GECCO 2017 Tutorial: Evolutionary Computation in Network Management and Security

Nur Zincir-Heywood & Gunes Kayacik
Dalhousie University
Halifax, Canada
zincir@cs.dal.ca

http://gecco-2017.sigevo.org/

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
GECCO '17 Companion, July 15-19, 2017, Berlin, Germany
ACM 978-1-4503-4939-0/17/07.
http://dx.doi.org/10.1145/3067695.3067726

1

Instructors

- Nur Zincir-Heywood is a Professor of Computer Science at Dalhousie University, Canada. Her research interests include computational intelligence and data analytics for network operations and cyber security. She currently works on traffic and behavior analysis for network / service management and cyber-security.

- Gunes Kayacik is a Research Scientist at Silicon Valley, USA. His research interests have always been found in the middle ground between computer security and machine learning. Dr. Kayacik has worked at Silicon Valley start-ups, developing machine learning methods for botnet detection and data leak prevention, which protected several thousand end users and hosts.

2

Agenda

- Introduction to Network Management and Cyberseucity
- Network monitoring
- System monitoring
- Streaming data analysis
- Security data analysis
- Overview and Examples
- Questions and Discussion

3

Network Management is….

- Deployment, Integration & Coordination
- of Hardware, Software & Human elements
- for Configuring, Monitoring, Analyzing, Testing, & Controlling
- to meet Real-Time Operational Performance & Quality of Service at Reasonable Cost

4

Network Management Framework

- Managing server(s)
- Managed device(s)
- Management Information Base
- Management agent
- Management protocol

5

Network Management Tasks

- Configuration Mng.
  - Topology
  - Discovery
  - Reconfiguration
- Performance Mng.
  - Capacity
  - Traffic
  - Throughput
  - Delay / Response time
- Fault Mng.
  - Identification
  - Reactive
  - Proactive
- Accounting Mng.
  - Cost
  - Efficiency
  - Planning

6

Network Security is…

- Protection
- of Data and Resources
- For Confidentiality, Integrity, Availability
- To keep up with latest technology
- At reasonable cost

7

Network Security Systems

- Firewalls
- Intrusion Detection / Prevention Systems
  - Signature based
  - Anomaly based
- Vulnerability Analysis
- Penetration Analysis

8

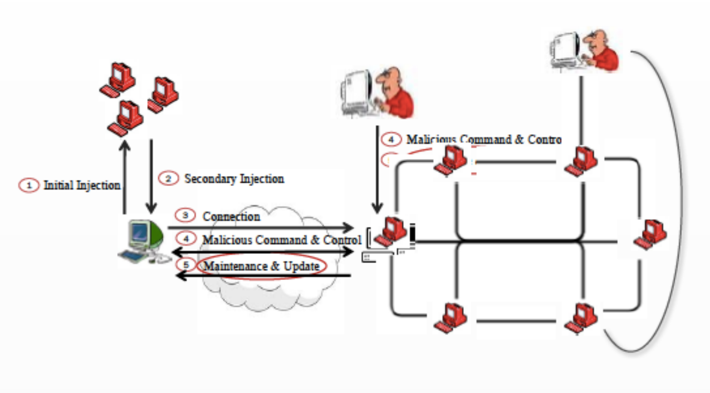
Malicious Programs - Malwares

- Does need host program
  - Trojan horses
  - Logic bombs
  - Viruses*
- Does not need host program
  - Worms*
  - Zombies*

How do computers get infected?

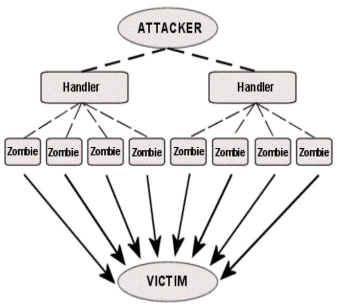
- Attack vectors
  - Web pages
  - Malicious e-mails
  - Attachments
- Payloads
  - Virus
  - Trojan
  - Spyware

Botnets



Distributed Denial of Service (DDoS) – Darknets

- Set of unallocated addresses on a network
- Goal is to collect the attack traffic
- Captured packets might be:
  - Infection attempts
  - Misconfiguration
  - User-based action



Challenges

- ❖ Superposition of behaviours
- ❖ Mix of stationary and non-stationary
- ❖ Diversity
- ❖ Dependencies
- ❖ Dynamics
- ❖ Volume

13

For Evolutionary Computation

- ❖ How to represent data?
- ❖ How to sample training data set?
- ❖ How to represent objectives?
- ❖ How to measure performance?
- ❖ How to incorporate visualization?
- ❖ How much prior knowledge?

14

Network Monitoring

- ❖ Behavioural analysis of network traffic data
  - Packet header
- ❖ Behavioural analysis of application data
  - Packet payload

15

A Fuzzy-Genetic Approach to Network Intrusion Detection

Fries, 2008

Feature Name	*Description*
duration	length of connections (in secs.)
protocol_type	type of protocol
service	network service on destination
src_bytes	number of data bytes from source to destination
dst_bytes	number of data bytes from destination to source
flag	status of connection: normal or error
land	1 if connection is from/to same port
wrong_fragment	number of "wrong" fragments
urgent	number of urgent packets

Feature Name	*Description*
hot	number of "hot" indicators
num_failed_logins	number of failed login attempts
logged_in	1 if successfully logged in
num_compromised	number of "compromised" conditions
root_shell	1 if root_shell is obtained
su_attempted	1 if "su root" command attempted
num_root	number of root accesses
num_file_creations	number of file creation operations
num_shells	number of shell prompts
num_access_files	number of operations on access control files
num_outbound_cmds	number of outbound commands in an ftp session
is_hot_login	1 if login belongs to hot list
is_guest_login	1 if login is a guest

Feature Name	*Description*
count	number of connections to same host in past 2 seconds
	Note: The following features refer to these same-host connections.
serror_rate	% of connections with SYN errors
rerror_rate	% of connections with REJ errors
same_srv_rate	% of connections to same service
diff_srv_rate	% of connections to different services
srv_count	number of connections to same service in past 2 seconds
	Note: The following features refer to these same-service connections.
srv_serror_rate	% of connections with SYN errors
srv_rerror_rate	% of connections with REJ errors
srv_diff_host_rate	% of connections to different hosts

Presents a fuzzy-genetic approach to intrusion detection that is shown to provide performance superior to other GA-based algorithms.

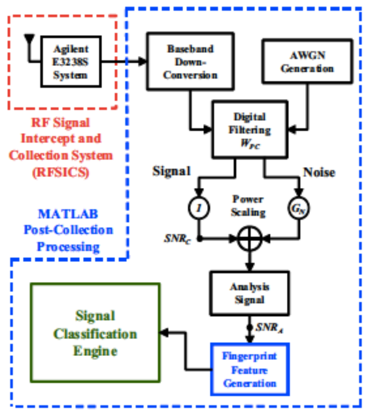
	Intrusion Detection Rate	False Positives
Genetic Clustering	60%	0.4%
Rule Optimization	94%	0%
Fuzzy Inference System	98%	6%
Fuzzy-Genetic IDS	99.6%	0.2%

16

Using Differential Evolution to Optimize 'Learning from Signals' and Enhance Network Security

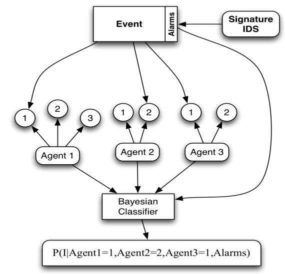
Harmer et al., 2011

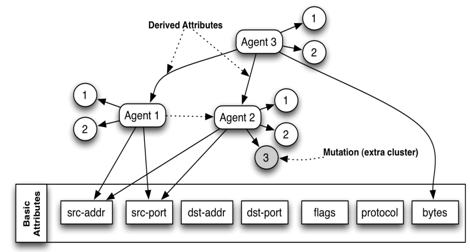
- Aims at developing a physical layer Radio Frequency air monitoring capability to limit unauthorized Wireless Access Point access and improve network security
- uses Differential Evolution to optimize the performance of a "Learning from Signals" (LFS) classifier implemented with Radio Frequency "Distinct Native Attribute" (RF-DNA) fingerprints
- comparative assessment is made using both Time Domain and Spectral Domain fingerprint features



An Evolutionary Multi-Agent Approach to Anomaly Detection and Cyber Defense

Carvalho et al. 2011





- An evolutionary multi-agent approach for anomaly detection based on adaptive clustering and classification.
- An evolutionary algorithm is proposed to allow agents to self-organize and cluster the data using different subsets of attributes, and dynamically created meta-attributes.

Classifying SSH Encrypted Traffic with Minimum Packet Header Features using Genetic Programming

Alshammari et al., 2012

	in-class	out-class
1	tcp.seq	tcp.seq
2	tcp.ack	tcp.ack
3	tcp.flags	tcp.flags
4	ip.flags.df	ip.flags
5		ip.ttl
6		ip.checksum
7		ip.checksum_bad
8		tcp.hd_len
9		tcp.flags.urg
10		tcp.flags.reset
11		tcpwindow_size

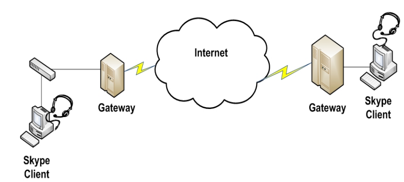
	in-class	out-class
1	frame.cap_len	frame.cap_len
2	tcp.ack	tcp.ack
3	ip.ttl	ip.ttl
4	tcp.window_size	tcp.window_size
5	tcp.seq	tcp.seq
6	tcp.nxtseq	tcp.nxtseq
7	frame.pkt_len	frame.pkt_len
8	tcp.flags	tcp.flags
9	frame.time_delta	frame.time_delta
10	tcp.len	tcp.len
11		ip.flags
12		tcp.flags.fin
13		ip.flags.df

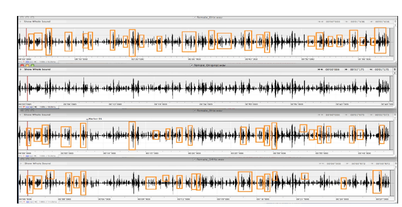
- Investigated the identification of SSH encrypted trac based on packet header features without using IP addresses, port numbers and payload data.
- Evaluation of C4.5, AdaBoost and the Symbiotic Bid-based paradigm of team-based Genetic Programming under data sets common and independent from the training condition indicates that SBB based GP solutions are capable of providing simpler solutions without sacrificing accuracy.

The Impact of Evasion on the Generalization of Machine Learning Algorithms to Classify VoIP Traffic

Alshammari et al., 2013

- Evasion Attacks against VoIP classifiers
- Altered data by padding/morphing
- Altering bit rate
- Altering format
- Shows not easy to evade





On Botnet Behaviour Analysis using GP and C4.5

Haddadi et al., 2014

	Data Set	Score	Legitimate		Botnet		Complexity		
			TPR	FPR	TPR	FPR	Time (sec)	Solution	Feature
C4.5	Zeus-1 (NIMS)	87%	90%	16%	84%	10%	0.24	457	9
	Zeus-2 (NIMS)	97%	97%	3%	97%	3%	0.01	35	9
	Zeus (NETRESEC)	96%	97%	6%	94%	3%	0.01	29	8
	Zeus (Snort)	98%	97%	1%	99%	3%	0	11	5
	Conficker (NIMS)	94%	93%	5%	95%	7%	3.41	365	10
	Torpig (NIMS)	99%	99%	1%	99%	1%	0.04	17	5
SBB	Zeus-1 (NIMS)	78%	73%	18%	82%	27%	188.252	51	8
	Zeus-2 (NIMS)	97%	94%	0%	100%	6%	161.87	14	6
	Zeus (NETRESEC)	90%	87%	7%	93%	13%	36.80	48	8
	Zeus (Snort)	100%	100%	0%	100%	0	8.22	41	8
	Conficker (NIMS)	91%	90%	9%	91%	10%	192.44	41	9
	Torpig (NIMS)	100%	100%	0%	100%	0%	109.23	60	11

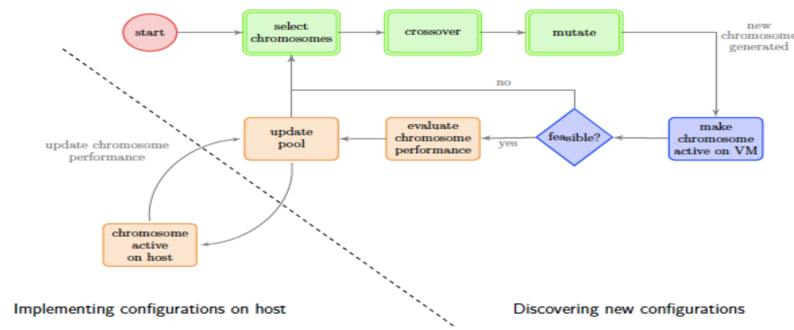
- Genetic programming and Decision trees to detect distinct behaviours in various botnets: Zeus, Conficker, Torpig
- Traffic flows
- HTTP protocol

Softflowd set.1 & 2	Softflowd set.2 only
Duration	Flag-A
Total number of packets (Pkts)	Flag-P
Total number of bytes (Byts)	Flag-R
Flows	Flag-S
Type of Service (TOS)	Flag-F
Bits per second (bps)	Flag-U
Packets per second (pps)	
Bytes per packet (Bpp)	

21

Evolutionary Based Moving Target Cyber Defense

John et al., 2014



Evolution-based algorithms, which formulate better solutions from good solutions, can be used to create a Moving Target Defense. New configurations are created based on the security of previous configurations and can be periodically implemented to change the system's attack surface.

22

Securing the Internet of Things with Responsive Artificial Immune Systems

Greensmith, 2015

- The Internet of Things -- One application is the `smart house', with components including household appliances, networked with the user able to control devices remotely.
- However, the security inherent in these systems is added as somewhat of an afterthought.
- Artificial Immune Systems may be extremely useful.
- Limitations -- focusing on detection without providing automatic responses.
- Opportunity to advance AIS -- A responsive version of the deterministic Dendritic Cell Algorithm is proposed through the incorporation of a model of T-cell responses.

23

Botnet Detection System Analysis on the Effect of Botnet Evolution and Feature Representation

Haddadi et al., 2015

Data Set	Legit	Botnet
CVUT-5	1046254	1046254
Zeus-1 (NIMS)	43460	43460
Zeus-2 (NIMS)	1547	1547
Zeus-3 (NIMS)	40236	40236
Zeus-4 (NIMS)	10678	10678
Zeus (NETRESEC)	401	401
Zeus (Snort)	144	144

- Evaluate genetic programming and decision trees to explore two questions:
- Does the representation of non-numeric features effect the detection rate?
- How long can a machine learning based detection system can perform effectively?

Data sets publicly available at: https://web.cs.dal.ca/~haddadi/data-analysis.htm

	Data Set	Score	Botnet		Legitimate		Complexity	
			TPR	FPR	TNR	FNR	Time (sec)	Solution
C4.5	CVUT-5	99.95%	100%	0.1%	99.9%	0%	2620.01	1199
	Zeus-1 (NIMS)	99.8%	99.8%	0.2%	99.8%	0.2%	26.97	399
	Zeus-2 (NIMS)	99.87%	100%	0.3%	99.7%	0%	0.23	9
	Zeus-3 (NIMS)	100%	100%	0%	100%	0%	12.2	43
	Zeus-4 (NIMS)	99.95%	99.9%	0%	100%	0.1%	2.21	41
	Zeus (NETRESEC)	97.63%	98.0%	2.7%	97.3%	2.0%	0.15	25
	Zeus (Snort)	100%	100%	0%	100%	0%	0.06	5
SBB	CVUT-5	98.66%	99.29%	1.97%	98.03%	0.71%	1047.53	23
	Zeus-1 (NIMS)	98.58%	97.26%	0.1%	99.9%	2.73%	372.256	47
	Zeus-2 (NIMS)	100%	100%	0%	100%	0%	229.486	26
	Zeus-3 (NIMS)	99.99%	99.98%	0%	100%	0.2%	197.21	17
	Zeus-4 (NIMS)	99.98%	100%	0%	99.97%	0%	327.256	67
	Zeus (NETRESEC)	99.17%	98.33%	0%	100%	1.67%	378.048	74
	Zeus (Snort)	100%	100%	0%	100%	0%	147.017	2

24

Benchmarking the Effect of Flow Exporters and Protocol Filters on Botnet Traffic Classification

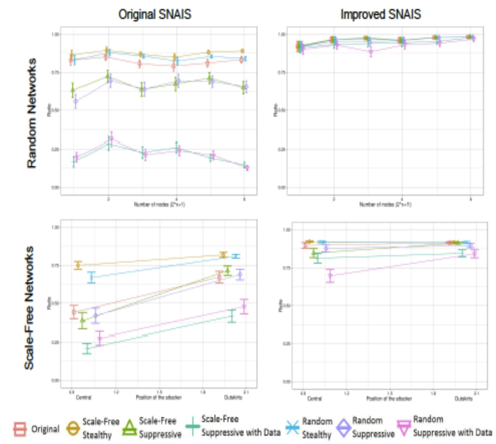
Haddadi et al., 2016

- Botnet traffic analysis
- Using:
  - Five different traffic flow exporters: Tranalyzer, Sfotflowd, Netmate, Yaf and Maji
  - Two different traffic protocol filters: HTTP and DNS
  - Five different classifiers
  - Eight different botnet data sets
- All publicly available at:
  - https://projects.cs.dal.ca/projectx/

Evolving Attackers against Wireless Sensor Networks

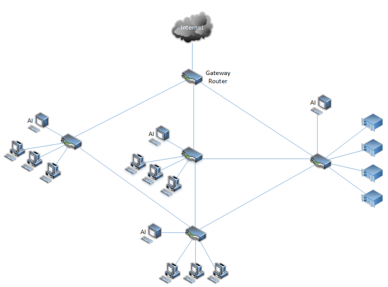
Mrugala et al., 2016

- Genetic Programming to evolve attacks against Internet of Things devices
- Goal is to identify vulnerabilities before systems are attacked
- Used a wireless sensor network setting with an IDS
- The GP attackers succeeded in suppressing significantly more legitimate messages than a hand-coded attacker



Initiating a Moving Target Network Defense with a Real-time Neuro-evolutionary Detector

Smith et al., 2016



Actual Label	Genetic Programming		
	Normal	LAND	IntProp
Normal	92.3 (99.8)	4.0 (0.02)	3.7 (0.14)
LAND	0.27 (0)	99.5 (100)	0.2 (0)
IntProp	0.42 (0)	0.23 (0)	99.3 (100)

Actual Label	NEAT		
	Normal	LAND	IntProp
Normal	80.8 (90.8)	10.7 (0.5)	8.5 (8.7)
LAND	4.5 (10.0)	91.5 (82.1)	4.0 (7.9)
IntProp	15.9 (0.6)	14.5 (6.8)	69.5 (92.6)

- The moving network target defense based approach aims to develop capabilities to dynamically change the attack surfaces, e.g. dynamically change IP addresses
- Denial of Service (LAND attack) and Worms (Internal Propagation) represent examples of attacks
- Evaluated Neuro-Evolution of Augmented Topologies (NEAT) and Genetic Programming represent examples of detectors

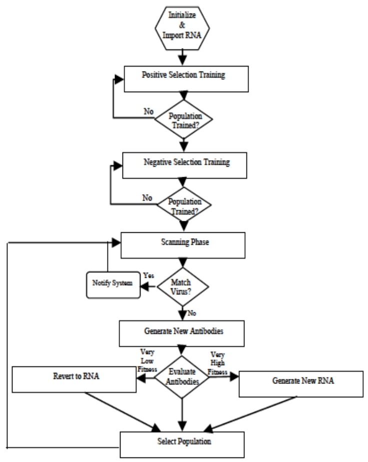
System Monitoring

- User behaviour analysis
  - Applications used
  - Web sites visited
- Device behaviour analysis
  - Syslog files
  - Application log files
  - Sensor log files

A Retrovirus Inspired Algorithm for Virus Detection & Optimization

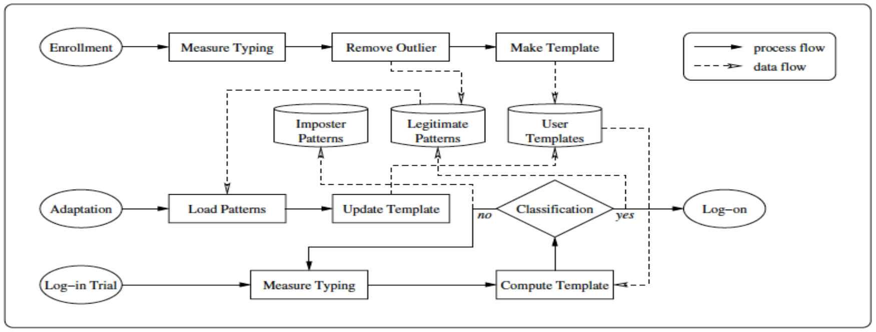
Edge et al. 2006

- Proposes an artificial immune system genetic algorithm (REALGO) based on the human immune system's use of reverse transcription ribonucleic acid (RNA).
- The REALGO algorithm provides memory such that during a complex search the algorithm can revert back to and attempt to mutate in a different "direction" in order to escape local minima.



An Evolutionary Keystroke Authentication Based on Ellipsoidal Hypothesis Space

Lee et al., 2007

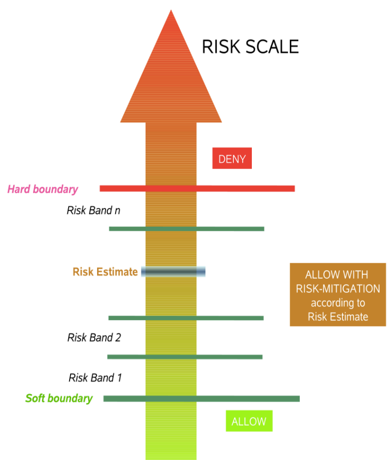


- Keystroke authentication is a biometric method utilizing the typing characteristics of users.
- Proposes an evolutionary method for stable keystroke authentication.

MLS Security Policy Evolution with Genetic Programming

Lim et al., 2008

- investigates how policies can be inferred automatically using Genetic Programming from examples of decisions made.
- This allows to discover a policy that may not formally have been documented, or else extract an underlying set of requirements by interpreting user decisions to posed "what if" scenarios.
- Three proof of concept experiments on MLS Bell-LaPadula, Budgetised MLS and Fuzzy MLS policies have been carried out.



On Evolving Buffer Overflow Attacks Using Genetic Programming

Kayacik et al., 2008

Evolved Program	Core Attack	Sub-goals
PUSH 0x68732f2f		
MUL EAX		
PUSH EBX		
MUL EDX		
CDQ		
CDQ		
SUB EAX, EAX	XOR EAX, EAX	(d)
MUL EDX	CDQ	(d)
PUSH EDX		
MOV CL, 0x0b		
PUSH EDX		
DEC ECX		
DEC ECX		
MOV EBX, ESP		
PUSH 0x6e69622f		
PUSH EDX	PUSH EAX	(a)
PUSH 0x68732f2f	Same	(a)
PUSH 0x6e69622f	Same	(a)
MOV EBX, ESP	Same	(b)
MOV ECX, EDX	PUSH EAX (step 1)	(c)
CDQ		
MUL EDX		
PUSH ECX	PUSH EAX (step 2)	(c)
PUSH EBX	Same	(c)
MOV ECX, ESP	Same	(c)
MOV AL, 0x0b	Same	(e)
INT 0x80	Same	(e)
PUSH EDX		
PUSH 0x6e69622f		
MOV DL, 0x0b		

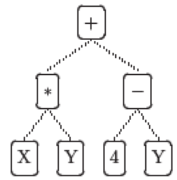
- Employs genetic programming to evolve a "white hat" attacker for providing better detectors
- Variants of buffer overflow attacks
- Appropriate fitness function and partnering instruction set
- intron behavior helps to obfuscate the true intent of the code
- Able to evade Snort intrusion detection system

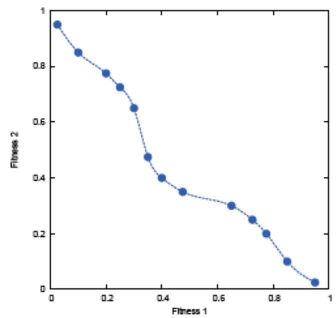
Code and Data Sets at: http://www.kayacik.ca/index.html

Dynamic Security Policy Learning

Lim et al. 2009

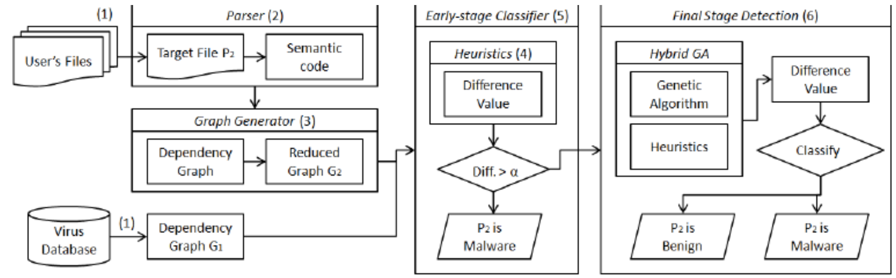
- Research has shown that security policies can be learnt from examples using machine learning techniques.
- Given a set of criteria of concern, one can apply these techniques to learn the policy that best fits the criteria.
- Proposes two dynamic security policy learning frameworks
  - Genetic Programming
  - Multi-Objective Evolutionary Algorithms





Malware Detection based on Dependency Graph using Hybrid Genetic Algorithm

Kim et al., 2010

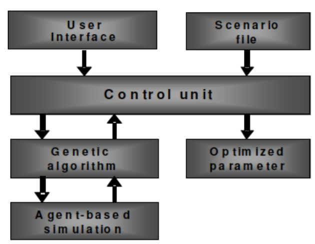


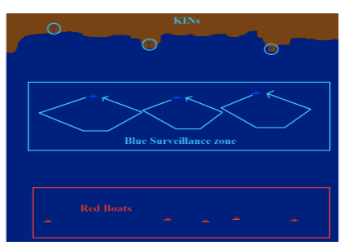
- propose a detector for script malwares, using dependency graph analysis
- present efficient heuristic approaches using for maximum subgraph isomorphism, which improve detection accuracy and reduce computational cost

Analysis of Key Installation Protection using Computerized Red Teaming

Ranjeet et al., 2011

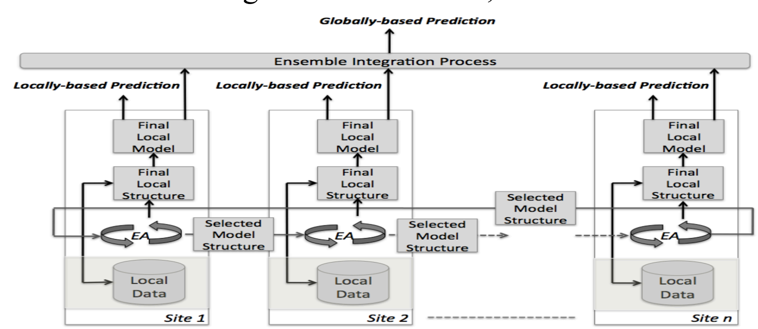
- Use of genetic algorithms for computerized red teaming applications, to explore options for military plans in specific scenarios.
- The proposed technique incorporates a genetic algorithm in conjunction with an agent-based simulation system
- Both enemy forces (the red team) and friendly forces (the blue team) are modelled as intelligent agents and tested on many simulated scenarios
- The aim of these experiments is to explore the red tactics to penetrate a fixed blue patrolling strategy.





Privacy-Preserving Approach to Bayesian Network Structure Learning from Distributed Data

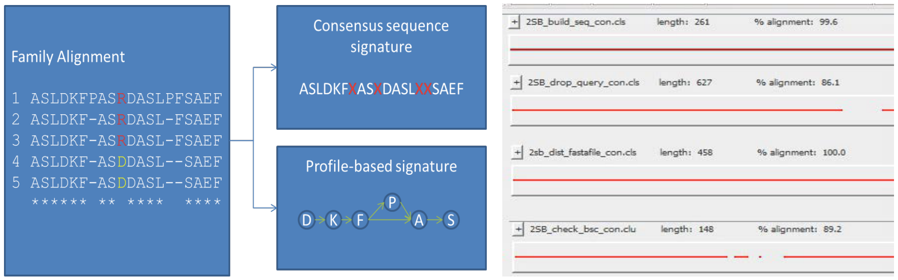
Regnier-Coudert et al., 2011



- present a new approach to learning Bayesian Networks structures from multiple datasets
- based on the use of Ensembles and an Island Model Genetic Algorithm
- Aims to ensure no data is shared during the process

Evolutionary Drift Models for Moving Target Defense

Oehmen et al. 2012



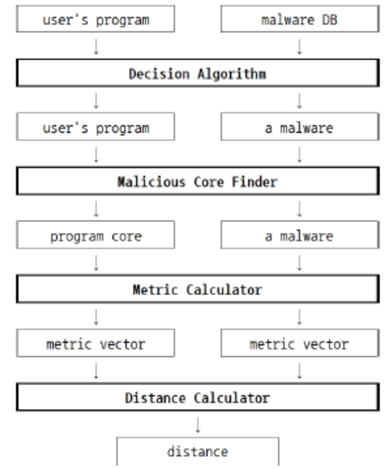
- applied sequence-based and profile-based evolutionary models and report the ability of these models to recognize highly volatile code regions
- "signature" being used to detect sequence-based behaviors is not a fixed signature but one that can recognize new variants of a known family

New Malware Detection System Using Metric-Based Method and Hybrid Genetic Algorithm

Kim et al., 2012

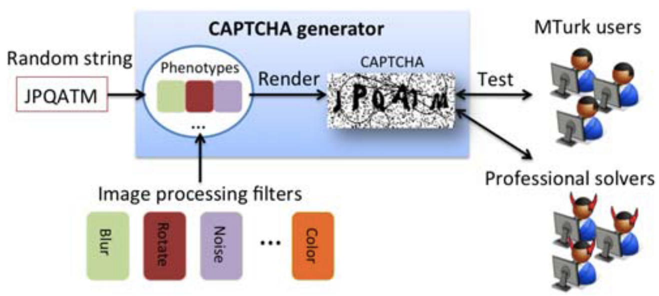
- propose a new approach to detect disguised malware, focusing on the malware scripts
- proposed system consists of a metric-based detection algorithm and a hybrid genetic algorithm.
- The genetic algorithm aims further detection by extracting the main core of a program

	Proposed system	Anti-viruses
Benign codes	100%	98.36%
Known variants	80%	62.79%
Generated malware scripts	100%	34.54%
Overall	92%	58.58%



Darwin: A Ground Truth Agnostic CAPTCHA Generator Using Evolutionary Algorithm

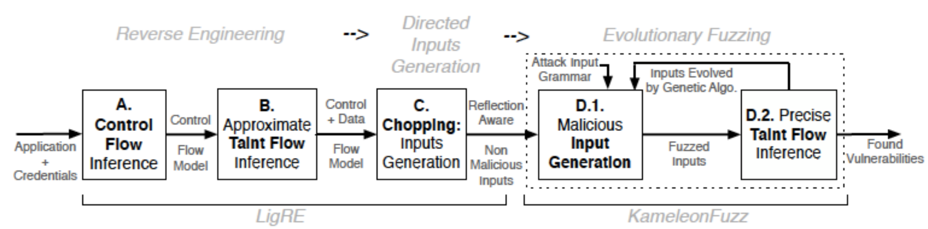
Chen et al, 2014



- CAPTCHA generator using evolutionary algorithm.
- Evaluated with MTurk users (non-attackers) and Antigate workers (attackers).
- Due to the ground-truth agnostic fitness function, discover a new category of CAPTCHAs in which attackers answer correctly but non-attackers answer incorrectly.

KameleonFuzz: Evolutionary Fuzzing for Black-Box XSS Detection

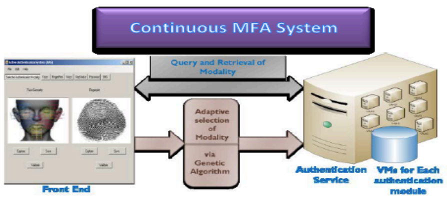
Duchhene et al. 2014

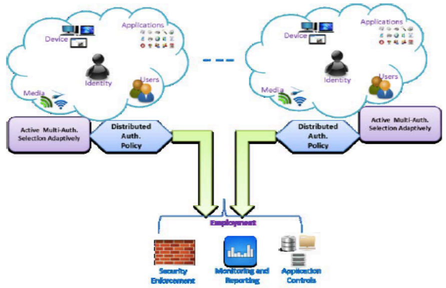


- Fuzz testing consists automatically generating and sending malicious inputs to an application in order to trigger a vulnerability.
- propose KameleonFuzz, a blackbox Cross Site Scripting (XSS) fuzzer for web applications.
- The malicious inputs generation and evolution is achieved with a genetic algorithm, guided by an attack grammar.

An Adaptive Approach for Continuous Multi-factor Authentication in an Identity Eco-System

Nag et al. 2014





- ❖ Multi-factor Authentication (MFA) is the current trend to identify the legitimate users in cyber eco-system through an active authentication process
- ❖ Focus on the design and development of a framework for continuous MFA where authentication modalities are selected adaptively (Genetic Algorithms) through sensing many characteristics of the user's operating environment.

Towards Automated Malware Creation: Code Generation and Code Integration

Cani et al., 2014

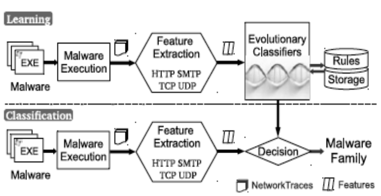
SPLIT.EXE	
offset interval	[0,43000)
Evaluations	300
Type I (zones found)	1
Type I (largest)	334
Type II (zones found)	32
Type II (largest)	1,511

TESTDISK.EXE			
offset interval	[0,43000)	[0,10000)	[0,2000)
Evaluations	15,000	2,000	300
Type I (zones found)	-	1	1
Type I (largest)	-	33	25
Type II (zones found)	3	4	3
Type II (largest)	179	167	183

- ❖ proposes two different ways for exploiting an evolutionary algorithm to devise malware:
- ❖ the former targeting heuristic-based anti-virus scanner;
- ❖ the latter optimizing a Trojan attack.

Evolutionary Algorithms for Classification of Malware Families through Different Network Behaviors

Rafique et al., 2014



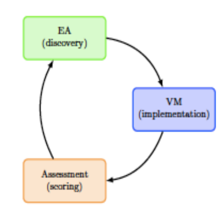
- ❖ malware family classification system that models the protocol-aware and state-space features
- ❖ a comprehensive study of 4 evolutionary and 4 non-evolutionary classification algorithms

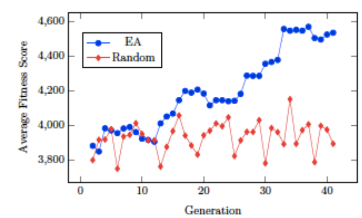
Classifiers	Evolutionary								Non-Evolutionary							
	GAssist-ADI		SLAVE		UCS		XCS		C4.5		C-SVM		KNN		NB	
Malware Families	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
1.Cleaman	9.98	10.20	0	0	100	100	0	0	0	0	100	100	0	0	100	100
2.Coinminer	9.72	12.50	0	0	100	100	0	0	100	100	100	100	100	100	27.78	0
3.Cridex	99.92	99.83	0	0	100	100	51	51.23	100	100	100	94.55	100	100	100	100
4.Cutwail	56.85	33.30	22.22	20	94.75	88.87	1	2.56	64.44	66.67	95.19	69.88	59.75	60	65	65
5.Drstwex.A	100	100	73.33	72.92	100	100	20.70	21.57	100	100	98.04	98.04	100	100	96.08	96.08
6.Foreign	7.41	0	0	0	72.22	83.33	0	0	0	0	100	100	66.67	66.67	0	0
7.Malagent	0	0	0	0	100	73.33	0	0	0	0	100	40	100	100	100	100
8.Onescan	50.20	48.24	0	0	98.17	76.47	0	0	100	100	100	76.47	44.71	44.71	100	98.82
9.Qakbot-AE	41.71	41.54	0	0	100	95.38	0	0	100	98.46	100	95.38	99.66	100	100	100
10.Ramnit	59.26	66.67	100	100	99.07	91.67	0	0	100	100	73.15	70.83	33.33	33.33	58.33	58.33
11.Simda	5	8.33	0	0	96.85	8.33	3.33	5	100	100	100	0	63.70	63.33	100	100
12.Spybot.bfr	54.94	55.56	50	50	100	100	50	50	100	100	100	100	100	100	100	100
13.Spyeye	0	0	0	0	95.96	66.19	1.52	0	98.30	73.86	95.62	79.17	76.33	78.69	24.35	14.39
14.Suspectcrc	68.46	67.71	0	0	97.22	79.17	0.75	0	37.73	35.94	100	64.06	84.84	83.85	99.48	98.96
15.Waledace.C	0	0	0	0	82.54	64.29	0	0	0	0	100	78.57	45.24	50	0	0
16.Waledace.R	2.68	3.45	0	0	93.10	79.31	0	0	0	0	100	62.07	69.73	72.41	100	62.07
17.Webprotection	19.61	23.53	0	0	100	100	0	0	0	0	100	100	0	0	70.59	70.59
18.Winwebsec	99.98	99.88	0	0	99.97	99.35	86.64	86.72	100	99.97	100	99.82	99.91	99.85	100	100
19.Zbot	99.97	99.96	100	100	100	99.99	96.84	96.96	100	100	100	99.57	99.96	99.96	99.83	99.79
20.Zeroaccess	99.77	99.72	49.49	49.49	100	99.88	48.98	98.61	100	100	100	98.93	99.45	99.45	99.33	99.33
Avg.(per Family)	44.27	43.52	17.82	17.71	96.49	85.28	18.04	20.63	65.02	63.75	98.10	81.37	72.16	72.61	77.04	73.17
Avg.(Samples)	99.14	99.19	84.90	84.91	99.96	99.70	94.53	94.53	99.49	99.42	99.99	99.01	99.56	99.55	99.73	99.70

An Initial Framework for Evolving Computer Configurations as a Moving Target Defense

Lucas et al. 2014

- ❖ describes an initial Python-based framework that creates an evolutionary inspired MTD for computers.
- ❖ The framework consists of three interacting components:
  - An evolutionary component discovers computer configurations based on previous configurations.
  - A second component vets new configurations by instantiating them using virtual machines.
  - A third component uses a combination of penetration software as well as reports from actual attacks to assess the configurations.

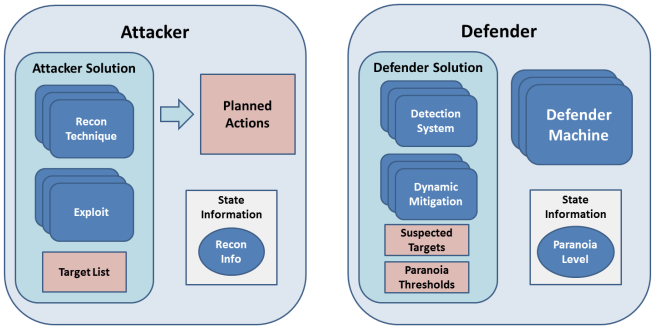




(a) Average population fitness.

Coevolutionary Agent-based Network Defense Lightweight Event System (CANDLES)

Rush et al., 2015

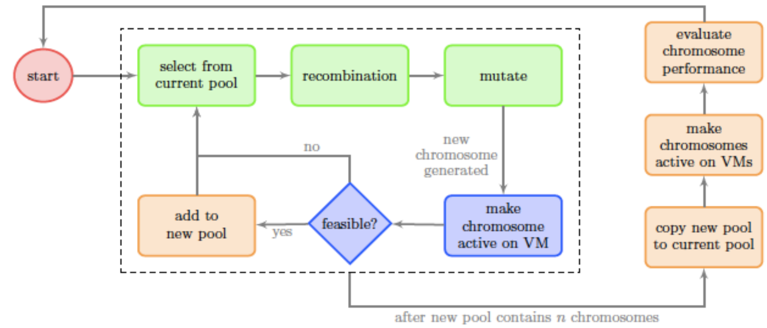


- a framework designed to coevolve attacker and defender agent strategies
- provide a proof of concept for the applicability of coevolution in planning for, and defending against, novel attacker strategies in computer network security

45

Using Probability Densities to Evolve more Secure Software Configurations

Odell et al. 2015



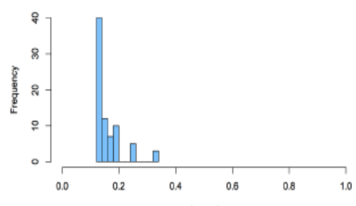
- use of Evolutionary Algorithms (EAs) is one method for securing software configurations in a changing environment.
- configurations are modeled as biological chromosomes, and a continual sequence of selection, recombination, and mutation processes is performed.

46

Malware Obfuscation through Evolutionary Packers

Gaudesi et al., 2015

- Describes a new obfuscation mechanism based on evolutionary algorithms
- an evolutionary core is embedded in the malware to generate a different, optimized hiding strategy for every single infection
- Such always-changing, hard-to-detect malware can be used by security industries to stress the analysis methodologies and to test the ability to react to malware mutations.

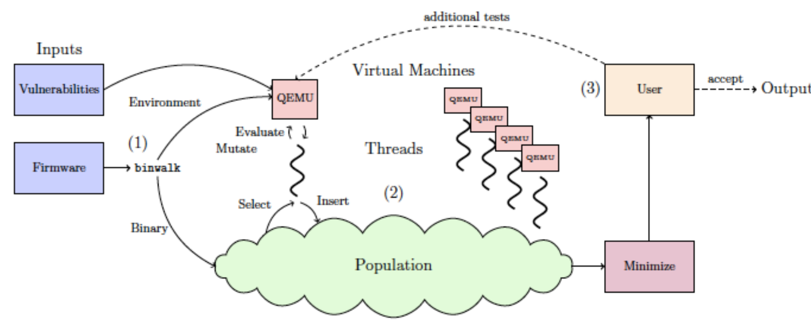


	Uncoded	Evo 1	Evo 2	Evo 3
Virus Total	35/57	2/57	2/57	1/57
Metascan Online	25/44	4/44	3/44	1/44

47

Repairing COTS Router Firmware without Access to Source Code or Test Suites: A Case Study in Evolutionary Software Repair

Schulte et al., 2015

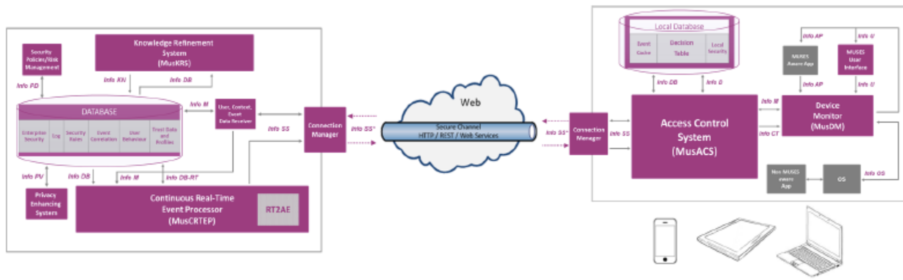


- propose a solution: an interactive evolutionary algorithm that searches for patches that resolve target vulnerabilities while relying heavily on post-evolution difference minimization to remove most regressions
- approach does not require access to source code, regression tests, or any participation from the software vendor.

48

Enforcing Corporate Security Policies via Computational Intelligence Techniques

Mora et al., 2014

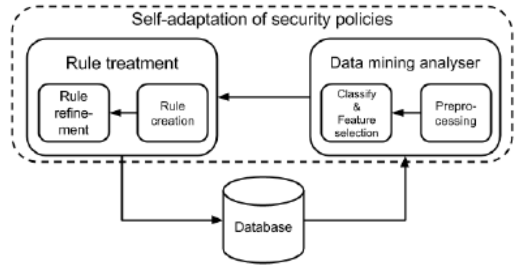


❖ Aims to analyse the user’s behaviour (modelled as events) when interacting with the company’s server, accessing to corporate assets, for instance.

❖ As a result -- Corporate Security Policies will be adapted to deal with new anomalous situations, or to better manage user’s behaviour.

Soft Computing Techniques Applied to Corporate and Personal Security

de las Cuevas et al., 2015



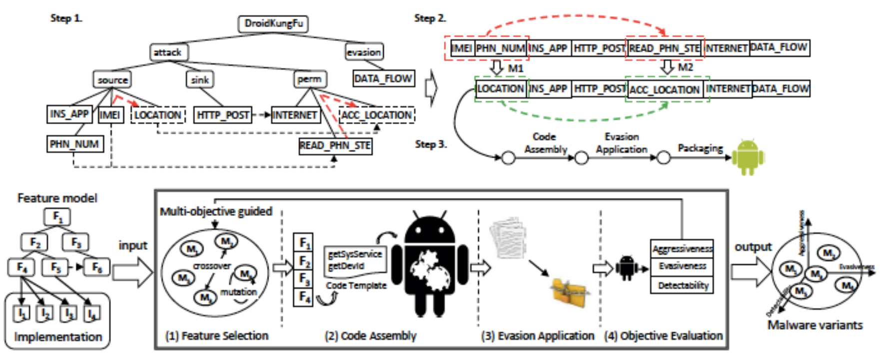
❖ Inside a \Bring Your Own Device" environment -- a new situation is risky or not?

❖ proposes the use of a variety of techniques from Data Mining to Evolutionary Algorithms for refining a set of existing securitypolicies

❖ Case study – URL access lists

Mystique: Evolving Android Malware for Auditing Anti-Malware Tools

Meng et al, 2016



❖ Propose a meta model for Android malware to capture the common attack features and evasion features in the malware.

❖ Develop a framework, MYSTIQUE, to automatically generate malware covering four attack features and two evasion features, by adopting the software product line engineering approach.

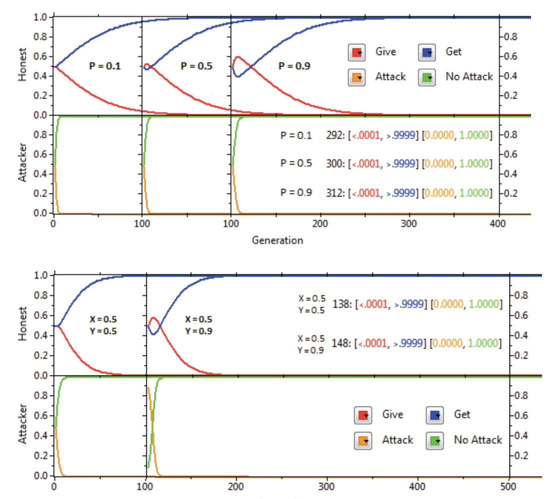
Solving Sybil Attacks Using Evolutionary Game Theory

Saab et al., 2016

❖ Recommender systems are vulnerable to several types of attacks that target user ratings.

❖ One such attack is the Sybil attack where an entity masquerades as several identities with the intention of diverting user ratings.

❖ Propose evolutionary game theory as a possible solution to the Sybil attack in recommender systems.



Streaming Data Analysis

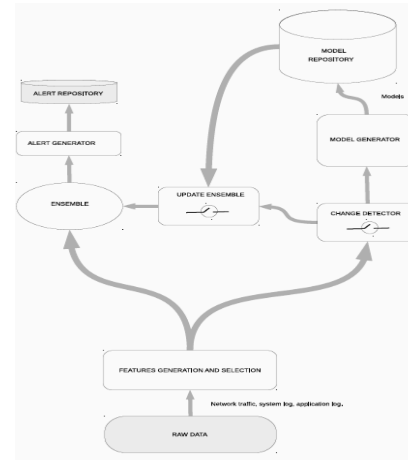
- Challenges not available in offline methods
- No training and test partitions
- Several champions throughout the stream - anytime operation
- Labelling is expensive - limited label budget
- Non-stationary processes
  - Sudden shifts or gradual drifts
  - Imbalance class distributions

53

An Incremental Ensemble Evolved by using Genetic Programming to Efficiently Detect Drifts in Cyber Security Dataset

Folino et al., 2016

- Unbalanced classes, the ability to detect changes in real-time, the speed of the streams – challenges with cyber security datasets
- To overcome these issues, they propose an ensemble-based algorithm, using a distributed Genetic Programming framework to generate the function to combine the classifiers and efficient strategies to react to changes in datasets

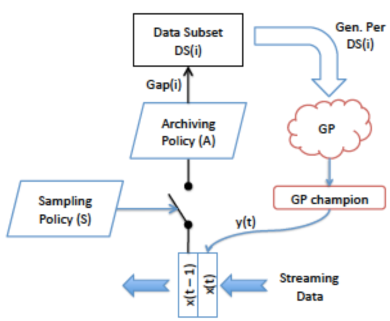


54

Properties of a GP Active Learning Framework for Streaming Data with Class Imbalance

Khanchi et al., 2017

- Genetic Programming (GP) active learning as applied to streaming data
- fitness evaluation is performed against a data subset
- also investigate the capability of the framework to actively balance (or not) the distribution of exemplars in the data subset during the course of the stream



55

DEMOS

56

DISCUSSION

57

Tranalyzer
Burschka et al., 2008

- Flow based forensic and network troubleshooting traffic analyzer
- URL: https://tranalyzer.com

58

Argus
qosient.com

- Network flow generator and traffic analysis
- URL: https://qosient.com/argus/

59

Softflowd
Miller et al., 2005

- Flow-based network traffic analyzer capable of Cisco NetFlow data export
- URL: https://code.google.com/archive/p/softflowd/
- Moved to Google code in 2011

60

NetMate FlowCalc
Arndt et al., 2011

- NetMate flow exporter, extended NetAI module by Dalhousie NIMS Lab, 2011: URL: https://dan.arndt.ca/projects/netmate-flowcalc/
- Originally by Zander et al., 2005 : https://sourceforge.net/projects/netmate-meter/files//netmate-meter/netmate-0.9.5/netmate-0.9.5-ChangeLog/view

61

Publicly Available Data Sets

- CAIDA: http://www.caida.org/data/
- Dalhousie University NIMS Lab: https://projects.cs.dal.ca/projectx/
- CTU-13: http://mcfp.weebly.com/the-ctu-13-dataset-a-labeled-dataset-with-botnet-normal-and-background-traffic.html
- Uvic: http://www.uvic.ca/engineering/ece/isot/datasets/
- UNB: http://www.unb.ca/cic/research/datasets/botnet.html
- Kayacik: http://www.kayacik.ca/data.html
- Rice University LiveLab: http://livelab.recg.rice.edu
- MIT Human dynamics Lab: http://realitycommons.media.mit.edu/realitymining.html
- MIT Lincoln Lab: https://www.ll.mit.edu/ideval/data/

62

Open Source Monitoring and CyberSecurity tools

- Wireshark: https://www.wireshark.org
- TcpDump: http://www.tcpdump.org
- TcpReplay: http://tcpreplay.appneta.com
- Snort: http://realitycommons.media.mit.edu/realitymining.html
- Bro: https://www.bro.org
- Corsaro: http://www.caida.org/tools/measurement/corsaro/
- Iatmon: http://www.caida.org/tools/measurement/iatmon/
- CVSS – NVS: https://nvd.nist.gov/vuln-metrics/cvss
- Nmap: https://nmap.org
- Kali: https://www.kali.org
- Metasploit: https://www.metasploit.com
- Argus: https://qosient.com/argus/
- Security tools: http://sectools.org

63

Acknowledgements

- The content of this tutorial has benefited from a host of collaborations over the years including, but not limited to:
  - All past and present Dalhousie University NIMS Lab members
- NZH would like to acknowledge the funding for aspects of research reported in this tutorial from the NSERC, DRDC and Raytheon.

64

References

- Alshammari, R., Lichodzijewski, P., Heywood, M. I., and Zincir-Heywood, A. N., "Classifying SSH Encrypted Traffic with Minimum Packet Header Features Using Genetic Programming", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 2539-2546, 2009
- Alshammari, R., and Zincir-Heywood, A. N., "How Robust Can a Machine Learning Approach be for Classifying Encrypted VoIP?", *Network system Management*, 23(4), pp. 830-869, 2015
- Cani, A., Gaudesi, M., Sanchez, E., Squillero, G., and Tonda, A., "Towards Automated Malware Creation: Code Generation and Code Integration", *Proceedings of the Symposium on Applied Computing,* pp. 157-158, 2014
- Carvalho, M., and Perez, C., "An Evolutionary Multi-Agent Approach to Anomaly Detection and Cyber Defense", *Proceedings of the Cyber Security and Information Intelligence Research Workshop*, pp. 1-4, 2011

References

- Chen, E. Y., Huang, L., Mengshoel, O. J., and Lohn, J. D., "Darwin: A Ground Truth Agnostic CAPTCHA Generator Using Evolutionary Algorithm", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 165-166, 2014
- De las Cuevas, P., Merelo, J. J., and Garcia-Sanchez, P., "Soft Computing Techniques Applies to Corporate and Personal Security", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1193-1196, 2015
- Duchene, F., Rawat, S., Richier, J., and Groz, R., "KameleonFuzz: Evolutionary Fuzzing for Black-Box XSS Detection", *Proceedings of the Conference on Data and Application Security and Privacy,* pp. 1-12, 2014
- Edge, K. S., Lamont, G. B., and Raines, R. A., "A Retrovirus Inspired Algorithm for Virus Detection and Optimization", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 103-110, 2006

References

- Folino, G., Pisani, F. S., and Sabationo, P., "An Incremental Ensamble Evolved by using Genetic Programming to Efficiently Detect Drifts in Cyber Security Datasets", *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 1103-1110, 2016
- Fries, T. P., "A Fuzzy-Genetic Approach to Network Intrusion Detection", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 2141-2146, 2008
- Gaudesi, M., Marcelli, A., Sanchez, E., Squillero, G., and Tonda, A., "Malware Obfuscation Through Evolutionary Packers", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 757-758, 2015
- Greensmith, J., "Securing the Internet of Things with Responsive Artificial Immune Systems", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 113-120, 2015

References

- Haddadi, F., Runkel, D., Zincir-Heywood, A. N., and Heywood, M. I., "On Botnet Behaviour Analysis using GP and C4.5", *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 1253-1260, 2014
- Haddadi, F., and Zincir-Heywood, A. N., "Botnet Detection System Analysis on the Effect of Botnet Evolution and Feature Representation", *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 893-900, 2015
- Haddadi, F., and Zincir-Heywood, A. N., "Benchmarking the Effect of Flow Exporters and Protocol Filters on Botnet Traffic Classification", *IEEE Systems Journal,* 10(4), pp. 1390-1401, 2016
- Harmer, P. K., Temple, M. A., Buckner, M. A., and Farquahar, E., "Using Differential Evolution to Optimize 'Learning from Signals' and Enhance Network Security", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1811-1818, 2011

References

- John, D., Smith, R. W., Turkett, W. H., Canas, D. A., and Fulp, E. W., "Evolutionary based Moving Target Cyber Defense", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1261-1268, 2014
- Kayacik, H. G., Heywood M. I., and Zincir-Heywood, A. N., "On Evolving Buffer Overflow Attacks using Genetic Programming", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1667-1673, 2006
- Kayacik, H. G., Zincir-Heywood, A. N., and Heywood M. I., "Can a Good Offense be a Good Defense? Vulnerability Testing of Anomaly Detectors through an Artificial Arms Race", *Applied Soft Computing,* 11(7), pp. 4366-4383, 2011
- Khanchi, S., Heywood, M. I., and Zincir-Heywood, A. N., "Properties of a GP Active Learning Framework for Streaming Data with Class Imbalance", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp., 2017

References

- Kim, K., and Moon, B., "Malware Detection based on Dependency Graph using Hybrid Genetic algorithm", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1211-1218, 2010
- Kim, J., and Moon, B., "New Malware Detection System using Metric-Based Method and Hybrid Genetic Algorithm", *Proceedings of the Genetic and Evolutionary Computation Conference Companion,* pp. 1527-1528, 2012
- Lee, J., Choi, S., and Moon, B., "An Evolutionary Keystroke Authentication Based on Ellipsoidal Hypothesis Space", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 2090-2097, 2007
- Lim, Y. T., Cheng, P., Rohatgi, P., and Clark, J., "MLS Security Policy Evolution with Genetic Programming", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1571-1578, 2008
- Lim, Y. T., Cheng, P., Rohatgi, P., and Clark, J., "Dynamic Security Policy Learning", *Proceedings of the Workshop on Information Security Governance,* pp. 39-48, 2009

References

- Lucas, B., Fulp, E. W., John, D. J., and Canas D., "An Initial Framework for Evolving Computer Configurations as a Moving Target Defense", *Proceedings of the Cyber and Information Security Research Conference,* pp. 69-72, 2014
- Meng, G., Xue, Y., Mahinthan, C., and Narayanan, A., "Mystique: Evolving Android Malware for Auditing Anti-Malware Tools", *Proceedings of the Asia Conference on Computer and Communications Security,* pp. 365-376, 2016
- Mora, A. M., de las Cuevas, P., Merelo, J. J., Zamarripa, S., and Esparcia-Alcazar, A. I., "Enforcing Corporate Security Policies via Computational Intelligence Techniques", *Proceedings of the Genetic and Evolutionary Computation Conference,* pp. 1245-1251, 2014
- Mrugala, K., Tuptuk, N., and Hailes, S., "Evolving Attackers against Wireless Sensor Networks", *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 107-108, 2016

References

- Nag, A. K., and Dasgupta, D., "An Adaptive Approach for Continuous Multifactor Authentication In an Identity Eco-System", *Proceedings of the Cyber and Information Security Research Conference,* pp. 65-68, 2014
- Odell, C. A., McNiece, M. R., Gage, S. K., Gage, H. D., and Fulp, E. W., "Using Probability Densities to Evolve More Secure Software Configurations", *Proceedings of the Workshop on Automated Decision Making for Active Cyber Defense,* pp. 27-32, 2015
- Oehmen, C., Peterson, E., and Dowson, S., "An Organic Model for Detecting Cyber Events", *Proceedings of the Cyber Security and Information Intelligence Research Workshop,* pp. 1-4, 2010
- Oehmen, C., Peterson, E., and Teuton, J., "Evolutionary Drift Models for Moving Target Defense", *Proceedings of the Cyber Security and Information Intelligence Research Workshop,* pp. 1-4, 2012

References

- Rafique, M. Z., Chen, P., Huygens, C., and Joosen, W., "Evolutionary Algorithms for Classification of Malware Families through Different Network Behaviours", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1167-1174, 2014
- Ranjeet, T. R., Hingston, P., Lam, C., and Masek, M., "Analysis of Key Installation Protection using Computerized Red Teaming", *Proceedings of the Australasian Computer Science Conference,* pp. 137-144, 2011
- Regnier-Coudert, O., and McCall, J., "Privacy Preserving Approach to Bayesian Network Structure Learning from Distributed Data", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 815-816, 2011
- Rush, G., Tauritz, D. R., and Kent, A. D., "Coevolutionary Agent based Network Defense Lightweight Event System (CANDLES)", *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 859-866, 2015

References

- Saab, F., Kayssi, A., Elhajj, I., and Chehab, A., "Solving Sybill Attacks Using Evolutionary Game Theory", *Proceedings of the Symposium on Applied Computing,* pp. 2195-2201, 2016
- Schulte, E., Weimer, W., and Forrest, S., "Repairing COTS Router Firmware without Access to Source Code or Test Suites: A Case Study in Evolutionary Software Repair", *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 847-854, 2015
- Smith, R. J., Zincir-Heywood, A. N., Heywood, M. I., and Jacobs, J. T., "Initiating a Moving Target Network Defense with a Real Time Neuro-Evolutionary Detector", *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 1095-1102, 2016