[Proceeding] Network Highlighter

Original Citation:
Giordano, Danilo; Traverso, Stefano; Grimaudo, Luigi; Baldi, Mario; Baralis, Elena; Mellia, Marco (2014). *Network Highlighter*. In: Traffic Monitoring and Analysis (TMA), Londra, 14 April 2014.

Availability:
This version is available at: http://porto.polito.it/2675282/ since: June 2017

Publisher:
IFIP

Terms of use:
This article is made available under terms and conditions applicable to Open Access Policy Article ("Creative Commons: Attribution 3.0"), as described at http://porto.polito.it/terms_and_conditions.html

Porto, the institutional repository of the Politecnico di Torino, is provided by the University Library and the IT-Services. The aim is to enable open access to all the world. Please share with us how this access benefits you. Your story matters.

(Article begins on next page)
Network Highlighter is fundamental to spot unusual and unknown behavior.

**Paramount task of network highlighter**
- Security
- Performance/Troubleshooting
- Traffic monitoring

**Network behavior and infrastructure change very fast**
- How to spot anomalies? What is normal and what is not?
- Reactive manual approach completely fails
- Need of automatic tools for anomaly detection in large scale networks
- CDNs/cloud systems make network even more complex: Akamai, YouTube, Amazon

**Our proposal is a distributed and comprehensive framework**
- To automatically spot anomalous traffic
- To provide administrators with a tool to "understand what is happening" in their networks
  
  E.g.: Capture sudden change in CDN (YouTube, Facebook, etc.) traffic patterns

---

**Our network highlighter workflow**

1. **Prediction**
2. **Classification**
3. **Filtering** (Feature extractor)
   - IP address, RTT, TTL, Port Number, service, device, etc.

**Anomalous**
- Security issue, performance problem, unusual redirect, etc.

**Normal**
- Useful to build baselines and normal traffic patterns

**Classification**
- Data mining and Clustering techniques: DBScan, Multidimensional Subspacing, Ad-Hoc clustering algorithms

**Filtering**
- Kalman filter, Linear/Gaussian Regression

---

**Preliminary Results on YouTube infrastructure**

**Clustering Technique**
- Three different clusters
  - A single IP address can be present in two clusters
- Four distinct clusters
  - A single client creates an outlier cluster
  - The outlier cause a wrong normalization
  - Automatic crosscheck still needed

**Multi-Dimensional Visual Technique**
- Easier to detect server classic behavior
- Harder to identify anomalies

---

**Classic clustering techniques are not adequate for network modelling, new ad-hoc solutions have to be developed**