

Machine Learning for Enterprise Security

Pratyusa K. Manadhata Hewlett Packard Labs manadhata@hpe.com

1 © Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.



Evolution of	Uses of machine	Strengths and	Trends and research					
enterprise security	learning	weaknesses	opportunities					
Focus on problems, not on ML techniques								



Enterprise security



A Security framework









© Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

1st Generation: Point products



Sev.	Sensor	Source IP	Destination IP	Event Signature	Timestamp
2	qa-eth0:eth0	172.16.116.234	DE 217.160.	51.31 ET POLICY curl User-Agent Outbound	7:41 PM
2	qa-eth0:eth0	217.160.51.31	172.16.11	6.234 GPL ATTACK_RESPONSE id check returned root	7:41 PM

2nd Generation: Security information and event management systems (SIEM)

De Cent	Converter Converter and	Code Status) 😳 Debbase Performer In Events Last Inner	ka Shitaka 🔯 Ardight Dav Sieta	(Ellerit Trouppet)	Stift Lotes Information (1) Int	View Exercise Land Houre	Total Events: 5,998	fort inpacts						
ami	Mark Series - New 2013 State field Times - New 2013 State Filters:	ana ang Indonesi Ang Ang Ang Ang Ang Ang Ang Ang Ang Ang						Orryteen Carrisonaded	Source IP	Destination IP	Event Signature			
ight Adhenistretion ight Poundation ight Solutions	Index Piller in Pile						time 1.000				ЛИМ			
igH System Shet	Ratar							-	- 172.16.116.2	84 - 1 2 10 3	1.3 V PC DY	iri U -4 in Strou	und	
onar k signed														
	I Henager Recopt Time &	tane 8	The Same	Taget Law Name 2	Single Attacher Target	Callegery Hest Application: Exercite Responses: Auccess	Parts 2	4	DE 217.160.51	.31 - 1/2.16.116	234 GPLATIA	K_RESPONSE Id check	k returned root	
	*340-3013 DR-36-30 POT	Tap value count data mentor value to				Artenational (Recents 24		Event Desals Avenues						
	4 Sep 2013 (Hotel all PUT	Top-selve count data menter value su	-			Pretikpicatur Bunculu/Response Success Antonatural Reserve 34		- Inter Inter						
	4 Sep 2010 294 06: 20 PDT	Tap value count data montor value lut				Peet, Applicature: Biological Response: Auconst	6	- Israi A		TTC				
						Informational / Deserty 21 Prost. Registration: Rescenter Rescenter	-	Name Query Russing Time						
	*140 2012 De De DE DE TOT	Tap value courd data menter value ou	ha i			Anternational Advanta bet		Type Consistent And Taxe Type 2010 (2010) PCT						
	+Sep.2010.00.00.00.PDF	Top value court data monter value cu	-			Pret/Application (Eventual) Response (Success) Informational (Security Informational		Application Produced						
	+1an 2012 (MIR) (5 PC*	ACCURATE AND ADDRESS	Dary Lowes Tree		- ARADIN	Peet, Application (Multhy Camilguration /Success		Turreport Pretacal Untercability Resource						
					(bdre-ean.halinp.com)	Janual Japania Institution Math. Partnet Partner		Settes In				i di i s s		
	+5ep 2013 10 00 15 PDT	Additional Success	Query Running Time		→ 15.25.117.25	Menatorial Jupices	(form on Design Complement Tem	Source	Des No	re Si at			
	CONTRACTOR OF	COMPOSED IN A	0.0755029.11	1		Peeckploter Publicarligation Success	1	Cutme feaure	ooul ii					
	A DECKING AND A	ACTVELUE OF THY MORE	Onu i ground peak		-> 13-23-127-29 Brite-em/tel/br.com	Annal Apploits		- Cenar						
	4 Sep 3012 SHORE 15 PDF	Additional Success	Query Rurring Time		- 3.31.17.39	Institution Medification Success	6	Denier D Denier URI	- 172.16.116.1	234 21 60		curren A nt too	und	
				-	(Dore-ean.rpl.rp.com)	Anthonistic Auth-Conformation Success		Denue Extend D						
	*See 2013 DR DR 15 PDT	Activitiant entity address	Query Running Tana		-> (5.25.117.29 0dm-em/splay.com)	Partial Application		Convert Ressure te						
	A Designment of the local states	And the Advance	Diana Roman Terra		A 18.78 UT 18	Peel/Application /Medfly/Content /Success	-	Appropriate (Trent C., 1	ns 17	2 116	RA DE ALA	RESPONSE id cher	ck returned root	
	Sector sector	A presentation of the		1	(Index wan hpl.7p.com)	(Monetonel (Application		Converse Event Co 1						
	+Sep 310 24 36 15 PCF	Activitial erroy willed	Query fluoring Time		-> 18.28.117.28	Peet, Approxime (ModPy Carligoration /Success		- Category						
			Concernance in the	-	(bde-een.hpDp.com)	Instantion Built Control Ductors		Category between						
	+Sec 2013 20100, 15 Pp?	Additional Ducines	Query Running Term	1 1	+ 15.25.117.25	Mention Dekiller	6	Calegory Technique						
		and the second se	Contraction (Institution MultiContent Success		Cohegory Device Group						
	Colo A Crock Crock	AD D.F. LCH	don't round me	1 1	(10 00 00 00 00 00 00 00 00 00 00 00 00 0	Anternetianal Application	-	Category Colores						
	4549 2012 2010 2010 15 PCT	Activation to added	Dues Running Time		-> 18.28.117.25	Phot/Applicature /HedPy/Configuration /Success	0	Category Tupe Desi						
	a construction of the			1	(hos-qr/lq/ma-a40d)	7014 34050	·	threat						
	+Se 203 (NOK (2707	QueryTenier QueryTurcentel	Correctors - Dropping Events	the S	15-25-126-294 (15-25-136-294) ->		C	Incento 0						
					13.25.117.25 (004-404.56.56.50 day)			Reference 10						
	*Sep 2110 19(36 (2 PD7	Query Henrer Query Succeeded	Convectors - Down - Shard Terri	Max.	15.25 (16.104 (15.25 (16.304) -> 15.25 (17.25 Julie min by An cont		1	Asset Orloadly 0						
		A LOS AND ALL			CONTRACTOR & S.			Parts 3						
	4.240 30 D 100 R 15 PCF	Query Running Stee	Conectors - Dropping theme				C	+ April						





3rd Generation: Security Operations Centers



1st Generation Point Products



Do one thing and do it "well"

- Anti-malware products
- Data leakage prevention products
- Firewalls
- Application firewalls
- ...

- Intrusion detection systems (IDS/IPS)
- Domain/IP Blacklists
- Sandboxes
- Web proxies
-



Product structure





Anti-Malware products





Sample labelling

Similarity

Machine Learning: Detect similarity at scale



12 © Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Similarity is key in other detection mechanisms too

No "true" indicators of good/bad exist



13 © Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Static analysis: looks similarity



- Features: PE Headers, instruction sequences,...
- Classification/Clustering
- False positives
- False negatives easy to evade

Dynamic analysis: behavior similarity



- Features: system call sequences, API sequences, ...
- Classification/clustering
- Computational challenges
- Hard to run samples

Reputation analysis: behavior similarity



- Program behavior data from end points, no need for samples
- Features: popularity, source, destinations,...
- Lightweight data collection
- Lagging detection



Front end detection: False positive-false negative trade-off

File hash

Strings

Behavior signatures

Almost no machine learning on end devices



Complex and **Fragile**

Malicious domain detection

Many Methods/Papers	Features						
 Classification/clustering/regression 	 Syntactic properties of a domain name string 						
Graph analysis	 HTTP/DNS protocol properties 						
 Pattern mining and matching 	Access patterns						
Statistical analysis	 Registration information 						
•	 Association with malware 						
	•						

A command & control domain used in the OPM breach

opmsecurity.org

source: https://www.threatconnect.com/opm-breach-analysis/



Observations and opportunities

ML helps with scalability, but not much with accuracy

Complex and fragile systems

ML on end devices instead of signature matching



2nd Generation Security Information and Event Management



Correlation



Reduce alerts by grouping Reduce false alarms



Security Information and Event Management





How to generate **correlation rules**?

Market-basket Analysis







Observations

Rules are heuristics driven/manually generated

Correlation window order of minutes



3rd Generation Security Operations Centers

A large enterprise network



Reducing false negatives





Advanced persistent threat (APT) detection





The data analysis approach holds promise

- Compromised account detection
- Lateral movement detection
- Anomalous user behavior
- Insider threat
- Preemptive detection detecting early stages of an attack

Security Analytics









Scalable, Reliable, and Timely Detection



Challenges





4th Generation Remediation & Recovery

Analyst sees an alert

192.168.0.23:43987 -> 203.45.65.201:1433 SQL Injection Attack 23Mar09 1930:003 user=Calvert

Analyst builds a context





Analyst follows a remediation plan

Quarantine the infected machine

Schedule/ run clean up tools Schedule/ run reimaging





A few minutes per alert



© Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Target Ignored Data Breach Alarms

Target's security team reviewed -- and ignored -- urgent warnings from threat-detection tool about unknown malware spotted on the network.

http://www.darkreading.com/attacks-and-breaches/target-ignored-data-breach-alarms/d/d-id/1127712

SOCs don't scale

Repetitive, Manual, and Error Prone

Machine learning to the rescue???





45 © Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Context building as a learning problem

Predict the information needed for each alert

- alert characteristics
- source/destination
- protocol
- vulnerability
- ...

Remediation plan as another learning problem

Predict remediation action for each alert

- context information
- device characteristics
- user characteristics
- location
- connectivity
- business needs



Incorporate analyst feedback



Summary

Current state

- ML has played a key role since early days of enterprise security
- ML helps achieve scale, but FPs/FNs remain
- One of the better tools at our disposal for enterprise security

Future: ML targeted toward enterprise security





Point products: Old problems require new solutions

Targeted new techniques

False positives/False negatives

Resource aware – power

Advances in hardware and systems software



Security analytics and remediation: New problems

Targeted ML techniques for scalable and reliable detection

Holistic approach to detection and remediation

Incorporate user feedback



And finally, be judicious in using ML





Thank You

Pratyusa K. Manadhata manadhata@hpe.com



55 © Copyright 2015 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.