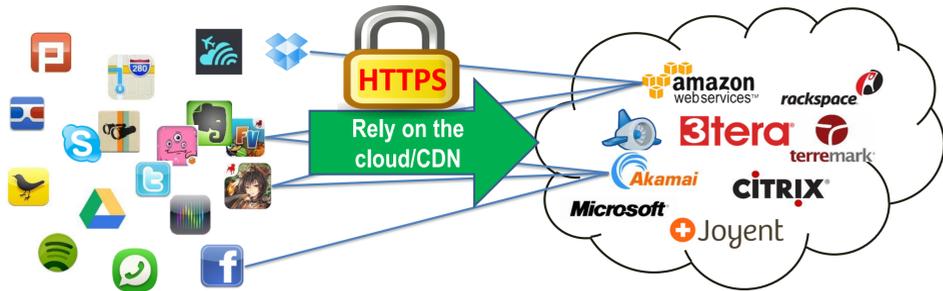


A Big Data approach for Web Service traffic classification and management

Martino Trevisan
Supervisor: Prof. Marco Mellia

Research Context

Cloud services have changed the way computing power is delivered to customers. The complexity of Web service management is outsourced to third-parties.

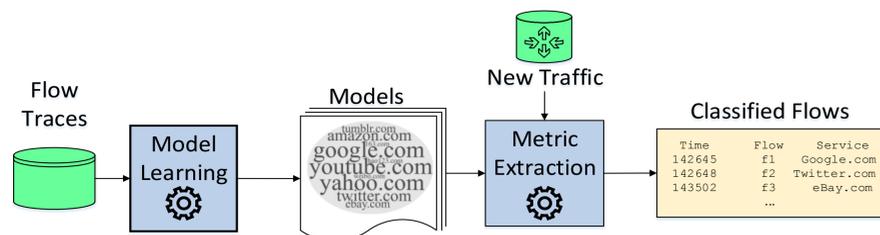


Web Service traffic classification

The Web Helper Accounting Tool (WHAT) classifies Internet traffic to unveil traffic generated by Web services. It combines big data and machine learning approaches to process large volumes of network flow measurements.

It is composed of two modules

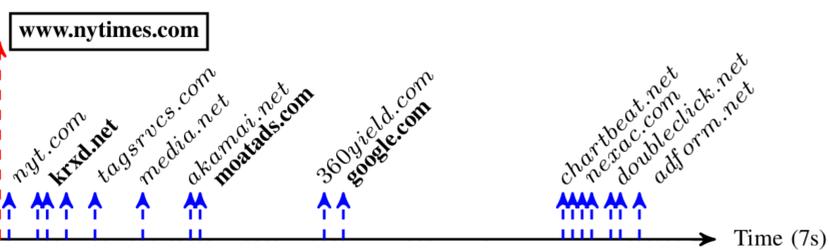
1. Automatically discovery of new Web Services
 - Characterization of the services to learn a model of the traffic generated by each of them
2. Use the models to account live traffic to the corresponding service



- Mines information from flow records datasets
- It leverages Machine Learning techniques to create the models
- It is implemented in state-of-art Big Data platforms (Hadoop and Spark) to scale with the size of datasets

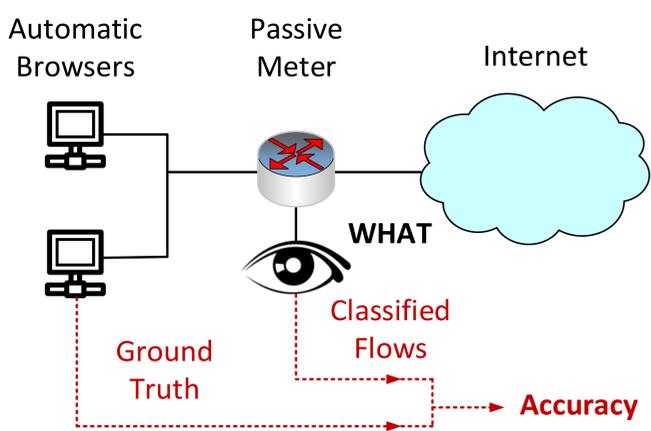
Key Idea: exploit temporal correlation among TCP flows

- When retrieving a Web Page several connections are issued
 - To retrieve images, CSS, JavaScripts
 - They can be used to define a **signature** for a given service



The goal of WHAT is to account all the traffic generated by a Web page as triggered it

Validation

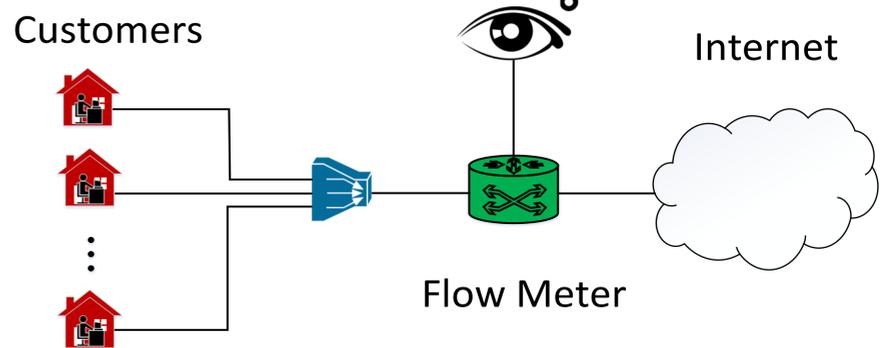


- We use real browsing histories from 30 volunteers
 - Different traffic mixtures are produced (number of users, services)
- The accuracy of WHAT is higher than **90%**. It means that almost all traffic is classified to the correct Web site that generated it.

Addressed Questions

Discovering and isolating the traffic of a specific service is a non-trivial task, which currently requires thorough and time-consuming manual work. Network administrators lose visibility about usage of Web Services

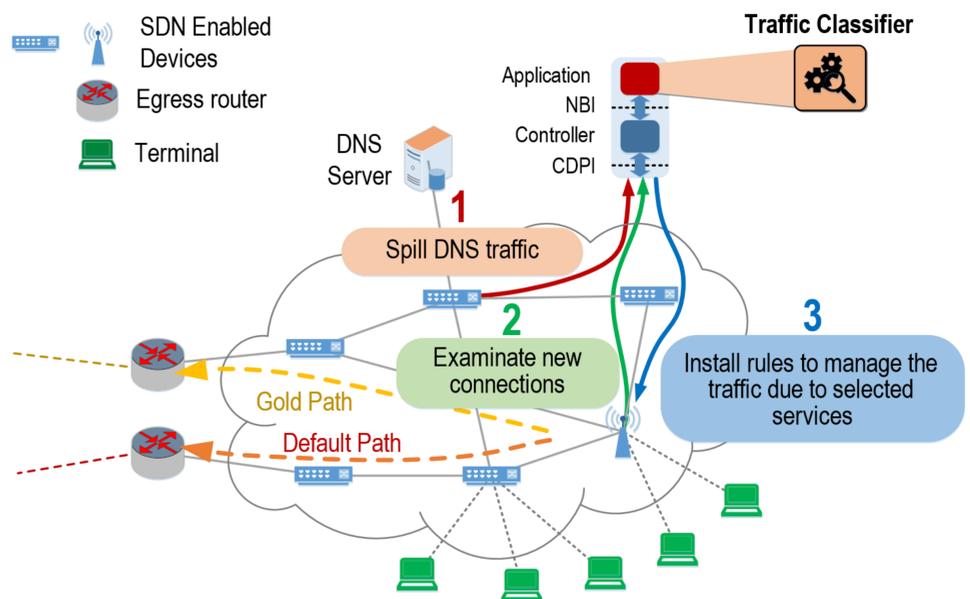
- Which services are users using?
- Are using only accredited ones?
- How can the network administrator manage traffic due to selected Web Services?



Per-Service Traffic Management

Combine Web Service traffic classification with Software Defined Network paradigm to allow per-Service traffic management

- The network administrator imposes policies based on the service being accessed, e.g., give priority to accredited Cloud Services while segregating News portals.
- The system identifies groups of flows related to the managed services, and, consequently, steers them along with the traffic due to related CDNs and cloud providers.
- Compliant with the basic SDN architecture and the latest version of OpenFlow.
- It requires no changes to existing APIs and SDN controllers, hence allowing adoption to existing SDN platforms to be simple.



Future work

- The traffic model can be refined to include services running **in background** and **mobile** applications
- Traffic patterns and correlation among TCP flows can be exploited to assess **Quality of Experience** enjoyed by Web Users
 - Can my ISP detect poor Web browsing performance?
 - How per-Service traffic management impacts QoE?
- Deployment in real operational networks of the Traffic Management prototype
 - Evaluate benefits and bottlenecks of the whole system