BigData Security Analytics

Giovanni Apruzzese
PhD Candidate in Information and Communication Technologies
University of Modena and Reggio Emilia
giovanni.apruzzese@unimore.it

Dartmouth
Security and Artificial
Intelligence Laboratory

Scuola di Ingegneria
Dipartimento di Ingegneria
"Enzo Ferrari"

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

Context: **CyberThreats current situation**

- Cyber threats are on the rise…

More than **4 billion** records compromised in 2016
→ a 566% increase from 2015

- …they become more advanced…

Examples of recent cyber attacks:
- BlackEnergy (2015)
- MEDJACK (2016)
- Archimedes (2017)
- *Wannacry (2017)*
- *Meltdown & Spectre (2018)*

- …and the penalties are steep…

3.6$ Million avg cost of a single (!) data breach

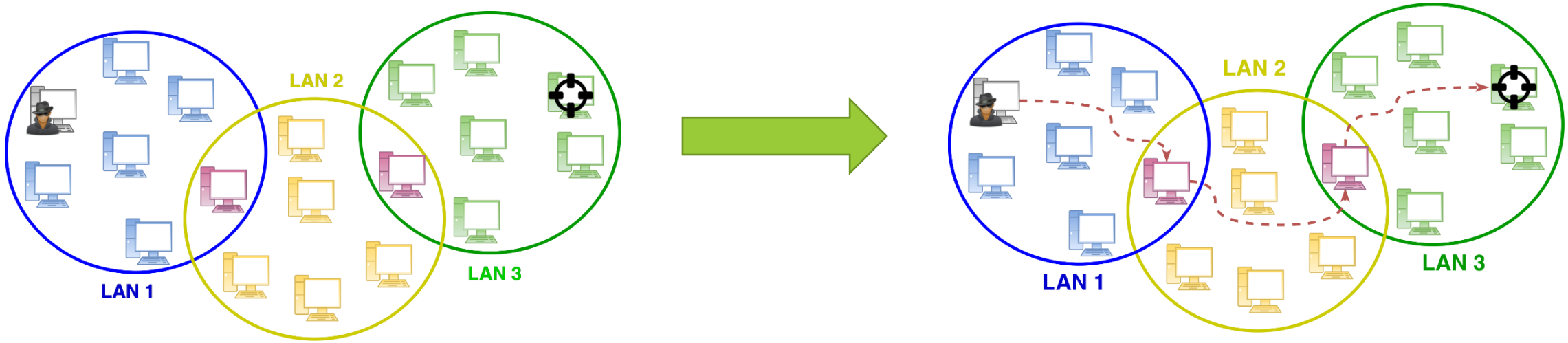
Solution: **(Big Data) Security Analytics**

Definition: process of using data collection, aggregation, and analysis tools for (enterprise) security monitoring and threat detection

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

Early Work: **Graph Analysis for *Lateral Movement* Detection through *Pivoting***

Attackers want to control hosts with **higher privileges** or **more valuable data.**



Pivoting: any action in which a *command propagation tunnel* is created among three or more hosts

Intuition: pivoting activities can be modelled through *Flow-sequences*

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

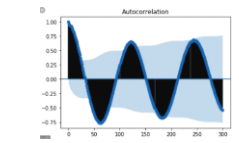
Shift to Machine Learning:
Time Series + Clustering to identify malicious external hosts

Novel malware variants are likely to evade detection...
...but some features of malware behavior persist and are shared even by novel variants

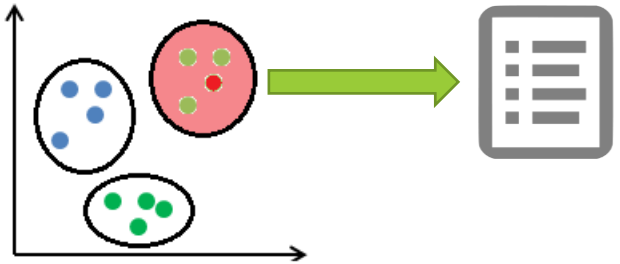
External hosts behaving similarly to a known malicious external host are likely to also be malicious

USE ONE TO FIND MANY:

- Generate temporal communication patterns
- Create clusters of similar communications
- Use NIDS alerts to find malicious external hosts
- Label as suspicious all clusters containing malicious external hosts
- Build *graylist* with external hosts belonging to suspicious clusters



Network communications



Giovanni Apruzzese
giovanni.apruzzese@unimore.it

Industrial Landscape of Machine Learning & CyberSecurity…

FORTINET

FortiGuard Artificial Intelligence (AI) Delivers Proactive Threat Detection at Machine Speed and Scale

JUNIPER NETWORKS

Machine learning moves to the front lines of defense against an expanding threat surface.

Machine Learning: New Frontiers in Advanced Threat Detection

Symantec

SOPHOS

Sophos Adds Advanced Machine Learning to Its Next-Generation Endpoint Protection Portfolio

MACHINE LEARNING HELPS US FIND NEW ATTACKS

F-Secure

KASPERSKY lab

Machine learning in Kaspersky Endpoint Security 10 for Windows

TREND MICRO

The truth is Trend Micro has been using machine learning since 2005.

CYBERARK

MACHINE LEARNING PREVENTS PRIVILEGE ATTACKS AT THE ENDPOINT

FireEye

McAfee is evolving its machine learning cybersecurity technology

McAfee

Rapid7 Attacker Behavior Analytics Brings Together Machine Learning and Human Security Expertise

RAPID7

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

…but all that shines is not gold

Main issues of ML for CyberSecurity:

Model training
- Where and how to find high quality and labeled training dataset?

Model deployment
- Is a pre-trained model applicable to my environment?

Model evaluation and selection
- How to compare different ML approaches?

Evolution over time (concept drift)
- How frequently should the model be re-trained?

Explainability
- Results are not explainable (yet)

False positives and false negatives
- 1% false positive rate in large organization = **thousands** of daily false alarms

Adversarial attacks
- More on this later…

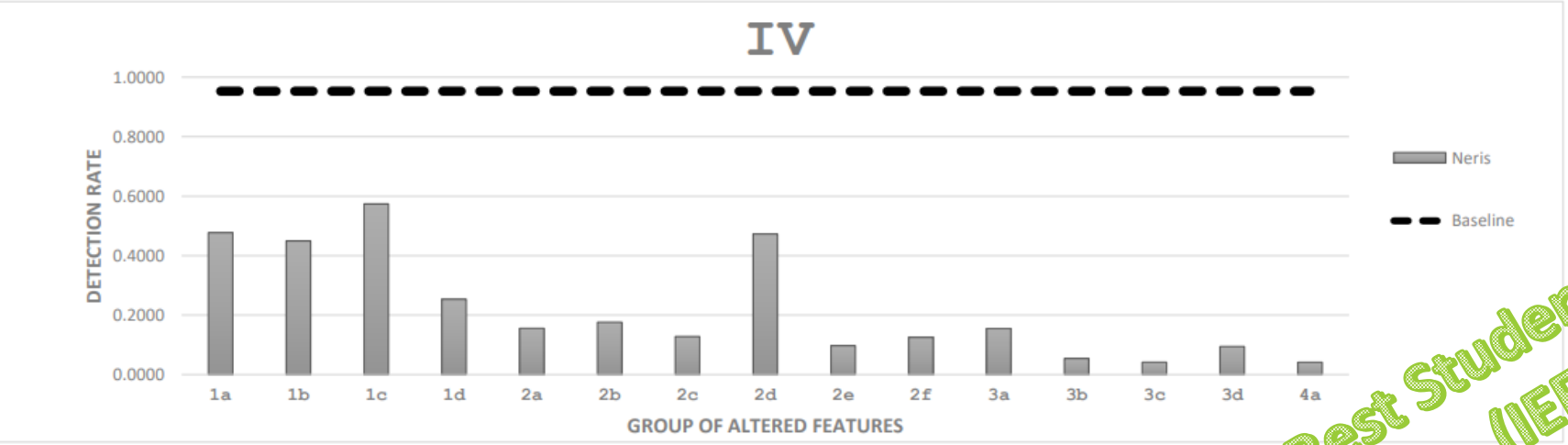
Giovanni Apruzzese
giovanni.apruzzese@unimore.it

Last Year: **Adversarial Attacks against Cyber Detectors**

We performed extensive experimental evaluations showing that the *detection rate* of a random forest-based botnet detector drops to values **lower than 5%** just by introducing small and targeted modifications to the network communications of the infected machines.

Experiments outline

1. Develop a botnet detector with good performance
2. Generate **realistic** adversarial samples
3. Evaluate the detector against the generated adversarial samples



Best Student Paper Award (IEEE NCA18)

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

Last Year: **Adversarial Attacks against Cyber Detectors**

Solutions?

- **Re-training with adversarial samples**

Requires the **availability** and **mainteance** of a **realistic** adversarial dataset

	Recall (before the attack)	Recall (after the attack)	Attack Severity
Random Forest	0.9684	0.3429	**0.6459**
Multi-layer Perceptron	0.9438	0.3012	**0.6809**
K-Nearest Neighbour	0.9375	0.3121	**0.6671**

- **Use feature-sets that cannot be modified by the attacker**

Decreases the performance of the detector against unmodified samples

	Precision		*Recall*		*F1-score*		*Accuracy*	
	Original	New	Original	New	Original	New	Original	New
Random Forest	0.9774	**0.8561**	0.9684	**0.8885**	0.9729	**0.8719**	0.9978	**0.9711**
Multi-layer Perceptron	0.9616	**0.7934**	0.9438	**0.7561**	0.9526	**0.7743**	0.9912	**0.9816**
K-Nearest Neighbour	0.9558	**0.8298**	0.9375	**0.8091**	0.9466	**0.8193**	0.9909	**0.9838**

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

References

- Giovanni Apruzzese, Fabio Pierazzi, Michele Colajanni, Mirco Marchetti, "**Detection and Threat Prioritization of Pivoting Attacks in Large Networks**", *IEEE Transactions in Emerging Topic in Computing (TETC)*
- Giovanni Apruzzese, Michele Colajanni, "**Evading Botnet Detectors based on Flows and Random Forest with Adversarial Samples**", *Proc. of the 17th IEEE International Symposium on Network Computing and Applications (IEEE NCA18)*, Cambridge, MA, USA, November 2018 **[BEST STUDENT PAPER AWARD]**
- Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Alessandro Guido, Mirco Marchetti, "**On the Effectiveness of Machine and Deep Learning for CyberSecurity**", *Proc. of the 10th NATO International Conference on Cyber Conflicts (CyCon 2018)*, Tallinn, Estonia, May 2018
- Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Mirco Marchetti, "**Addressing Adversarial Attacks against Security Systems based on Machine Learning**", *Proc. of the 11th NATO International Conference on Cyber Conflicts (CyCon 2019)*, Tallinn, Estonia, May 2019
- Giovanni Apruzzese, Mauro Andreolini, Michele Colajanni, Mirco Marchetti, "**Hardening Random Forest Cyber Detectors against Adversarial Attacks**", submitted to *IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI)*
- Giovanni Apruzzese, Mirco Marchetti, Michele Colajanni, Gabriele Gambigliani Zoccoli, Alessandro Guido, "**Identifying Malicious Hosts Involved in Periodic Communications**", *Proc. of the 16th IEEE International Symposium on Network Computing and Applications (IEEE NCA17)*, Cambridge, MA, USA, November 2017
- Fabio Pierazzi, Giovanni Apruzzese, Michele Colajanni, Alessandro Guido, Mirco Marchetti, "**Scalable Architecture for Online Prioritisation of Cyber Threats**", *Proc. of the 9th NATO International Conference on Cyber Conflicts (CyCon 2017)*, Tallinn, Estonia, May 2017
- Giovanni Apruzzese, Michele Colajanni, Mirco Marchetti, "**Evaluating the Effectiveness of Adversarial Attacks against Botnet Detectors**", *Proc. of the 19th IEEE International Symposium on Network Computing and Applications (IEEE NCA19)*, Cambridge, MA, USA, September 2019

Giovanni Apruzzese
giovanni.apruzzese@unimore.it

Why am I here?

Networking

Learning

Planning

Contact Information

Giovanni Apruzzese

PhD Candidate in Information and Communication Technologies

University of Modena and Reggio Emilia

Department of Engineering "Enzo Ferrari"

giovanni.apruzzese@unimore.it

https://weblab.ing.unimo.it/people/apruzzese