# Interactive Data Mining for Root Cause Analysis of Performance Issues in Networks

Luis Galárraga, Elisa Fromont, Alexandre Termier

April 26, 2018

## 1 Context

AdvisorSLA is a French company headquartered in Cesson-Sévigné, a city located in the outskirts of Rennes in Brittany. The company is specialized in software solutions for network monitoring. For this purpose, the company relies on techniques of *network metrology*. AdvisorSLA's customers are carriers and telecommunications/data service providers that require to monitor the performance of their communication infrastructure as well as their QoE (quality of service). Network monitoring is of tremendous value for service providers because it is their primary tool for proper network maintenance. By continously measuring the state of the network, monitoring solutions detect events (e.g., overloaded router) that may degrade the network's operation and the quality of the services running on top of it (e.g., video transmission could become choppy).

## 2 Problem

When a monitoring solution detects a potentially problematic sequence of events, it triggers an alarm so that the network manager can take actions. Those actions can be preventive or corrective. Some statistics show that only 40% of the triggered alarms are *conclusive*, that is, they manage to signal a well-understood problem that requires an action from the network manager. This means that the remaining 60% are *presumably* false alarms. While false alarms do not hinder network operation, they do incur an important cost in terms of human resources. Thus, in this thesis we propose to *characterize* conclusive and false alarms. This will be achieved by designing automatic methods to "learn" the conditions that most likely precede the fire of each type of alarm, and therefore predict whether the alarm will be conclusive or not. This can help adjust existing monitoring solutions in order to improve their accuracy. Besides, it can help network managers automatically trace the causes of a problem in the network.

## 3 Methodology

The aforementioned problem has an inherent temporal nature: we need to learn which events occur *before* an alarm and *in which order*. Moreover, metrology models take into account the measurements of different components and variables of the network such as latency and packet loss. For these two reasons, we

resort to the field of multivariated time sequences and time series. The fact that we know the "symptoms" of an alarm and whether it is conclusive or not, allows for the application of supervised machine learning and pattern mining methods. In the realm of machine learning, detecting the *class* of an alarm is a classification problem. Since machine learning methods have traditionally not been concerned with the interpretability of their verdicts, we envision to enhance our methods with discriminative pattern mining techniques. Such techniques, for example, can find comprehensible sequences of events that occur more frequently before false alarms, e.g., data transmission between two network components A and B. In a classical setting, discriminative pattern mining approaches deal with static data. Thus, they have no ways –other than statistical– to evaluate the real relevance of the discovered patterns. In our scenario, however, we can establish a feedback loop for our pattern mining algorithm: The events represented by the patterns can be reproduced in the network in order to either verify/reject the pattern's validity or refine it with additional context information, e.g., figuring out that the faulty transmission between components A and B occurs for video packets. This will be the first effort to integrate a pattern discovery algorithm inside a feedback loop, and study the actual relevance of the extracted patterns. The scientific challenge lies on the design of such feedback loop. There will be many patterns to test and each test will incur some cost in terms of time and network bandwith. Hence, the core problem is how to identify the real cause of an issue with a limited test budget, or in other words, how to prioritize the patterns for testing.

Such feedback scheme can constitute the base for the development of online methods for alarm classification and root cause analysis of faults. In this spirit the monitoring system can automatically learn the relevant set of patterns that characterize network faults and adjust the behavior of the alarms as the network operation evolves. Finally, in a different line of thought, the system could also immerse the user in the process by providing detailed information about inconclusive patterns and asking for user's feedback.

# 4 Working Environment

The PhD candidate will work part time at the IRISA research center in Rennes and part time at AdvisorSLA in Cesson-Sévigné. The exact distribution of time spent in each of the facilities will be discussed during the interview process. At IRISA, the candidate will be hosted by the LACODAM[1] (Large Scale Collaborative Data Mining) team and will be supervised by:

**Luis Galárraga.** Luis is a full-time INRIA researcher working in the LACODAM team since October 2017. His research interests can be described in terms of two axes: pattern mining and knowledge representation. His current and past work experience includes pattern-based regression algorithms, data mining in RDF knowledge bases, efficient RDF query processing, and reasoning with rules and metadata. More details about his research work are available at `http://luisgalarraga.de`.

[1] `http://team.inria.fr/lacodam`

**Elisa Fromont.** Elisa is a professor in computer science at the University of Rennes 1, and member of the LACODAM team since September 2017. Before, she was an associate professor at the University of Saint-Etienne. Her research interests comprise the development of machine learning algorithms for specific applications. In particular, she works on new deep learning methods for semantic segmentation and anomaly detection. More details are avaialble at`http://people.irisa.fr/Elisa.Fromont/`.

**Alexandre Termier.** Alexandre is a professor in Computer Science at the University of Rennes 1, and the head of the LACODAM team. He is also an associate member of the Grenoble Informatics Laboratory (LIG), where he worked as an associate professor and researcher at the SLIDE [2] group. His research speciality is pattern mining, with focus on condensed representations and parallelism. More details are avaialble at `http://people.irisa.fr/Alexandre.Termier/`.

# 5 Required skills

We look for a very motivated candidate who meets the following requirements.

1. A master's degree in Computer Science
2. Proficiency in spoken and written English
3. The ability to work autonomously

# 6 Application instructions

Send your application to ALL the following email addresses: luis.galarraga@inria.fr, elisa.fromont@irisa.fr, alexandre.termier@irisa.fr

Your application must contain:

1. A CV
2. Your last grade certificate (if you are currently finishing your Master's degree, we need an official list of the grades you obtained so far in this degree with your rank among your peers)
3. At least two recommendation letters
4. A motivation letter

[2] Scalable Information Discovery and Exploitation