ICT15-129 - Big-DAMA: Big Data Analytics for network traffic Monitoring and Analysis

Abstract

The vast amount of big data in communication networks has opened a new era of data-driven solutions which will shape the future of the Internet. However, the complexity of the Internet itself and our dependence on it as a society have so drastically increased in the last few years, that more than ever, we need better and more scalable network measurement and analysis techniques and tools to understand it and manage it.

Critical applications such as network security, anomaly detection or dynamic network management require fast mechanisms for online analysis of thousands of events per second, as well as efficient techniques for offline analysis of massive historical data. Besides characterization, making operational sense out of the ever-growing amount of network measurements is becoming a major challenge.

Both networking for big data management and big data analytics in networking applications pose difficult challenges for industry and academic researchers. In a nutshell, handling large amounts of heterogeneous sources of data, and analyzing them in (near) real-time with big data analytics and machine learning approaches requires specialized hardware and software platforms to make it happen. The problem is, that such a specialized and tailored hardware and software require strong engineering to actually perform as needed, not only in terms of analysis speed, but also in terms of complexity of the analysis tasks which can be addressed.

The Big-DAMA project has conceived novel scalable techniques and big-data analytics frameworks (BDAFs) capable to analyze both online network traffic data streams and offline massive traffic datasets. The team has developed scalable online and offline machine learning-based techniques to monitor and characterize extremely large network traffic datasets. Using off-the-shelf machine learning libraries and parallelization techniques, we have conceived multiple data analytics algorithms for anomaly detection and cyber security, using supervised and unsupervised machine learning models.

Using the Big-DAMA BDAF, one can store and analyze big amounts of both structured and unstructured heterogeneous data sources, in either (near) real-time or in an off-line manner, using advanced big data analytics and machine learning models.

Examples of use cases we have tackled within the project include (i) network security, where we have implemented 0-day attacks-detection based on unsupervised analysis techniques; (ii) anomaly detection, where we have conceived approaches for online detection and classification of network and service anomalies; and (iii) Quality of Experience, where we have integrated machine learning techniques to monitor networks from a user-centric perspective.

While the outcomes of the Big-DAMA project have direct impact and application in the network monitoring and analysis domain, the techniques and technology developed within the project are highly applicable to address other data-driven domains of high relevance to society, where similar data analysis problems and requirements arise. Just to name a few of them: analysis of big data associated to smart cities, including intelligent transportation systems and smart mobility, smart energy management, environmental sensing analysis, and more; financial technology analysis, including fraud

detection, blockchain monitoring, price forecasting and analysis, consumer protection, etc.; and IoT data analytics, where trillions of devices will be connected to the Internet for an unprecedented level of sensing.

Being Artificial Intelligence and Big Data analytics a fast-growing worldwide market, the development of analysis techniques, technologies, as well as strong know-how in the area directly benefits the local Vienna market, by taking a leading position in such a timely and highly relevant topic, especially in the years to come.

Scientific disciplines:
Telecommunications (50%) | Database systems (20%) | Machine learning (30%)

Keywords:
Big Data, Analytics, Data Stream Processing, Data Stream Warehousing,Machine Learning, Data Mining, Network Traffic Monitoring and Analysis,Anomaly Detection, Network Security

---

Principal Investigator:	Pedro Casas
Institution:	AIT Austrian Institute of Technology GmbH
Co-Principal Investigator(s):	Tanja Zseby (TU Wien)

Status: Completed (01.03.2016 - 28.02.2019)
GrantID: 10.47379/ICT15129

Further links to the persons involved and to the project can be found under
https://www.gmbh.wwtf.at/funding/programmes/ict/ICT15-129/