Politecnico di Torino

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

COMPUTER AND CONTROL ENGINEERING

PNRR/SERICS - Protect-IT – Distributed platform for cybersecurity data collection and analysis

Funded By	MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	MELLIA MARCO - marco.mellia@polito.it
Contact	GIORDANO DANILO - danilo.giordano@polito.it MELLIA MARCO - marco.mellia@polito.it
Context of the research activity	The project aims to respond to the deficiency of the current network security solutions, dramatically shown by the ever-increasing cyberattacks. The goal is to build an open and distributed platform leveraging AI-ML and policy-based enforcement solutions to identify new threats quickly, automatically, and reliably, and to deploy prompt and effective network security countermeasures. Progetto finanziato nell'ambito del PNRR M4C2, Investimento 1.3 - Avviso n. 341 del 15/03/2022 - PE0000014 Security and Rights in the CyberSpace (SERICS) - CUP E13C22001850001.
	The project consists of the design, implementation and testing of a flexible and distributed infrastructure to collect and process data related to the identification of novel cybersecurity threats. Thanks to technologies based on containers, darknets, and honeypots, the infrastructure will support the collection of data in a scalable manner. The usage of AI and ML solutions will allow one to extract actionable information to identify already known and novel attacks. To avoid data sharing, Federated Learning solutions will allow the training of a common model by leveraging the data each vantage point will collect seamlessly. The research will focus on extending machine learning solutions to automate and assist security analysts in the process (i) of identifying new attacks and (ii) setting up countermeasures. In the first phase, the candidate will study the state of art of platforms for cybersecurity data collection and machine learning, including the problem of data collection in distributed scenarios, and of federated learning solutions.

Objectives

In the second phase, the candidate will create and set up the distributed platform leveraging servers and virtual machines offered by partners of the PROTECT-IT/SERICS project. This will include the joint deployment of (i) data collected from honeypots and (ii) darknets.
Then he/she will investigate machine learning models that can identify new threats. For this, he/she will leverage mechanisms to create generic data representation using embeddings that will be then specialised to solve custom downstream tasks. Via federated learning, the candidate will investigate how to train a shared embedding model suitable to then solve specific tasks such as threat classification, anomaly detection, or identifying new attack patterns.

The project will involve a collaboration with partners in the Protect-IT project including the University of Brescia, Naples, and Milano among others.

-- Outline of the research work plan

1st year
- Study of the state-of-the-art security log analysis and state-of-the-art data collection platforms and machine learning models in ML.
- Data collection platform design with inclusion of security monitors such as honeypots, darknets, IDS, etc.

2nd year
- Deployment of the distributed platform for data collection and initial model training using federated learning.
- Propose and develop innovative solutions to the problems of cyber threats analysis with machine learning solutions.
- Propose multi-modal embeddings for network raw data and security logs.

3rd year
- Tune the developed techniques and highlight possible strategies to counteract the various threats.
- Application of the strategies to new data for validation and testing.

-- References:

- Gioacchini, Luca, Mellia, Marco, Vassio, Luca, Drago, Idilio, Milan, Giulia, Houidi, Zied Ben, Rossi, Dario (2023). Cross-network Embeddings Transfer for Traffic Analysis. IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, p. 1-13, ISSN: 1932-4537, doi: 10.1109/TNSM.2023.3329442
- Luca Gioacchini, Luca Vassio, Marco Mellia, Idilio Drago, Zied Ben Houidi, Dario Rossi (2023). i-DarkVec: Incremental Embeddings for Darknet Traffic Analysis. ACM TRANSACTIONS ON INTERNET TECHNOLOGY, vol. 23, p. 1-28, ISSN: 1533-5399, doi: 10.1145/3595378
- Huang, Kai, Gioacchini, Luca, Mellia, Marco, Vassio, Luca, Dynamic Cluster Analysis to Detect and Track Novelty in Network Telescopes, 9th International Workshop on Traffic Measurements for Cybersecurity (WTMC 2024), Vienna, Austria, July 2024
- Huang, Kai, Gioacchini, Luca, Mellia, Marco, Vassio, Luca, Incremental Federated Host Embeddings for Network Telescopes Traffic Analysis, IEEE International Workshop on Generative, Incremental, Adversarial, Explainable AI/ML in Distributed Computing Systems (AI-DCS), Jersey City, New Jersey, USA, July 2024

-- List of possible venues for publications

	Security venues: IEEE Symposium on Security and Privacy, IEEE Transactions on Information Forensics and Security, ACM Symposium on Computer and Communications Security (CCS), USENIX Security Symposium, IEEE Security & Privacy;
	AI venues: Neural Information Processing Systems (NeurIPS), International Conference on Learning Representations (ICLR), International Conference on Machine Learning (ICML), AAAI Conference on Artificial Intelligence, ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD);
	Computer networks venues: Distributed System Security Symposium (NDSS), Privacy Enhancing Technologies Symposium, The Web Conference (formerly International World Wide Web Conference WWW), ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT), USENIX Symposium on Networked Systems Design and Implementation (NSDI);

Skills and competencies for the development of the activity	- Good programming skills (such as Python, Torch, Spark) - Excellent Machine Learning knowledge - Knowledge Federated Learning and Machine Learning - Basics of Networking and security