Elaborati DAC

- **Elaborato 1: Explainable AI Dashboards for CyberSecurity in IoMT Scenarios**
- **Elaborato 2: Explainable AI Dashboards for CyberSecurity in SDN Scenarios**
- **Elaborato 3: Few-Shot Class Incremental Learning for 0-day attacks in IoMT Scenarios**
- **Elaborato 4: Few-Shot Class Incremental Learning for 0-day attacks in SDN Scenarios**
- **Elaborato 5: Federated Class Incremental Learning for 0-day attacks in IoMT Scenarios**
- **Elaborato 6: Federated Class Incremental Learning for 0-day attacks in SDN Scenarios**

—----------

Elaborato 1/2: Explainable AI Dashboards for CyberSecurity

Obiettivi

Progettare e sviluppare ***dashboard XAI*** per il supporto alla cybersecurity, in grado di spiegare il comportamento di modelli di Deep Learning (DL) nel rilevamento di minacce. L'elaborato mira a integrare tecniche di spiegabilità pre- e post-hoc, eventualmente con l'utilizzo di metriche per la valutazione delle spiegazioni. Inoltre, l'elaborato ha l'obiettivo di analizzare l'efficacia di tecniche XAI per la realizzazione di adversarial attacks.

Task / Domande di Ricerca

- Perchè un AI-based NIDS prende una particolare decisione osservando un flusso di traffico?
- Quali tecniche XAI risultano più efficaci nel rendere spiegabili i modelli di DL nel contesto della sicurezza informatica?
- Possono le spiegazioni pre-hoc (es. analisi di similarità tra biflussi di test e training set) fornire insight sul traffico prima della fase operazionale?
- Quali metriche possono essere utilizzate per valutare le spiegazioni fornite (es. fidelity, stabilità, coerenza)?
- Come posso sfruttare gli insight sul NIDS per realizzare adversarial attacks xai-based/informed?

Valutazione sperimentale

- Tecniche di XAI per NIDS Interpretability
- Metriche XAI: coerenza, stabilità, fedeltà, ecc.
- Integrazione con dashboard interattive per analisi visiva.
- Adversarial attacks sfruttando insight sul modello ottenuti con analisi XAI

Scenario

- **IoMT (Internet of Medical Things)**
  - Dataset:
    - https://www.unb.ca/cic/datasets/iomt-dataset-2024.html
- **Attacks in SDN environments:**
  - Dataset
    - https://ieeexplore.ieee.org/document/9187858

Metodi

- Similarity-based Analysis, Exploratory Data Analysis
- SHAP, Integrated Gradients, Counterfactuals
- Metodi di valutazione di diversi aspetti delle Explanations

Risorse

- Nascita, A., et al., 2021. XAI meets mobile traffic classification: Understanding and improving multimodal deep learning architectures. IEEE Transactions on Network and Service Management, 18(4), pp.4225-4246.
- Nascita, A., et al., 2024. A Survey on Explainable Artificial Intelligence for Internet Traffic Classification and Prediction, and Intrusion Detection. IEEE Communications Surveys & Tutorials
- …

Elaborato 3/4: Few-Shot Class Incremental Learning for 0-day attacks

Obiettivi

Applicare tecniche di Few-Shot Class Incremental Learning (FSCIL) per l'integrazione progressiva di nuove classi di attacchi nel NIDS, mantenendo le prestazioni sulle classi precedenti e apprendendo correttamente le nuove classi.

Task / Domande di Ricerca

- Riusciamo a distinguere i *pattern degli attacchi nuovi* rispetto a quelli già conosciuti?
- Come si comportano *diverse architetture di deep learning* nel contesto FSCIL applicato alla cybersecurity?
- Qual è l'impatto di utilizzare più *informazioni* sul traffico di rete (*single-modal vs multi-modal*) sulle prestazioni dell'apprendimento incrementale?
- Quali strategie di aggiornamento incrementale sono più adatte quando i nuovi attacchi sono rappresentati solo da pochi flussi di traffico (*few-shot*)?
- Come variano le prestazioni del modello incrementale al variare del numero di nuovi

attacchi aggiunti al modello (*diversi scenari di incremento*)?

Valutazione sperimentale

- Definizione di scenari pratici: numero di classi base, numero di task incrementali, quantità limitata di esempi per nuove classi.
- Caratterizzazione degli attacchi nuovi rispetto a quelli già conosciuti
- Analisi comparativa tra metodi FSCIL più recenti e baseline tradizionali.
- Confronto tra diverse architetture DL single modal.
- Confronto tra modelli single-modal e multi-modal per la classificazione incrementale.

Scenario

- **IoMT (Internet of Medical Things)**
  - Dataset:
    - https://www.unb.ca/cic/datasets/iomt-dataset-2024.html
- **Attacks in SDN environments:**
  - Dataset
    - https://ieeexplore.ieee.org/document/9187858

Metodi

- Approcci Few-Shot e Class Incremental Learning
- Architetture single-modal (CNN, RNN)
- Architetture multi-modali (features+payload, ecc.)

Risorse

- Masana, M., et al., 2022. Class-incremental learning: survey and performance evaluation on image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5), pp.5513-5533.
- Cerasuolo, F., et al., 2025. Attack-adaptive network intrusion detection systems for IoT networks through class incremental learning. Computer Networks, 263, p.111228.
- Aceto, G., et al., 2019. MIMETIC: Mobile encrypted traffic classification using multimodal deep learning. Computer networks, 165, p.106944.

Elaborato 5/6: Federated Class Incremental Learning for 0-day attacks

Obiettivi

Sviluppare soluzioni basate su Federated Class Incremental Learning (FedCIL) per la classificazione di attacchi in scenari distribuiti, con attenzione alla privacy dei dati, alla scalabilità e alla capacità del sistema di apprendere nuove classi senza dimenticare le precedenti.

Task / Domande di Ricerca

- Come combinare apprendimento federato e class incremental learning per l'aggiornamento continuo dei modelli in ambienti distribuiti?

- Qual è l'effetto della *non-IIDness* (distribuzione non indipendente e non identica) dei dati nei nodi sul processo di apprendimento incrementale?
- Come progettare una strategia di aggiornamento dei modelli che bilanci l'apprendimento locale con la generalizzazione globale?
- È possibile mitigare il *catastrophic forgetting* in contesti federati, in cui i dati delle classi precedenti non sono più disponibili in forma centralizzata?
- Quali tecniche di *mitigation del forgetting* si adattano meglio al contesto federato?
- Come garantire la privacy e la sicurezza dei dati pur consentendo l'aggiornamento del modello con nuove classi?

Valutazione sperimentale

- Simulazione di scenari federati realistici con più nodi (client) e distribuzioni di dati eterogenee.
- Analisi dell'impatto del numero di classi per task e della quantità di dati disponibili per ciascun nodo.
- Confronto tra strategie FedAvg, FedProx, e altre varianti recenti.
- Studio delle prestazioni in ambienti con attacchi mirati (adversarial client, data poisoning).

Scenario

- **IoMT (Internet of Medical Things)**
  - Dataset:
    - https://www.unb.ca/cic/datasets/iomt-dataset-2024.html
- **Attacks in SDN environments:**
  - Dataset
    - https://ieeexplore.ieee.org/document/9187858

Metodi

- Algoritmi di Federated Learning: FedAvg, FedProx, FedDyn, ecc.
- Approcci Class Incremental Learning: rehearsal, regularization, prototype-based
- Strategie ibride Fed-CIL
- Architetture deep learning per traffico di rete (single e multi-modal)

Risorse

- Agrawal, S., et al., "Federated learning for intrusion detection system: Concepts, challenges and future directions.", In Computer Communications, 195, pp.346-361, 2021
- Jin, D., et al., "Federated incremental learning based evolvable intrusion detection system for zero-day attacks.", IEEE Network, 37(1), pp.125-132, 2023
- Nair, A.K., et al., "A robust analysis of adversarial attacks on federated learning environments." In Computer Standards & Interfaces, 86, p.103723, 2023