Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(35^{th}cycle)

Strengthening Privacy and Cybersecurity through Anonimization and Big Data

By

Thomas Favale

******

Supervisor(s):
Prof. Marco Mellia, Supervisor
Prof. Danilo Giordano, Co-Supervisor

Doctoral Examination Committee:
Prof. Chadi Barakat, Referee, Inria, Université Côte d'Azur
Prof. Jose Luis Garcia Dorado, Referee, Universidad Autónoma de Madrid
Prof. Giancarlo Ruffo, Università di Torino
Prof. Guido Marchetto, Politecnico di Torino
Prof. Robert Renè Maria Birke, Università di Torino

Politecnico di Torino
December 2022

Abstract

The problem of Cybersecurity, together with Privacy are becoming increasingly pervasive and tangled especially in recent years with the relentless expansion of the Internet: it is becoming more and more important in our lives and it is the foundation for every business. In this scenario, malicious entities develop very rapidly and create new and increasingly sophisticated threats. Since their success could cause various catastrophes for each of us, it is important to be able to monitor the network and analyze the traffic captured through passive sensors such as Darknets, or active systems, as Honeypots. In the first case we are able to observe unwanted traffic (often referred to as Internet Background Radiation), generally produced by heavy-hitter sources. Moreover, coupling this tool together with active Honeypots, helps to further enrich the visibility on malicious events. Honeypots, on the other hand, are able to reply to unsollicited requests, providing a broader knowledge on the threat scenario, by engaging the potential attacker.

However, this study scenario clashes with the need to guarantee user privacy: capturing traffic anywhere on the network can involve packets generated by not malicious users, therefore identity disclosure could be a serious problem. For all these reasons, in this thesis I demonstrate how it is possible to perform network monitoring safely. As a first step, I propose an anonymization system called α-MON which is able to capture network packets at high speeds (multiple Gb/s) and apply a desired level of obfuscation for potentially sensitive information. Here I refer not only to the network addresses in the headers, but also to the payload: many protocols are still completely in clear (e.g. DNS), or, if encrypted, they expose the name of the reference service (TLS). It is important to identify them and consequently act to hide the aforementioned names contained in the payloads. The z-anonymity algorithm, on which α-MON is based, is also innovative and has its foundations in the k-anonymity, but

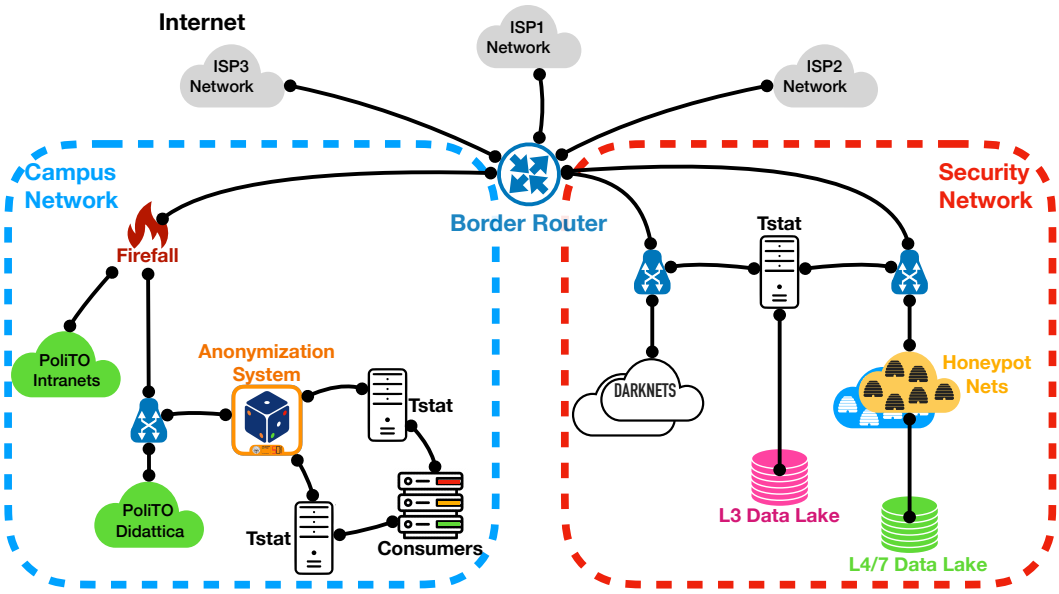

Fig. 1 Topology of the data collection system and cybersecurity environment.

working on a streaming fashion: the goal is still to hide uncommon identifiers that can easily be traced back to specific users, leaving the most common unchanged. The development and commissioning of α-MON made it possible to safely analyze the traffic within the Politecnico di Torino backbone during the Covid-19 pandemic, presenting a trend of use of the real network that has never been seen before. With this use case, I can demonstrate how this approach is promising and that the use of such tools does not degrade the quality of the data captured.

In this thesis, in addition to the privacy issue, I provide my contribution for the analysis of malicious traffic. In particular, I conducted studies on the performance of Deep Packet Inspection (DPI) systems with the aim of identifying the most suitable to operate in a streaming fashion. This is a fundamental point for creating a smart honeypot capable of understanding the protocol autonomously and actively responding consistently to attackers who do not necessarily use standard ports (for protocols) to perform their activities. I called this tool DPIPot. It is able to reveal attack patterns that would otherwise remain hidden because the most widespread and documented Honeypots tend to look only for the standard ports of the protocol that they are able to handle. Subsequently, I deployed DPIPot and a series of further responders within an address space dedicated to Darknets, demonstrating an increase in attacking

sources. This is interesting, since the infrastructure has allowed me to analyze all incoming traffic horizontally: my goal is to demonstrate that attackers are able not only to hit single services with considerable insistence, but also to interact with different protocols by cycling through all those left open at their disposal. I show that the purposes can be extremely varied, starting from Crawlers for the discovery of services to brute force attackers who try to login with usernames and passwords obtained from published leaks.

In conclusion, the objective of this thesis is to demonstrate that despite the recent enhancements regarding privacy, network monitoring is still possible without loss of data quality. In addition, I demonstrate how the use of mixed passive and active probe infrastructures allows the discovery of new and more visible attack patterns.