Intelligence Augmentation: The Future of Cyber Threat Intelligence

Ed Cabrera, Chief Cybersecurity Officer
Trend Micro

- 30 years focused on security software
- Headquartered in Japan, Tokyo Exchange Nikkei Index (4704)
- Annual sales over $1B US
- Customers include 45 of top 50 global corporations
- 6000 employees in over 50 countries

500k commercial customers & 250M+ endpoints protected
AI / ML / DL

- **ARTIFICIAL INTELLIGENCE**: A technique which enables machines to mimic human behaviour.
- **MACHINE LEARNING**: Subset of AI technique which use statistical methods to enable machines to improve with experience.
- **DEEP LEARNING**: Subset of ML which make the computation of multi-layer neural network feasible.
## What is Machine Learning

<table>
<thead>
<tr>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans</td>
</tr>
<tr>
<td>Computers</td>
</tr>
</tbody>
</table>

Machine Learning is the art of getting a computer to identify similar objects.
What is Machine Learning
What is Machine Learning

Unsupervised Learning
- Dimensionality Reduction
- Clustering
- Targeted Marketing
- Recommender Systems
- Customer Segmentation

Supervised Learning
- Classification
- Regression
- Population Growth Prediction
- Advertising Popularity Prediction
- Weather Forecasting
- Market Forecasting
- Customer Retention
- Diagnostics

Reinforcement Learning
- Real-time decisions
- Game AI
- Skill Acquisition
- Robot Navigation
- Learning Tasks
Similar or Identical?
Similar or Identical?
Machine Learning in Cyber Security
ML Types

Fig. 1. Types of Learning [2] [3]
Data – The Foundation of Machine Learning

- Malware and benign samples
- Behavior logs
- Honey pot emails
- Hosted Email Service
- Web URL
- Web page contents
- Exploit kits
- Device info & identification
- CVE database
- NetFlow logs
- Process behavior logs

- Statistics of File, IP, Domain
- Histogram, Prevalence, Distribution, etc.

Correlated Data + In-depth Knowledge
Data – The Foundation of Machine Learning

- More data & More processing needed
- Less data & Less processing needed
ML Training Process
Cyber Security

Initial Access
- Exploits – Web / Docs
- Malicious URL
- Phishing URL
- Business Email Compromise

Execution
- Malware – EXE / Scripts
- Fileless
- Command & Control
- IOT security

Lateral Movement
- Compromised Accounts
- Insiders
SVM Exploits

- Model for each Exploit Kit
- Cover Neutrino, Rig, Sundown, Kaixin, Hunter, etc.
- Run in 100G throughput

SVM Document

- Generic model for all malicious Macro
- Cover social engineering attack, downloaders, etc.

Boosted Tree Business Email Compromise

- Model for email-writing style of each VIP
- Cover CEO Fraud, payment hijack, etc.
- Under POC

SVM Phishing

- Protect 394 brands
- Detect 150K phishing URL per day

Deep Learning Malicious URL

- Protect 6 billion URL queries per day
- Detect 2.8 million malicious URL per day
• Pre-Execution & Post-Execution
• Proactively detected 98.8% of the WannaCry samples

• Behavior anomaly
• Unnatural-spelling domains
• Early warning service

• Model for known C&C connections
• Use NetFlow
• Under POC

• Models for known IoT bots
• Use NetFlow
• Detect Mirai with only 0.0032% FP
• Both IT & OT
• HTTPS OK

• Focus on OT
• Baseline for each device type
• Use NetFlow
• HTTPS OK
Markov Chain to build transition matrix
Belief propagation to calculate score based on transition probability
Highlight high-severity events, compromised hosts, and accounts
Provide kill chain for investigation
Paradigm Shift - 2019

Easy for Experts

0% 90% 100%

Signature, Expert Rules, Etc...

Machine Learning

Hard for Experts

Easy for ML

0% 90% 100%

Machine Learning

Signature, Expert Rules, Etc...

Hard for ML

Proactive

Lightweight
The Next Step: ML in Cyber OSINT/SOCMINT
HOW SOCIAL MEDIA CAN BE USED TO GATHER ACTIONABLE THREAT INTELLIGENCE

• Social media platforms allow users and organizations to communicate and share information.
• For security professionals, it could be more than just a networking tool. It can also be an additional source of valuable information on topics from vulnerabilities, exploits, and malware to threat actors and anomalous cyber activities.
• In fact, 44% of surveyed organizations cited the importance of social media intelligence (SOCMINT) to their digital risk protection solutions.
Where social media analysis can help

• Social Media can be used to gather Actionable Threat intelligence (IoCs, early alerts of ongoing campaigns)
• Some malware families use Social Media Platforms as (intermediate) C2
• Social Media poses Enterprise Reputational Risks and can be monitored for that
• Scammers impersonate Corporations and Individuals in Scam Campaigns on Social Media
• #ANONYMOUS! Influence on availability of business processes
• Mal-Actors investigations
Social Media Data Processing Workflow

- Extract the Data Slices From raw data
- Import into text search platform
- Enrich

Massage
Visualize
Convert to graphs
Identify
And document Discoveries
Data Slicing... 😊

athena_search_twitter_md5.py
athena_search_twitter_name_regex.py
athena_search_twitter_query.py
athena_search_twitter_regex.py
athena_search_twitter_source.py
athena_search_twitter_user.py
athena_search_twitter_word.py
search2_twitter.sh
search_twitter.sh
search_twitter_f.sh
search_twitter_source.sh
search_utwitter.sh

arch_twitter.sh "苏尔的开" anubis_hunt
arch_twitter.sh @day cvemon
arch_twitter.sh CVE- cvemon
ing #bettingexpert #soccer #cash #bitcoin #western #unibet
arch_twitter.sh "fixed%matches" fixedmatches
arch_twitter.sh "bettingtips" fixedmatches
arch_twitter.sh "unibet" fixedmatches
arch_twitter.sh "bet365" fixedmatches
arch_twitter.sh exploit cvemon
arch2_twitter.sh /data/ftr-share/fy/tools/q2 scam22
arch2_twitter.sh /data/ftr-share/fy/tools/q5 scam22
Natural Language Analytics

The text needs to be pre-processed, we perform automated keyword extraction and remove stop words.
Examining context of social media content
Situational Awareness: who to follow
AESDDoS bot exploits CVE-2019-3396 flaw to hit Atlassian Confluence Server

AESDDoS exploits CVE-2019-3396 flaw to hit Atlassian Con.. A new variant of the AESDDoS bot is exploiting a recent vulnerability in the Atlassian collaborative software Confluence. securityaffairs.co
Vulnerability lifecycle through twitter
Visualizations can be done in ES
ML Research Continued
Network Threats Examined: Clustering Malicious Network Flows with Machine Learning

To discover and analyze different kinds of network anomalies, flow data needs to be looked at as they contain information useful for analyzing traffic composition of different applications and services in the network.

Machine learning is then applied to cluster malicious network flows. This will help analysts obtain insights that can show them relationships between different malware families, and how they differ from one another.
Using Machine Learning to Detect Malware Outbreaks With Limited Samples

In collaboration with Federation University Australia researchers, conducted a study titled “Generative Malware Outbreak Detection,” which showed the effectiveness of the latent representations obtained through adversarial autoencoder for such situations. This ML model for malware outbreak detection uses generative adversarial network (GAN) to obtain smooth approximated nearby distributions from a small number of OS X training samples.
Thank You!