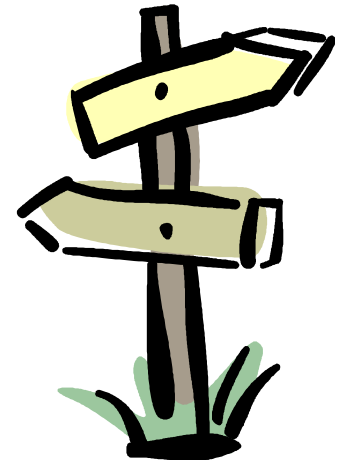
- Bots
 - Send spam, perform DDoS attacks, advertise fake-AVs, “click” on advertisements
 - Part of a “botnet”
 - Updated and controlled by a botmaster through a C&C infrastructure
- Blacklists & Take-downs
 - Render botnet useless by crippling the C&C infrastructure
 - Shut down C&C servers

Motivation

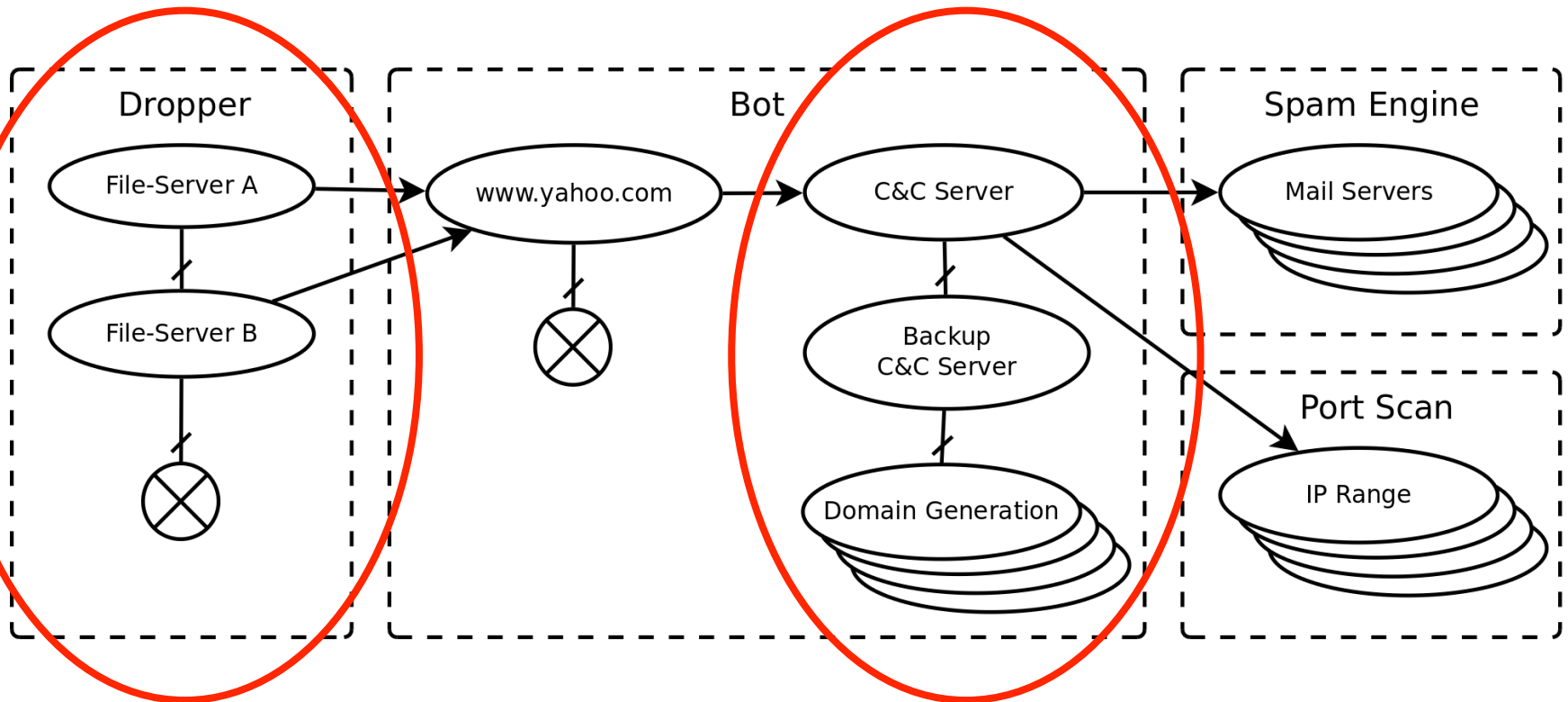


Outline

- Running Example
- Our Approach
 - Multipath exploration
 - Exploration strategies
- Squeeze
- Evaluation
- Conclusion



Running Example



Multipath Exploration

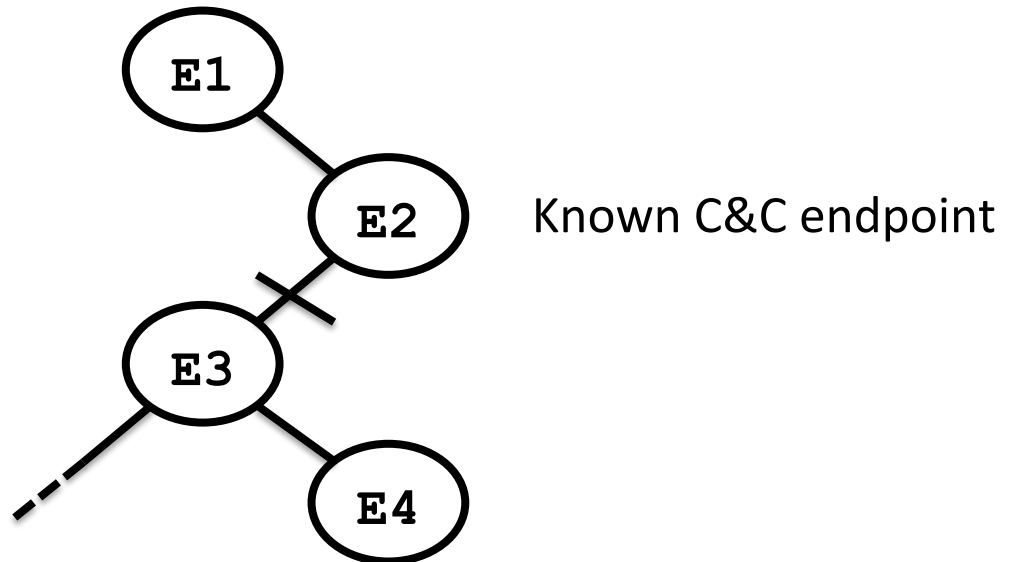
- Network analysis process can be represented as a binary tree
 - Endpoints are nodes
 - Decision whether or not to block is binary
- Standard tree traversal methods are not sufficient
 - Breadth-first: exhaustive exploration of initial behavior
 - Depth-first: get stuck in e.g. port-scan

Multipath Exploration

- C&C domain knowledge
 - C&C network traffic signature
 - Known C&C endpoints
- Limit exploration to “bot component” process of the malware
- C&C knowledge allows for different strategies

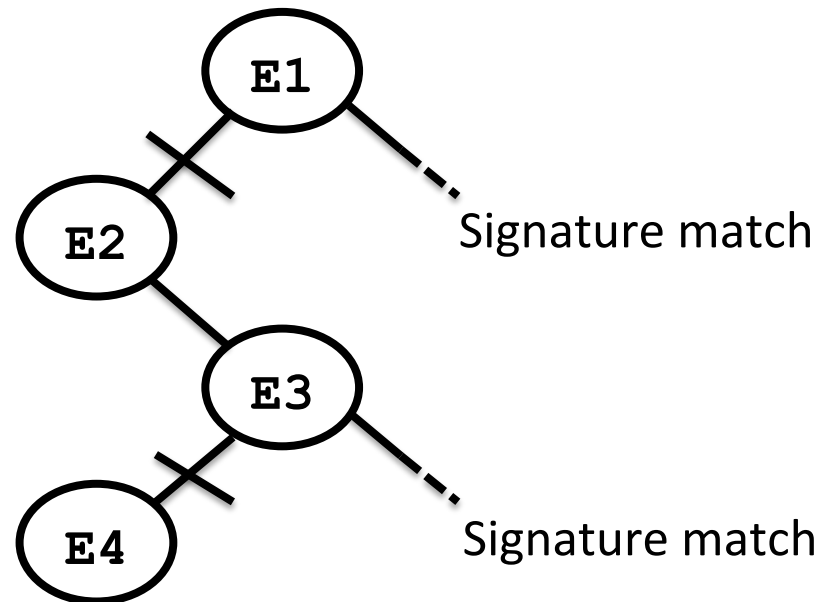
Exploration Strategy A

- Initially allow all connections
- Encountering a known C&C endpoint, switch to block-first, depth-first search
- Backtrack if there is no further network activity



Exploration Strategy B

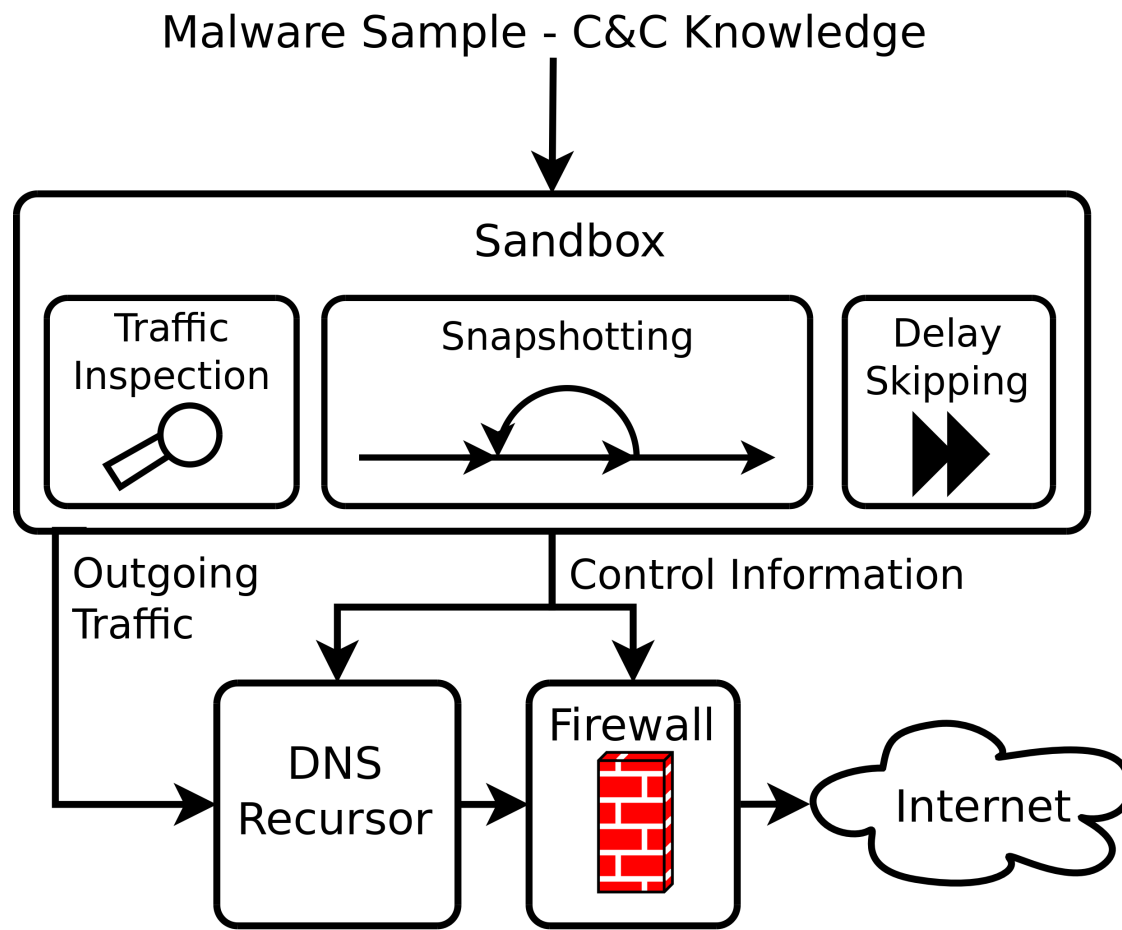
- Allow-first depth-first search
- Encountering C&C traffic, backtrack and block the corresponding endpoint



Strategy Comparison

	Strategy A	Strategy B
Pros	Potential to reveal more endpoints overall	Triggers even if the endpoint is previously unknown
Cons	C&C endpoints can only be confirmed if the search backtracks far enough	Relies on detecting C&C communication on the fly

Squeeze Architecture



Squeeze Sandbox

- Snapshotting
 - Extend Qemu snapshotting to include state of Anubis instrumentation
 - Keep snapshots in main memory
- Delay skipping
 - Limit `NtDelayExecution` to max 0.5s
 - Forward `GetTickCount` by exponentially growing interval
- Traffic interception
 - Device control of `afd.sys`: `AFD_CONNECT`, `AFD_SEND`
 - DNS client's `DnsQuery` function

Squeeze Networking

- Connection blocking
 - PowerDNS recursor to create NXDOMAIN replies
 - Netfilter firewall
 - TCP SYN → TCP reset
 - UDP packet → ICMP destination unreachable
- Contain harmful traffic
 - Redirect well-known ports to honeypot
 - Throttle network throughput and number of connections

Evaluation

- Input to Squeeze
 - Domain knowledge
 - Sample selection
- Malware behavior under Squeeze
- Endpoints revealed by Squeeze
- Fallback strategy insight
- Real-world deployment

Evaluation – Squeeze Input

- Domain knowledge: C&C network traffic signatures
 - Provided by a security company
 - Manually vetted by experts
- Sample selection
 - Matched signatures against traffic dumps
 - Discarded unreachable endpoints
 - Selected most recent samples

Evaluation – Dataset and Setup

- Dataset of 8,346 samples
 - 213 families
 - Most represented families account for only 21%
 - 14 of the top 20 malware families covered
 - Exceptions: Conficker, Storm, Bredolab
- Setup
 - Six minute analysis runs, max. 21 snapshots
 - 10 days in March 2011
 - 5 days Strategy A – Dataset A with 4,013 samples
 - 5 days Strategy B – Dataset B with 4,333 samples

Evaluation – Malware Behavior

	Strategy A	Strategy B	
Samples analyzed	4013	4333	
Initial C&C knowledge match	54%	58%	
No further activity	44%	43%	
Substantial delay skipping	34%	31%	
New endpoints	25%	23%	
New endpoints in bot component	19%	13%	upper bound
New endpoints with signature match	9%	8%	lower bound

Over 10% of the samples performed connectivity checks

Evaluation – Endpoints

Dataset A	Baseline			Strategy A			
	Dom.	IPs	Total	Dom.	IPs	Total	
Endpoints	3562	767	4329	661	942	1603	
Endpoints in bot component	2080	362	2432	454	260	714	29.4%
Endpoints in blacklists	1970	111	2081	391	36	427	20.5%
Endpoints with signature match	1489	110	1599	195	6	201	12.6%

Dataset B	Baseline			Strategy B			
	Dom.	IPs	Total	Dom.	IPs	Total	
Endpoints	2627	364	2991	534	325	859	
Endpoints in bot component	1330	211	1541	293	213	506	32.8%
Endpoints in blacklists	1336	81	1417	353	15	368	26.0%
Endpoints with signature match	885	53	938	184	1	185	19.7%

Evaluation – Qualitative Results

- Palevo/Butterfly
 - 2 static IPs, both matching signatures
 - DGA afterwards, none of the 42 domains registered
- Pakes
 - Idles for several minutes before contacting backups
 - 20 additional C&C servers revealed
- Koobface
 - Connection check using www.google.com
 - Up to 50 C&C servers revealed
- Piptea
 - One-stage fallback
 - Taken offline during our evaluation
- Virut, zBot: no backup strategy in place

Evaluation – Deployment

- Deployment using strategy B
 - June – August 2011
 - Over 32,000 samples analyzed
- Improvement over baseline:
 - 39% in bot component
 - 12% with signature match
- Reduced delay until re-analysis
 - Only several hours instead of days
 - Traffic from over 90% of the samples re-matched signatures

Conclusions

- Squeeze is effective in revealing C&C failover strategies
- Complementing malware analysis systems with Squeeze improves the coverage of automatically generated blacklists
- Squeeze is currently deployed as part of the Anubis malware analysis framework