ITEA 4

Project Results

ENTA

Privacy-preserving visibility into encrypted traffic

To address new challenges in cybersecurity and network operations, the ITEA project ENTA (Encrypted Network Traffic Analysis) developed a platform and artificial intelligence (AI) models to provide visibility into encrypted traffic without decryption, thereby preserving user privacy.

Over 90% of internet traffic is currently encrypted, offering greater privacy and security. However, as traditional techniques for traffic analysis and deep packet inspection (DPI) now fail to provide sufficient visibility, issues are arising in cyber-threat detection, network operations and law enforcement. Security operations centres, for instance, may be unable to detect malware, data exfiltration or rogue Internet of Things (IoT) devices, while law enforcement agencies may be hindered in forensics and data analysis and deprived of ongoing criminal transactions. From a business perspective, IT departments might be unable to enforce policy or quality of service. This generates the need to regain visibility while preserving privacy.

ENTA's answer lies in exploiting AI-based techniques that analyse the temporal and spatial properties of encrypted network traffic. To enable this, an encrypted network traffic analysis platform was developed using open-source components. This facilitates the creation and lifecycle management of solutions based on machine learning (ML) and deep learning (DL), through which users can develop their own models. Both ML and DL-based models were used in order to address solutions while improving accuracy under various conditions. Two use-cases – among many other possible cybersecurity-related use-cases – have demonstrated this approach: (1) the classification of encrypted traffic into applications and categories, and (2) the discovery of IoT devices and detection of rogue IoT devices. In addition, ENTA created 18 labelled datasets for model training and validation of the test results.

Technology applied

Developed around the open-source Kubeflow platform, ENTA's platform supports feature extraction, training/test dataset creation, and model training and evaluation. It is also scalable, extendable, and deployable either locally or in the cloud. The platform allows users to focus on developing solutions instead of the software infrastructure. Its support for ML/DL models has been enabled via the creation of the datasets with network packet capture, each addressing different elements of the use-cases such as IoT devices behaving normally versus when under attack. Three datasets have so far been made available at IEEE DataPort (with more to follow), providing third-party researchers with an open source of labelled data to train their models.

In the first use-case, six scenarios were explored for ML/DL-based network traffic classification according to traffic type, application name or activity detection. These AI approaches examine the statistical and/or sequential features of traffic without needing to inspect the payload, allowing them to recognise the difference between types of streams (like video, audio and chat) and the applications themselves (such as Netflix or WhatsApp). A mechanism was also developed for real-time classification using ML on a high-speed packet processing infrastructure. In the second use-case, five scenarios were developed for the detection and identification of (rogue) IoT devices by collecting and analysing traffic at the gateway. This resulted in ML models to classify both consumer and industrial IoT devices and to detect their operational or security status, including types of cyber-attack.

Making the difference

A major shortcoming in academic AI research is that model accuracies of 70-90% are usually achieved using the same datasets for both training and testing, which is how the academic baseline has been set and demonstrated. ENTA was thus motivated to enable a platform to facilitate testing with cross-datasets and to support the validation of results with different AI algorithms.

By making its datasets open source, the project addresses a current lack of public data and thereby lowers the barrier for AI research due to the laborious and partially manual nature of dataset creation. The first three datasets have been accessed over 2,200 times by researchers within a few months of their release. Meanwhile, six of ENTA's own models have achieved accuracies above 90% (including over 95% for IoT versus non-IoT classification), providing a highly promising foundation for further research while demonstrating the technology's potential as a DPI replacement/ complement.

Alongside the project's open-source element, commercialisation is ongoing. Solana and MTP are productising the traffic classification solution and the IoT solution respectively, each positioned as a service that can be adapted to user needs. Solana is already engaged in serious discussions with various possible customers, including large and small businesses and the Canadian government. The company is also exploring participation in a follow-up RD&I project to take the results from the prototype to a larger number of researchers and developers. Across the consortium, such activities have been supplemented by presentations at three tradeshows and the publication of 15 conference papers.

Traffic encryption in all forms of communication is only set to increase over time due to sophisticated cyber-threats. The need for traffic visibility will always be present for network and security monitoring to ensure IT networks operate robustly and resiliently. Obtaining visibility into traffic is also a must for law enforcement. As a result, ENTA expects the ML/DL-based approach to form an integral part of traffic inspection or classification solutions within the next five years. In turn, this will open up a variety of potential users beyond the original project scope, such as router/switch vendors that need to use traffic types and applications to support class of service, firewall vendors that need to differentiate between application types in the next generation of firewalls, or military operators that need to be aware of applications to achieve accurate network situational awareness. ENTA thus offers a springboard for far-reaching impacts on research, business and society as a whole.

Major project outcomes

Dissemination

- Three publicly available datasets from IEEE DataPort (https://ieee-dataport.org/datasets)
- One publicly available IoT dataset at https://zenodo.org/records/14802737
- Fifteen published papers in international conferences
- Showcased ENTA technology in three tradeshows during 2024

Exploitation (so far)

- New prodcuts:
  - Platform for development of AI-based network traffic analytics
  - Encrypted network traffic classification
  - IoT device discovery and threat detection
- New services:
  - IoT device datasets
  - Encrypted traffic classification datasets

Spin offs

- The outcome of encrypted network traffic classification has been productised and branded as TrafficWiz (https://trafficwiz.com/)

ITEA is the Eureka RD&I Cluster on software innovation, enabling a large international community of large industry, SMEs, start-ups, academia and customer organisations, to collaborate in funded projects that turn innovative ideas into new businesses, jobs, economic growth and benefits for society. ITEA is part of the Eureka Clusters Programme (ECP).

https://itea4.org

ENTA

20209

Partners

Austria
- Beia

Canada
- Dalhousie University
- Solana Networks

Spain
- Metodos y Tecnologia

Switzerland
- Ruag

Türkiye
- Karel Electronics

United Kingdom
- Centre for Factories of the Future

Project start
October 2021

Project end
March 2025

Project leader
Biswajit Nandy, Solana Networks

Project email
bnandy@solananetworks.com

Project website
https://project-enta.com/
https://itea4.org/project/enta.html

ITEA 4

eureka