Diagnosing Network-Wide Traffic Anomalies

Anukool Lakhina, Mark Crovella, and Christophe Diot

Abhishek Srivastava
Student ID: 861307778

May 31, 2017
CS 204, Spring 2017

Review:

This paper presents general method to detect, identify and quantify volume anomalies in real traffic. Author says that it is important to identify anomalies because they can create congestion in network and they also badly impact the end-user experience. The problem of detecting anomalies is not easy because it requires sophisticated backbone to monitor the traffic but ISP's only collect trivial information, processing collected data is also demanding task, ISP's do not have capability to process information fast to detect anomalies in real time and the collected data can be very high dimensional and noisy so it is hard to extract meaningful information from that.

The Authors main focus is identifying *volume anomaly* which is sudden positive and negative changes in an OD(Origin-Destination) flow's traffic. The reason being anomaly originating outside network will propagate from origin PoP(Point of Presence) to destination PoP and anomaly can be detected by collecting IP flow level traffic summaries on all input links at all PoPs and then applying temporal decomposition method to each OD flow but this is impractical. Author presented method which only observes link count to detect OD flow based anomalies.

Authors presented method utilizes PCA(Principal Component Analysis) to separate normal and anomalous network wide traffic and use it to do further diagnosis. After identifying principal axises, data is projected and normalized to capture temporal variation along those axis. Authors observed that for initial principal components only capture the typical patterns which are common across traffic on all links and lower principal components are better at exhibiting anomalous behavior. A simple threshold method was then used to separate normal and anomalous sets, projection on each principal component is analyzed and if found that it exceeds threshold then that and all subsequent principal axis are assigned to anomalous subspace and earlier ones are to normal subspace. After separating subspaces it is decomposed into normal and anomalous components then use this to diagnose volume anomalies. The link traffic is then projected onto these subspaces and separated into *modeled* and $residual\ traffic$. It was observed that occurrence of a volume anomaly will tend to result in a large change to $residual\ traffic$. It can be further used to identify the hypothesis and quantify it from set of possible anomalies.

Comments:

- This paper was focused on volume anomaly but anomalies can easily hide by not causing disruptions in traffic and will not detectable. It can rather be extended to do feature based anomaly detection which uses different feature information for detecting anomalies and i think it can make anomalies more detectable and not being able to hide in network traffic.