A MATLAB® Toolkit for Spatial and Temporal Analysis of Network Traffic Anomalies and a Simulator/Emulator for Network Traffic Anomalies

Vidarshana Bandara, Ali Pezeshki, and Anura P. Jayasumana
Department of Electrical and Computer Engineering
Colorado State University
Fort Collins, CO 80523-1373, USA
{vwb, pezeshki, anura}@engr.colostate.edu

Abstract- *An easily customizable toolkit used to reveal spatial and temporal properties of network traffic traces and a simulator/emulator that regenerates anomalies having statistically similar anomalies to real networks is developed. The analyzer toolkit is fed with network traces as inputs, and anomalies are identified along with their properties. The toolkit uses Fourier analysis to suppress prominent trends in traffic that could mask anomalies, and use graph wavelets to reveal the spatial information on a node. It also traverses along the network to reveal the global spatial information. The toolkit also contains non-MATLAB scripts to convert certain network trace formats to a generic format. The simulator is parameterized with a few anomaly parameters at each node. By employing parameters for Internet2 network the simulator and the model it uses is validated. Further, an emulator for Internet2 traffic anomalies is also presented.*

Keywords- *Performance evaluation, MATLAB® Toolkit, Simulator/Emulator*

I. INTRODUCTION

Analyzing network traffic anomalies serves many applications like: network monitoring, emulators/simulators, robust network design. Since anomalies are embedded in regular traffic, traffic has to be processed prior to anomaly detection. Network traffic, especially Internet traffic tends is difficult to be well described using common statistical properties. However declaring when an anomaly is present requires some kind of characterizing regular traffic. The method presented in [1] de-trends for regular traffic enabling anomalies to be recognized thru a simple thresholding.

Under this research work, a toolkit for processing network traffic to find spatial and temporal behavior anomalies and simulator/emulator for network traffic anomalies is developed. The toolkit is an implementation of the methodology explained in [1]. It scans and analyzes traffic traces. During the process it suppresses the prominent trends enabling masked anomalies to stand out. Once the anomalies are detected their temporal features extracted by studying the time they are supported. Then spatial properties of each anomaly detected are revealed locally at nodes, and globally over the entire network. The simulator/emulator is based on the model presented in [1]. Here, anomalies of the network are parameterized with a few node variables.

Phases of the toolkit and the simulator/emulator are implemented using mostly MATLAB® scripts. Other phases include a few Bash scripts and minor ANSI-C codes. All the source codes are easily customizable and made publically available.

II. DEMONSTRATION

The demonstration will cover the following:

1. Deriving results claimed in all section of [1].
2. Guide on how the toolkit could be customized.
3. A simulator employing the model presented in [1].
4. An emulator for Internet2 anomalies.

The demonstration will exhibit the results generated for [1] at each phase of the study. This includes pre-processing Internet2 dataset, applying graph-wavelets to discover spatial properties, and deriving spatial-temporal statistics of Internet2 anomalies. Further, how other network traces could be processed will also be explained. The toolkit uses simple data abstraction enabling a great deal of flexibility. Thus toolkit can easily be customized for other similar work. The demonstration session will include a guide in customizing the toolkit.

Based on the model proposed in [1], an anomaly simulator will also be presented. The simulator is parameterized by a few variables per each node. This simulator is validated by building it to an emulator of Internet2 traffic anomalies. The simulator/emulator captures the statistical behavior of traffic anomalies of real networks, which are hard to characterize.

III. THE TOOLKIT

The toolkit was developed to utilize historical internet2 throughput data [2]. The data used were volume measurements from October 16th, 2005. The Internet2 network observed had 11 nodes with 28 in- and out-bound links. Each link is probed at both ends every 5 minutes.

Preprocessing:

The preprocessing required to extract the traffic volume information contained in RRD format [3] datafile. This phase is implemented in a Linux environment with Bash scripts and ANSI-C codes. The phase returns row data in a text file with the timestamp, inbound volume and outbound volume.

Anomaly detection phase:

Phases here forth are implemented on MATLAB®. Row traffic measurements are lumped to sets of weeks and processed using 2048 point FFT. The prominent trends form significant frequency coefficients. When those are forced to zero, and returned back to time domain, the prominent trends are no longer present in the traffic. Thus regular traffic appears nearly around a line. Here, anomalies tend to generate significant spikes which enable to be detected using a thresholding scheme. End of the phase another set of data is generated only carrying the anomaly information but still keep the same format as the row datafiles.

Temporal analysis:

Anomaly datafiles are analyzed to reveal temporal properties of anomalies. Since anomalies are rare occurrences, data is processed in groups of 10 weeks to have a large enough sample set to obtain stable statistical properties. Statistical properties are derived for the parameters listed for the model in [1].

Spatial analysis:

Each node is connected to multiple links. The contribution of other links to an anomaly present in one link is analyzed using Haar-wavelets. Then the anomalies are traced forward and backwards over the network to find the spatial distribution of each anomaly. The statistics of depths of the anomalies are then assessed.

IV. THE SIMULATOR/EMULATOR

The model introduced in [1] is employed on an anomaly simulator. The simulator is validated using the Internet2 dataset. As shown in [1] the anomalies of the entire Internet2 network is characterized using a few parameter values at each node. This constitutes the emulator for Internet2 network anomalies. It regenerates anomalies having similar statistical properties as anomalies of Internet2 network does.

FACILITIES NEEDED FOR THE DEMONSTRATION

Equipment to be used for the demo:

i. A laptop

ii. A LCD screen

ii. A Poster

Space:

i. 2'x4' to keep the laptop and the screen on a table

ii. 4'x4' area on a wall to put the poster up

Setup time:

20 mins

Additional facilities needed:

i. a table

ii. two power outlets

iii. internet access (wired or wireless)

REFERENCES

[1] V. Bandara, A. Pezeshki and A. Jayasumana, "Modeling Spatial and Temporal Behavior of Internet Traffic Anomalies," to appear in LCN 2010.

[2] http://noc.net.internet2.edu/i2network/live-network-status/historical-abilene-data.html

[3] http://oss.oetiker.ch/rrdtool/