

Deep Machine Learning Meets Cybersecurity

The Problem



- * Malware growing exponentially
 - * Over 100K malware variants created every hour
- * Cyber defense is a big data problem
- * Bad actors embraced automation
 - * Create large amounts of malware
- * Good actors have not kept pace
 - * Still construct malware detection rules manually

The Solution:

Deep Machine Learning Applied to Cybersecurity





Gartner report: "Intelligent and Automated Security Controls Impact the Future of the Security Market", Oct 2015

Graphical Expression of Files



Step 1:

- Malware has one thing in common with all files: it is composed of code
- Software code is best expressed as a graph
- We characterize malware as a graph then feed it into our Deep Learning engine

Graphical Characterization of Malware



Big Data & Deep Learning Platform in the Cloud

Input

Output





Malware? What Family? Capabilities?

Graph-Based Malware Features

Cloud-Based Deep Learning Neural Network

Step 2:

• Our Deep Learning engine predicts malware with precision and real-time speed

Malware Prediction Using ML & Graphs

Unknown file

Predicted as malware



Neural network is trained to recognize malware

Machine Learning-Based Automated Malware Analysis

The Most Accurate and Fastest Platform



Accurately detects malware at 99.5%

Why Now?

- Deep Learning most accurate in AI industry
- HPC platforms readily available (e.g., AWS)
- Can provide comprehensive visibility

Deployed System



User Interface and Visual Analytics CISO / Security Leaders View

• Threat Landscape Specific to Your Enterprise



User Interface and Visual Analytics Analysts / Incident Responders View

Comprehensive Malware Analysis

	Cyber 20/20					
		Uploa	d Files Light		Update(s): 0	
≣ Q	Data Projector					
4 0	DATA			Show All Isolate 101 Data points	Clear selection	
	1 tensor found Financial Malw	1 tensor found Financial Malware 10K -		Smsagent Search .*	^{by} family 💌	
Þ	Label by family		•	shife hash 0dc0f5e12ee9f74ca682 neighbors ?	—— 100	
	Color by family		—	assembly.unctions 6484 distance COSINE assembly.blocks 22385 assembly.instructions 128947	EUCLIDIAN	
	Smsthief		960	assembly.calls 7245 Nearest points in the original assembly.macalls 29064 Smsthief	0.435	
	Smsagent		962	CSmsagent Injector assembly.jumps 11569 injector	0.512	
	Cutwail		962	Banker assembly.nmjumps 205 October Additioned Additional Addition	0.537	
	Shifu		958	Smsagent bytes.entropy 0.738897 Shifu	0.620	
	Banload		960	Shifu Ramnit Bouil a Charles Control of Cont	0.625	
	Banker		955	Andreineda Andreineda Andreineda Andreineda Andreineda Cutturail	0.640	
	 Andromeda 		964	Cutwain C	0.685	
	Zbot		974		0.608	
	Ramnit		966	Shifu Injector	0.734	
	Injector		971	Smsthief Banload	0.734	
	T-SNE	PCA	CUSTOM	Armit A	0.741	
				•Zbot •Zbot •Banker Banload Zbot	0.750	
				Andromeda	0.757	
	X	~		Banker «Ramnit einjector Banker Shifu	0.791	
	Component #1	- Com	oonent #2 📼	Andromeda Injector Zhot	0.800	
				Andromeda	0.802	
	Z Component #3	- V		Shifu	0.829	
	Total variance described: 18.1%.					

Class Projects

Machine Learning



Standardized Indicators of Compromise







Graphs

Visual Analytics

