OSINT-based Data-driven Cybersecurity Discovey

Fernando Alves, Pedro M. Ferreira, Alysson Bessani

LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

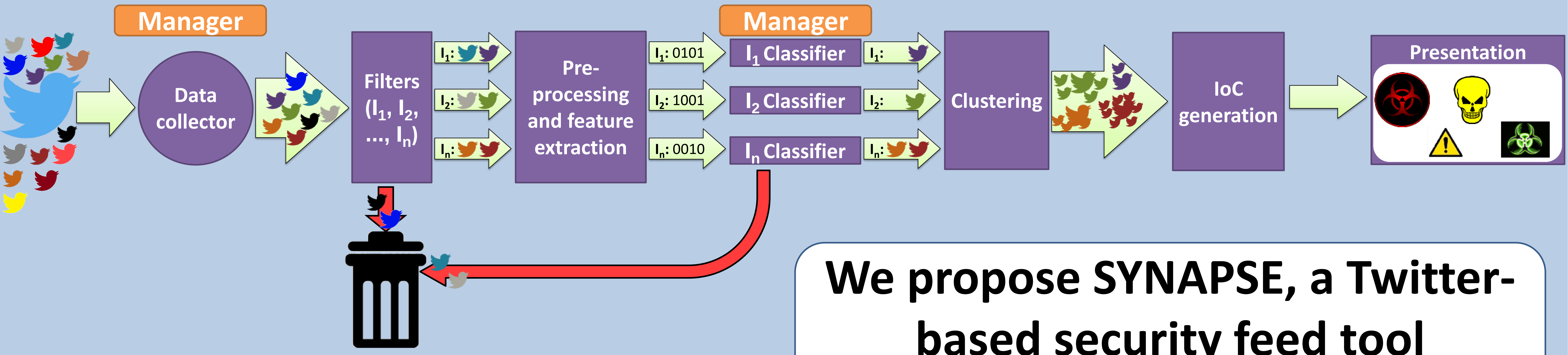
Security systems require timely updates and management

Cyberattacks are growing concern for all kinds of business, and even primary industry sectors. More that ever, cybersecurity **threat awareness** is of utmost importance due to the latest increase in vulnerability disclosure (from ~5k/year to 14.7k in 2017).

Companies can obtain news of the current threat landscape through paid curated feeds, or through harvesting **Open Source Intelligence (OSINT)**. The latter, although harder to employ, has taken both enterprises and the research community's attention.

A wealth of OSINT is published daily to the Internet

... and yet, companies do not take advantage from it. Security feeds, hackers, security analysts, researchers,, users, and others post their experiences and findings. Scraping such data can provide valuable and much needed security **awareness**. Twitter is a natural data aggregator, since it provides visibility to any kind of content.



Current security systems do not take advantage of OSINT feeds

The state-of-the-art in threat intelligence tools focuses on collecting from multiple OSINT sources. The research community's proposals do not consider a production environment.

There is a clear opportunity to research and develop a tool with advanced processing capabilities to deliver quality OSINT to threat intelligence systems.

We propose SYNAPSE, a Twitter-based security feed tool

SYNAPSE is designed to feed security systems with **Indicators of Compromise (IoC)** obtained from Twitter. **SYNAPSE**'s main features are:

- A supervised **machine learning** classifier to identify security-wise tweets
- An **innovative clustering** methodology to collate tweets referring the same threat
- A **tweet-to-IoC pipeline**, including automatic tagging, link exploration, and IoC extraction
- **Self-management** on the used data sources and classifier models

This work was supported by EC through funding of DiSIEM project, ref. G.A. nº 700692, and by FCT through funding of LASIGE Research Unit, ref. UID/CEC/00408/2013