D1: Malware Traffic Detection using Tamper Resistant Features

Research Lead: Dr. Patrick McDaniel, The Pennsylvania State University, mcdaniel@cse.psu.edu

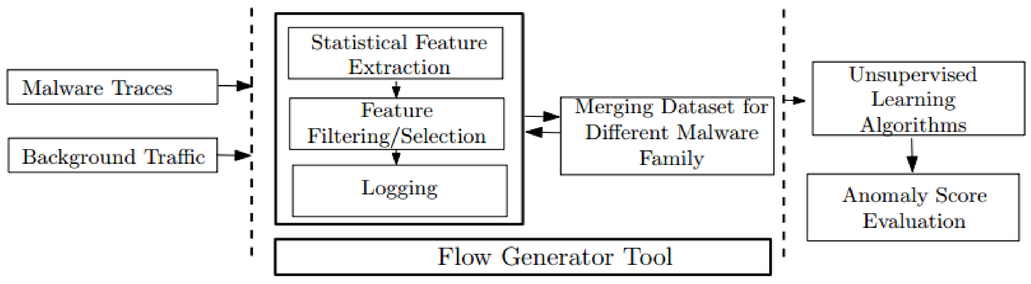
Research Goals

Evaluation of the transport layer feature space of malware heartbeat traffic to distinguish malware traffic from traffic generated by legitimate applications. In contrast to previous work:

- We eliminate features at risk of producing overly optimistic detection results (*e.g.,* port numbers)
- We rely only on tamper-resistant features making it difficult for sophisticated malware to avoid detection (*e.g.,* TCP flags and URLs)
- We detect previously unobserved anomalous behaviors

Terminology: Heartbeat traffic refers to the subset of the malware traffic that is either in a sleep or stealth state where the malware slowly/subtly generates traffic to send control, keep alive, command transfer messages, update requests, or peer list queries.

Technical Approach



1 Feature Extraction and Selection:

- Statistics of TCP network flows extracted from only five unidirectional or bidirectional sequence packets between two endpoints after a successful 3-way handshake is established.

2 Feature Calibration

- Calibrate the time-sensitive features of malware traces of timing-based features.
- Calibration process includes the sampling of the eligible RTTs (both client to server and server to client) from background traffic and changing the RTTs of malware traffic to provide consistency between the timing-based features of the two traces.

3 Novelty (Anomaly) Detection Algorithms

- Four anomaly detection algorithms (OCSVM, k-NN, LSAD, k-Means) based on the idea that anomalies are rare compared to the normal traffic are applied

Results

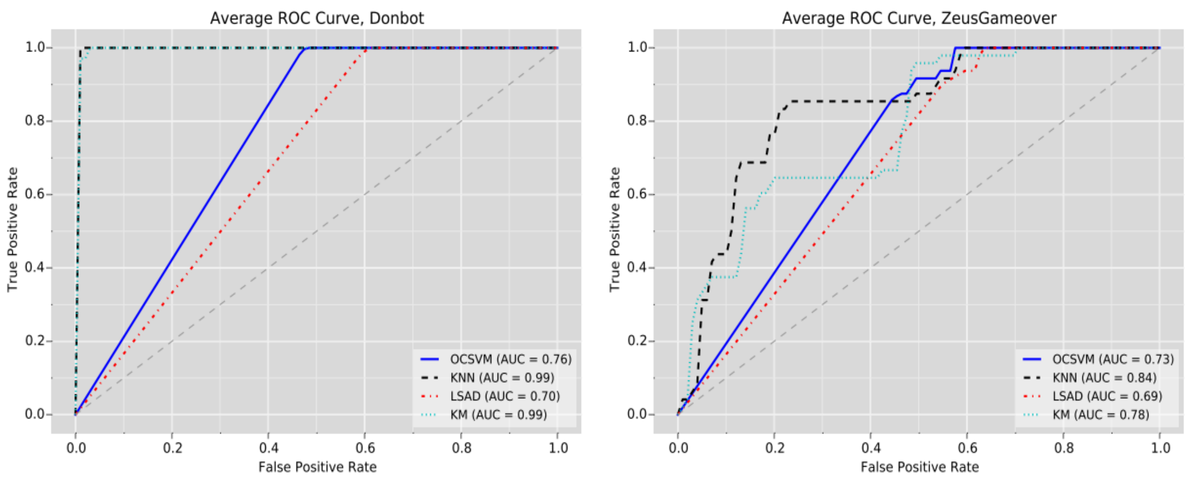
- **Dataset**

16 Malware Families are blended into university network traces.

Dataset			Method			
	Date	Number of Flows	OCSVM	k-NN	LSAD	k-Means
Agobot	2002	8	*0.7697*	**0.9779**	*0.7075*	*0.9505*
Donbot	2006	33	0.7632	*0.9979*	0.6987	**0.9983**
Kaiten	2007	49	*0.7726*	**0.7776**	*0.7141*	*0.4186*
ZeusV1	2007	50	*0.7864*	**0.8538**	*0.7207*	*0.7587*
Qbot	2008	126	*0.7994*	**0.9166**	0.7236	0.8410
Sality	2008	4	*0.7686*	**0.8567**	*0.7107*	*0.6196*
Torpig	2008	4	*0.7786*	**0.8412**	*0.7120*	*0.7611*
Neris	2009	1688	*0.8013*	**0.8337**	*0.7361*	*0.8112*
Kelihos	2010	8	0.7762	**0.9846**	0.7136	*0.9734*
Rbot	2010	806	*0.7664*	**0.8966**	0.6969	0.8304
Spyeye	2010	15	*0.7737*	*0.8183*	*0.7161*	**0.8271**
Zeroaccess	2011	363	*0.8065*	**0.8708**	0.7252	*0.7508*
ZeusGameover	2011	48	0.7347	**0.8373**	*0.6923*	*0.7769*
Tbot	2012	384	*0.8073*	*0.9131*	0.7239	**0.9161**
ZeusPonyloader	2012	8	*0.7754*	**0.8815**	*0.7144*	*0.6679*
ZeusV2	2013	6	*0.6868*	**0.7421**	*0.7239*	*0.7350*
Avg. Time			460.75	10.66	91.85	68.64

TABLE I: AUC results

TCP traffic is still commonly utilized (*i.e.*, mostly in HTTP(S) traffic) for malicious activities and heartbeat messages.



The performance of the detectors gradually drops with the evaluation of malware families

- **HTTP(S) Mimicry**

	Sality	Tbot	Spyeye	ZeusV1	ZeusGameOver	ZeusP.Loader	ZeusV2
web	0 (0%)	11 (0.025%)	4 (0.009%)	8 (0.018%)	8 (0.018%)	0 (0%)	0 (0%)

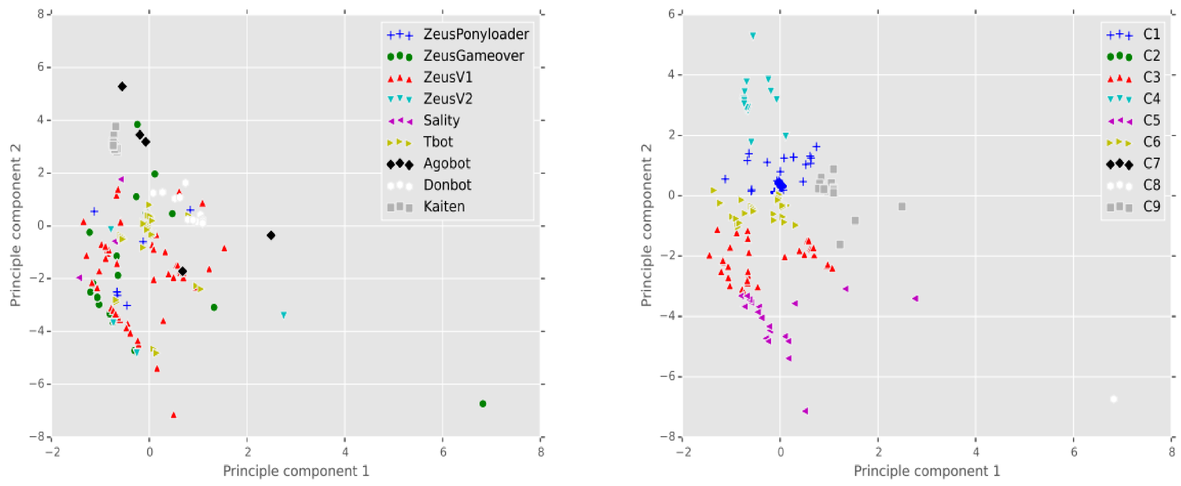
TABLE II: False positive counts

	Sality	Tbot	Spyeye	ZeusV1	ZeusGameOver	ZeusP.Loader	ZeusV2
web	4 (100%)	11 (0.029%)	9 (60%)	23 (45%)	22 (45.8%)	8 (100%)	6 (100%)

TABLE III: False negative counts

Recent malware variants mimic the legitimate HTTP(S) traffic in order to disguise their traffic.

- **Code Reuse**



The similar feature space of malware may be indicative of code reuse, or addition of new patches to previous versions.

Publications

- Malware Traffic Detection using Tamper Resistant Features, submitted to MILCOM'15

Primary Researchers

Z. Berkay Celik	PSU	zbc102@cse.psu.edu
Robert J. Walls	PSU	rjwalls@cse.psu.edu
Patrick McDaniel	PSU	mcdaniel@cse.psu.edu
Ananthram Swami	ARL	ananthram.swami.civ@mail.mil

Task Rotations (listed by PI)

Dr. Patrick McDaniel, PI (PSU), 10 days, ARL

Collaborations

PSU	Penn State University
ARL	Army Research Lab