



Data-Driven Performance Correlation

Project thesis proposal

Correlating networking signals is of paramount importance for making sense of the large amount of measurements usually available to a network operator. As an intuition, measuring the loss rate of hundreds of flows per prefix in parallel is feasible yet observing and debugging each of those individually is not. On the other hand, extracting accurate and useful information that combines all of them is usually very complex. Simple techniques like aggregation can easily destroy the signal. For instance, the loss rate of a prefix can be very low while there is a group of flows that experienced extremely increased loss. Other approaches such as splitting traffic to groups based on their forwarding paths will also not work for the following reasons. Paths in the Internet are more and more difficult to predict both because of the variety of actual paths as opposed to AS paths and because of the inconsistency between the Control plane (BGP) and the actual paths used.

To deal with this problem, while leveraging the accurate measurements provided nowadays by hardware devices, we propose a clustering scheme which will be based exclusively on observed variances of the data plane signal. As such, the primary goal of this project is to design an algorithm to cluster traffic based on observed performance. The secondary goal is to investigate which parts of the clustering procedure can be migrated or approximated by the data plane.

Requirements

- Good Programming skills
- Preferably familiarity with machine learning techniques

Contact

- Maria Apostolaki, apmaria@ethz.ch
- Prof. Dr. Laurent Vanbever, lvanbever@ethz.ch

Tentative schedule

Week 1	• Literature research for similar techniques
Week 2	
Week 3	• Machine Learning techniques for clustering flows that were affected by the same incidents.
Week 4	
Week 5	• Build Mininet [1] or ns-3 [2] simulation for testing.
Week 6	
Week 7	• Iteratively test and tune algorithms for clustering.
Week 8	
Week 9	• Apply and evaluate algorithm in RIPE ATLAS [3] measurements.
Week 10	
Week 11	• Partially design/implement the clustering algorithm in P4 [4]
Week 12	
Week 11	• Write report
Week 12	

References

- [1] Mininet. <http://mininet.org/>.
- [2] ns-3. <http://nsnam.sourceforge.net/wiki/index.php/>.
- [3] RIPE NCC. <https://atlas.ripe.net/>.
- [4] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, and D. Walker. P4: Programming protocol-independent packet processors. *SIGCOMM Comput. Commun. Rev.*, 44(3):87–95, July 2014.