Systems@ETH Zürich

Real-time Policy Monitoring for Large-scale Distributed Systems

Master Thesis Proposal

Introduction

Enterprises store and process vast amounts of data often containing confidential information about real persons and organizations. As a result, access to these data must be granted only to authorized system users. Such security requirements are prevalent in today's data infrastructures -- the most recent one is EU's GDPR -- and emerge also in medical and financial institutions, which often have to comply with acts like HIPAA and SOX.

Enforcing complex security policies has received considerable attention over the past years [1], leading to the development of several *access control* mechanisms that are now considered an essential feature of data-centric systems. Access control is typically *static* and determines the behavior of the system at a specific point in time (the current state), preventing unauthorized user actions. A major drawback of these mechanisms is that they cannot capture and verify *dynamic* dependencies between different states of the system *across time*, therefore, they cannot identify unexpected system behavior indicated by such causal dependencies [2]. This task is achieved via *policy monitoring* and is the focus of the thesis.

In practice, policy monitoring requires instrumented systems that log and periodically emit events related to user actions or component interactions, which are then processed by the monitor in a streaming fashion to identify event patterns violating the policies. Besides the security use case, this task has been proved extremely useful also in debugging, identifying hidden interactions between software modules, and, in general, understanding the behavior and performance of large-scale distributed systems [3].

Current Status

Current approaches for policy monitoring suffer from at least one of the following limitations. Firstly, they support policy languages that are not expressive enough to capture complex temporal dependencies. Secondly, they rely on pattern matching algorithms that do not scale with large volumes of data. Last but not least, they are based on batch processing systems that cannot meet the low-latency and high-throughput requirements of real-world applications.

To tackle the aforementioned problems we have recently built a novel monitor for online policy verification. The monitor supports an expressive language called MFODL [2], which allows users to express complex policies with non-trivial temporal dependencies, conditions, and event patterns. The major novelty of the monitor is that the policy verification tasks are automatically translated into *streaming dataflow computations*, which can then be parallelized and executed on Strymon [4], a streaming system we are actively developing in the Systems Group.

Thesis Goal

The thesis aims to contribute to our dataflow-based monitor for online policy verification in distributed systems. The student will build on top of the existing Strymon prototype and extend it in two directions: (i) *to ingest multiple streams of possibly out-of-order event logs* (instead of a single totally ordered stream), and (ii) *to evaluate temporal window operators incrementally*. Both extensions require changes in the verification algorithm as well as in the system architecture.

Evaluation

Real event logs are provided by our industry partners but for the purposes of the thesis we will also use synthetic data generators. The evaluation will focus on stressing the monitor to investigate its behavior, and ultimately improve its performance, with respect to three factors: (i) *input event rate*, (ii) *policy complexity*, and (iii) *level of parallelism*.

References

[1]: Elena Ferrari. *Access Control in Data Management Systems*, 2010.

[2]: G. Peycheva. *Real-time Verification of Datacenter Security Policies via Online Log Analysis*, Master Thesis, ETH Zurich, 2018.

[3]: P. Reynolds *et al. Pip: Detecting the Unexpected in Distributed Systems*, NSDI, 2006.

[4]: *Strymon: A Platform for Online Modelling of Enterprise Datacenter Behavior*, URL: strymon.systems.ethz.ch

If you are interested in this project please contact John Liagouris (liagos@inf.ethz.ch). The proposed thesis will be supervised by John Liagouris and Prof. Timothy Roscoe.