Mining Trends from Network Traffic Data

[1]Jennifer Li, [2]Saurabh Amin ,[2]Blaine Nelson, and [2]Suzanna Schmeelk

[1]Louisiana State University, Department of Computer Science

[2]University of California at Berkeley, Department of Electrical Engineering and Computer Science

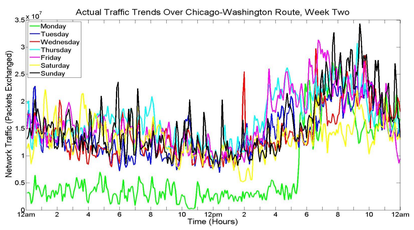
Abstract

In this project, we study the traffic behavior of a real Internet backbone, Abilene. Abilene is a complex communication network used by many universities and corporate institutions for educational and research purposes. Due to the heavy traffic flow of large networks such as Abilene, an attack could cause much destruction and therefore, security is of utmost concern. A crucial fundamental step to strengthening network security is to obtain, analyze, and understand the behavior of normal network traffic. To do this, we collected data on the number of packets exchanged on Abilene routes over an extended period of time. As a result, we used this data to map out the underlying traffic flow behaviors and found that the flows exhibit periodic trends. The insight we gained on the normal traffic flow behavior is important as it allowed us to emulate portions of Abilene traffic. This slightly varied but realistic traffic we generated will help us differentiate between normal and abnormal traffic, as well as develop successful defense strategies against attacks.

Methods

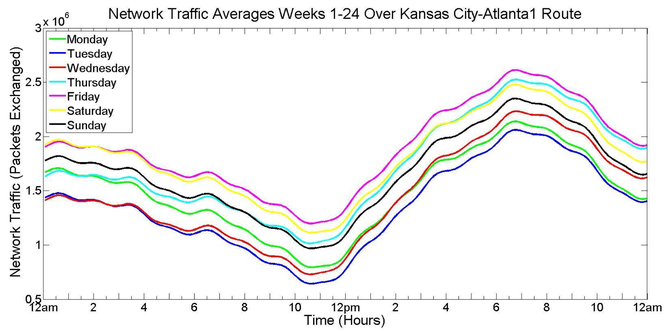
We obtained 24 weeks of actual traffic data from the Abilene network from 144 routes. Then we used Fourier's Transform to filter traffic noise and pull out the underlying trends of these routes, which were analyzed using the following methods:

1) Examination of general underlying trend behavior in data averaged over multiple weeks
2) Comparison of underlying trends from one network route over multiple weeks
3) Comparison of underlying trends from several network routes over a single week

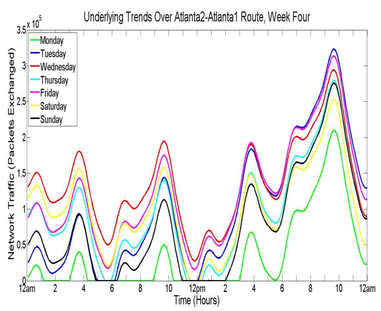


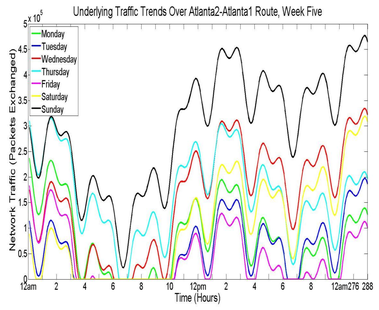
Observations

By averaging underlying traffic trend data from multiple weeks of a single route, we found that a general trend existed. The figure below shows this general trend over the average of 24 weeks of data

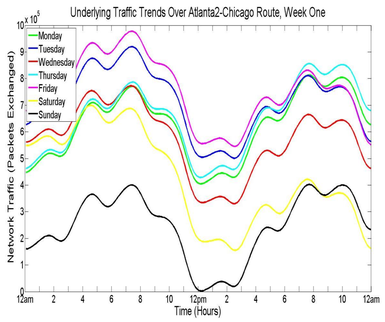


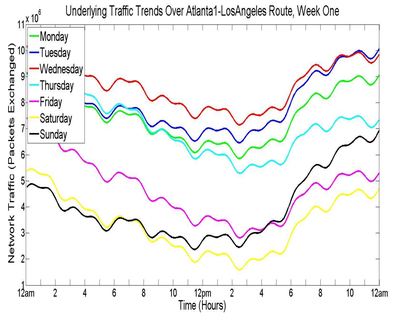
Analyzing the underlying trends of traffic data by separate weeks, we observed that daily traffic flow over each week exhibit periodic trends as well, but there was no weekly periodicity.

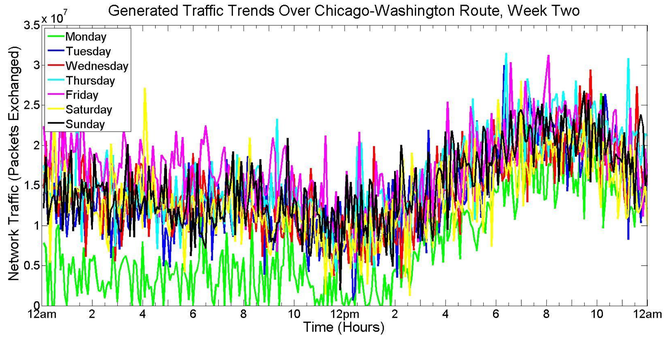


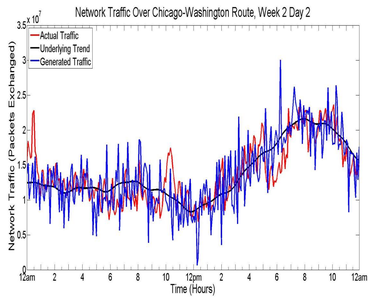


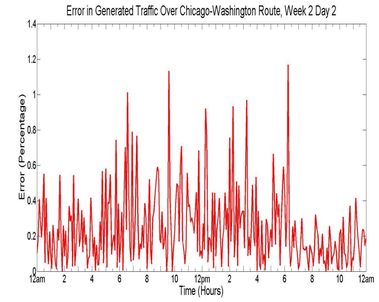
We also observed that the traffic on different routes over a single week exhibited different patterns.











Based on the periodic trends we found, we were able to generate Abilene Network traffic. We compared the generated traffic to the actual traffic and found the error percentage to be very minimal. Therefore, the portions of the Abilene Network traffic that we emulated are realistic generations.

Conclusions and Future Work

Daily traffic flow of the same route within each week do exhibit similar trends. This is an important find, since it will allow researchers to use these mappings for emulation of traffic models on a smaller scale. This realistic traffic emulation can then be used in simulated attack experiments, which will ultimately allow further exploration of defense tactics to make these systems more resilient to these threats.

Acknowledgements

Many thanks to my research advisors: Blaine Nelson, Suzanna Schmeelk, and Saurabh Amin, for their guidance and support. The programs used for realistic traffic generation were provided by Ben Rubinstein and the UC Berkeley SecML Research Group. This summer research opportunity was provided by the UC Berkeley TRUST REU Program.

This work was supported by the TRUST Center (NSF award number CCF-0424422)