

ML4SEC

Machine Learning for Cyber Security: 2020

Mr. Chamath Palihawadana: c.palihawadana@rgu.ac.uk

Dr M. Omar Al Kadri: o.alkadri@rgu.ac.uk

Dr Harsha Kalutarage: h.kalutarage@rgu.ac.uk

Overview

- Introduction
- Machine Learning for Cyber Security
- Challenges for Using ML in Cyber Security
- Malicious Use of Machine Learning
- Summary

Introduction

What is a cyber attack?

A malicious and deliberate attempt by an individual or organisation to breach the information system of another individual or organisation. Usually, the attacker seeks some type of benefit from disrupting the victim's network.

Why do they happen?

Who can be affected?

Think Individually

- What could happen if your information is compromised?
 - Email or data loss



Think Individually

- What could happen if your information is compromised?
 - Email or data loss
 - Much worse
 - Max out credit cards in your name, it will be near impossible to recover credit score, no mortgage, no leases, no loans.
 - Modify your information, end contracts and close bank accounts.
 - Commit crimes in your name online and offline.



Think Individually

- What could happen if your information is compromised?
 - Email or data loss
 - Much worse
 - Max out credit cards in your name, it will be near impossible to recover credit score, no mortgage, no leases, no loans.
 - Modify your information, end contracts and close bank accounts.
 - Commit crimes in your name online and offline.
 - Even issue death certificates in your name. In DefCon'15 , A session titled "I Will Kill You" from Australian security researcher Chris Rock, demonstrated how it could be done online.



Think Bigger

- On companies and organisational level, there is much to lose!
- Motives vary from self-gain to total destruction!



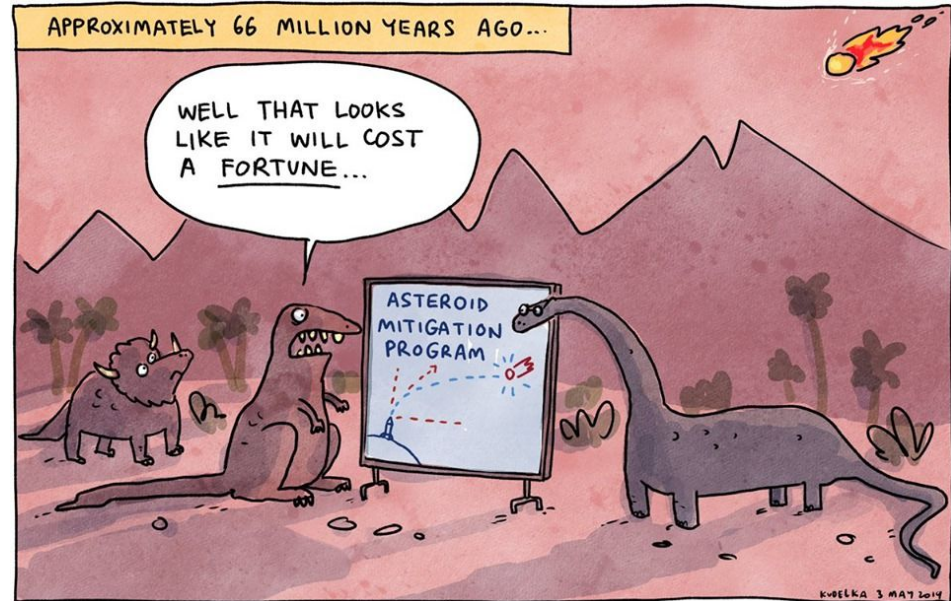
Think Bigger

- On companies and organisational level, there is much to lose
- Motives vary from self-gain to total destruction



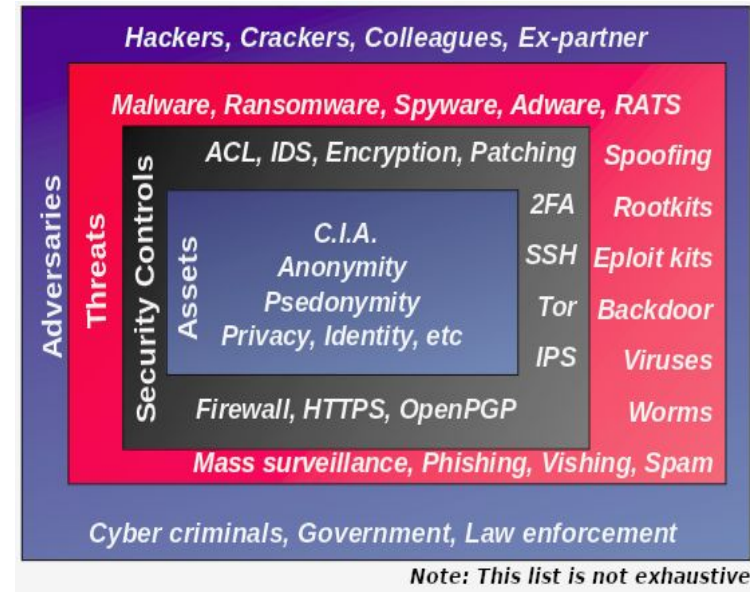
Think Bigger

- On companies and organisational level, there is much to lose
- Motives vary from self-gain to total destruction
- What about impact on countries?



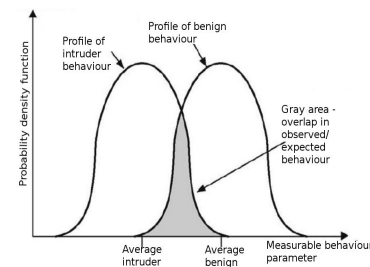
Cyber Security Stakeholders

- Exploit vulnerabilities
 - Weaknesses in HW/SW/Users
- Use various threats
 - Weakest link in security?
- Use intrusion detection



Limitation in Security Measures

- Signature/misuse based detection
 - Contain a database of recognised attacks
 - Activity is compared with signature database
 - Zero-day exploits go undetected
- Use tools borrowed from Machine Learning (ML)
 - Assumption → behaviour differ
 - Anomaly/behaviour based



Demo: Machine Learning for Cyber Security in Action (Spam Filter Example)

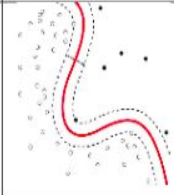
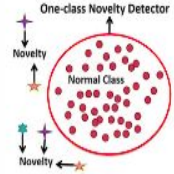
Machine Learning in Cyber Security

- Smartphone Security to Automotive Security
 - Automotive Security [1,2]
 - IoT Security [3]
 - Wireless Healthcare Network Security [4]
 - Software (Android) Security [5,6] <http://acidproject.org.uk>
 - Network Security [7,8]
 - Insider Threat Monitoring [9]
 - And more...

Challenges for using Machine Learning in Cyber Security

Anomaly Detection

- Need to have a perfect model of normality
 - Closed world assumption not hold in real life?
- Not certain all cases are covered
 - Normality changing over the time
- Malicious activities \leftrightarrow Anomalies
 - Do these assumptions hold?
- Resulting **high false alarm rates!**

	Classification	Anomaly detection
Training samples	Enough to distinguish two classes	Almost all from one class
		

Low Fault Tolerance Rates

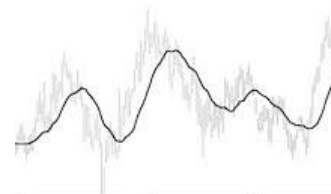
ML system	Cost of false negative	Cost of false positive
Product recommendation	Low: potential missed sales	Low: Continue shopping
Optical character recognition	Low: post processing fixes	Low: post processing fixes
Spam detection	Low: Spam lands in inbox	High: Missed important email
Intrusion detection	Extremely high: Critical damage to the business	High: Wasted analyst's time

- Cost of false alarms to the business is **extremely high**

Diversity of Cyber Data

- Even most basic characteristics (e.g. bandwidth, duration) has huge variability
- Difficult to find stable notion of “normality”
- One way to reduce the diversity
 - Reduce granularity (e.g. aggregation, moving average)
 - Pros: More stable series→easy to model
 - Con: Reducing the visibility
- Unpredictable over short time periods

OSI MODEL	TCP/IP MODEL
Application Layer	Application Layer
Presentation Layer	
Session Layer	
Transport Layer	Transport Layer
Network Layer	Internet Layer
Data Link Layer	Network Access Layer
Physical Layer	



Moving Average

Lack of Data Availability

- Goal of ML: Predict results based on incoming data
 - Three components: Training data, Features and Algorithms
 - If the data is crappy, even the best algorithm won't help
- Most cyber-security tasks are supervised learning tasks
 - Labelled data expensive to obtain
 - Legal, ethical and privacy issues
- Benchmark datasets - can they represent all possible situations?

Imbalanced Data

- Class imbalanced
 - Ratio between the majority class and the minority class are large
 - A ratio of 1:10 is considered imbalanced in ML community
- Malicious training examples are extremely rare in Cyber Security
 - Imbalance ratio of 1:10000 common in Cyber Security problems

Arms Race

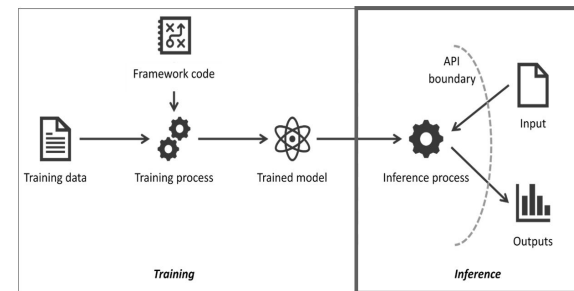
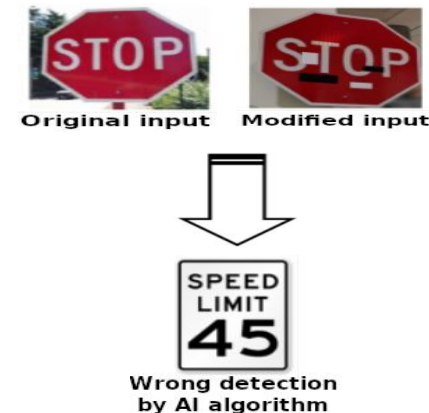
- Attacker-Defender arms race
 - Advancement of one side will not end the game
 - Will lead to a new round with different settings
- Model and insights becomes obsolete quickly
 - Volumes change, New protocols and domains appear
 - Malicious activities have trends, etc
- Concept drift
 - The model not change, but the world that changed

Semantic Gap

- Challenging to transfer results into actionable reports
- Interpret results from operator's point of view
 - What does it means?
- Unlikely to meet operational expectations
- Addressing the semantic gap
 - Incorporate local security policies

Adversarial Settings

- ML techniques originally designed for stationary environments
 - Stationary \rightarrow Prob. distribution does not change over the time
- Attacks against ML systems
 - Adversarial inputs - E.g. Attacks in spam filtering \rightarrow spam messages are obfuscated through misspelling of bad words or insertion of good words [10]
 - Data poisoning attacks
 - Model theft



ML Threat Matrix

Reconnaissance	Initial Access	Execution	Persistence	Model Evasion	Exfiltration	Impact	
Acquire OSINT information: (Sub Techniques) 1. Arxiv 2. Public blogs 3. Press Releases 4. Conference Proceedings 5. Github Repository 6. Tweets	Pre-trained ML model with backdoor	Execute unsafe ML models (Sub Techniques) 1. ML models from compromised sources 2. Pickle embedding	Execute unsafe ML models (Sub Techniques) 1. ML models from compromised sources 2. Pickle embedding	Evasion Attack (Sub Techniques) 1. Offline Evasion 2. Online Evasion	Exfiltrate Training Data (Sub Techniques) 1. Membership inference attack 2. Model inversion	Defacement	
ML Model Discovery (Sub Techniques) 1. Reveal ML model ontology – 2. Reveal ML model family –	Valid account	Execution via API	Account Manipulation		Model Stealing	Denial of Service	
Gathering datasets	Phishing	Traditional Software attacks	Implant Container Image	Model Poisoning	Insecure Storage 1. Model File 2. Training data	Stolen Intellectual Property	
Exploit physical environment	External remote services			Data Poisoning (Sub Techniques) 1. Tainting data from acquisition – Label corruption 2. Tainting data from open source supply chains 3. Tainting data from acquisition – Chaff data 4. Tainting data in training environment – Label corruption		Data Encrypted for Impact Defacement	
Model Replication (Sub Techniques) 1. Exploit API – Shadow Model 2. Alter publicly available, pre-trained weights	Exploit public facing application					Stop System Shutdown/Reboot	
Model Stealing	Trusted Relationship						

Demo: Malicious use of Machine Learning in Action (Captcha Example)

Summary

- What Are Cyber Attacks
- Potential Impact of The Attacks
- Machine Learning for Cyber Security
- Challenges for Using ML in Cyber Security
- Malicious Use of Machine Learning

Demonstration code: <https://github.com/RGU-Computing/ML4SEC-workshop>



ML4SEC

Thank You For Listening!
Questions?

Further References

- [1] Kalutarage, H. K., Al-Kadri, M. O., Cheah, M., & Madzudzo, G. (2019). Context-aware Anomaly Detector for Monitoring Cyber Attacks on Automotive CAN Bus. In ACM Computer Science in Cars Symposium (p.7). ACM.
- [2] Tomlinson, A., Bryans, J., Shaikh, S. A., & Kalutarage, H. K. (2018). Detection of Automotive CAN Cyber-Attacks by Identifying Packet Timing Anomalies in Time Windows. In 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W) (pp. 231-238). IEEE.
- [3] Zakariyya, I., Al-Kadri, M. O., & Kalutarage, H. (2020). Resource Efficient Boosting Method for IoT Security Monitoring. In IEEE Consumer Communications and Networking Conference (IEEE CCNC 2021)
- [4] Hajar, M. S., Al-Kadri, M. O., & Kalutarage, H. (2020). LTMS: A Lightweight Trust Management System for Wireless Medical Sensor Networks. In 2020 19th IEEE International Conference On Trust, Security And Privacy In Computing And Communications (IEEE TrustCom/IWCSS 2020)
- [5] Asavoe, I. M., Blasco, J., Chen, T. M., Kalutarage, H. K., Muttik, I., Nguyen, H. N., ... & Shaikh, S. A. (2016). Towards Automated Android App Collusion Detection. CEUR Workshop Proceedings, 1575, 29-37.
- [6] Kalutarage, H. K., Nguyen, H. N., & Shaikh, S. A. (2017). Towards a threat assessment framework for apps collusion. Telecommunication Systems, 66(3), 417-430.
- [7] Kalutarage, H. K., Shaikh, S. A., Wickramasinghe, I. P., Zhou, Q., & James, A. E. (2015). Detecting stealthy attacks: Efficient monitoring of suspicious activities on computer networks. Computers and Electrical Engineering, 47, 327-344.
- [8] Jia, G., Miller, P., Hong, X., Kalutarage, H., & Ban, T. (2019). Anomaly Detection in Network Traffic Using Dynamic Graph Mining with a Sparse Autoencoder. In 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 458-465). IEEE.
- [9] Palomares, I., Kalutarage, H., Huang, Y., Miller, P., McCausland, R., & McWilliams, G. (2017). A fuzzy multicriteria aggregation method for data analytics: Application to insider threat monitoring. In Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS), 2017 Joint 17th World Congress of International (pp. 1-6). IEEE.
- [10] Wickramasinghe, I., & Kalutarage, H. (2020). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Computing, 1-17