PERFORMANCE ANALYSIS OF INTERNET TRAFFIC DATA FOR PRECISE RECAPTURING

Poornima Kebbepura R C, Department Of Computer Science And Engineering, Akshaya Institute Of Technology, Tumakuru, Karnataka
Manjesh B N, Assistant Professor, Department Of Computer Science And Engineering, Akshaya Institute Of Technology, Tumakuru, Karnataka

ABSTRACT

The inference of traffic volume of whole network from partial traffic measurements becomes increasingly critical for various network engineering tasks, such as capacity planning and anomaly detection. Analysis of Network traffic has applications in wide comprehensive set of areas and has newly attracted significant number of studies. Different types of experiments are conducted to identify various problems in existing computer network applications. Internet traffic data analysis is a proactive approach to ensure reliable, secure and qualitative network communication. Various techniques have been proposed and experimented for analyzing internet traffic data including Data mining techniques and Neural Network based techniques. Also various Linear and Non-Linear models have been proposed Network traffic prediction. In this paper we analyze performance of Internet traffic data based on the characteristics of Internet traffic data.

Keywords: Internet traffic data, Data mining

I. INTRODUCTION

Gaining a full knowledge of the traffic data volume between a set of origin and destination (OD) pairs in the networks becomes increasingly critical for a wide variety of network engineering tasks , including capacity planning, load balancing, path setup, dimensioning, provisioning, anomaly detection, and failure recovery. Although important, it is impractical to collect measurement data from a very large number of points in a large network at the fine time-scales. To reduce the cost, an alternative measurement strategy usually adopted by the network monitoring system is to take random measurement samples from the full traffic data. The actual data collected can be even less when experiencing data loss under severe communication and system conditions, such as network congestion, node misbehavior, transmission interference , and monitor failure. As many network engineering tasks require the complete traffic information or they are highly sensitive to the missing data, the accurate reconstruction of missing values from partial traffic measurements becomes a key problem, and we refer this problem as the traffic data recovery problem. Various studies have been made to handle missing traffic data. As most of the known approaches are designed based on purely spatial or purely temporal information, their data recovery performance is low. To utilize both spatial and temporal information, several recent studies model the traffic data as traffic matrices and propose matrix based algorithms to recover the missing traffic data. Although these approaches present good performance when the data missing ratio is low, their performance suffers when the missing ratio is large, especially in the extreme case when the traffic data on several time intervals are all lost. Tensors are the higher-order generalization of vectors and matrices. Tensor-based multilinear data analysis has shown that tensor models can take full advantage of the multilinear structures to provide better data understanding and information precision. Tensor-based analytical tools have seen applications for web graphs, knowledge bases, chemo metrics, signal processing, computer vision, anomaly detection.

II LITERATURE SURVEY

Traffic engineering and traffic matrix estimation are often treated as separate fields, even though one of the major applications for a traffic matrix is traffic engineering. In cases where a traffic matrix cannot be measured directly, it may still be estimated from indirect data (such as link measurements), but these estimates contain errors. Yet little thought has been given to the effects of inexact traffic estimates on traffic engineering. How well traffic engineering works with estimated traffic matrices in the context of a specific task; namely that of optimizing network routing to minimize congestion, measured by maximum link-utilization. Our basic question is: how well is the real traffic routed if the routing is only optimized for an estimated traffic matrix? We compare against optimal routing of the real traffic using data derived from an operational tier-1 ISP. We find that the magnitude of errors in the traffic matrix estimate is not, in itself, a good indicator of the performance of that estimate in route optimization. Likewise, the optimal algorithm for traffic engineering given knowledge of the real traffic matrix is no longer the best with only the estimated traffic matrix as input. Our main practical finding is that the combination of a known traffic matrix estimation technique and a known traffic engineering technique can get close to the optimum in avoiding congestion for the real traffic. We even demonstrate stability in the sense that routing optimized on data from one day continued to perform well on subsequent days. This stability is crucial for the practical relevance to off-line traffic engineering, as it can be performed by ISPs today.

III RESEARCH METHODOLOGY

Existing System:

It is impractical to collect measurement data from a very large number of points in a large network at the fine time-scales. To reduce the cost, an alternative measurement strategy usually adopted by the network monitoring system is to take random measurement samples from the full traffic data. The actual data collected can be even less when experiencing data loss under severe communication and system conditions, such as network congestion, node misbehavior, transmission interference, and monitor failure. As many network engineering tasks require the complete traffic information or they are highly sensitive to the missing data, the accurate reconstruction of missing values from partial traffic measurements becomes a key problem, and we refer this problem as the traffic data recovery problem.

Proposed System:

Based on the analysis of real traffic trace, we reveal that traffic data have the features of temporal stability, spatial correlation, and periodicity. To fully exploit the hidden structures for the data recovery, we model the traffic data as a 3-way traffic tensor, which allows us to combine and utilize the multi-mode (i.e. OD pair-mode, time-mode, and day-mode) correlations of data to better infer the missing data. To reduce the computation cost of the traffic recovery, Sequential Tensor Completion algorithm so that the tensor can be decomposed for the current data based on the tensor decomposition result of the previous traffic data.

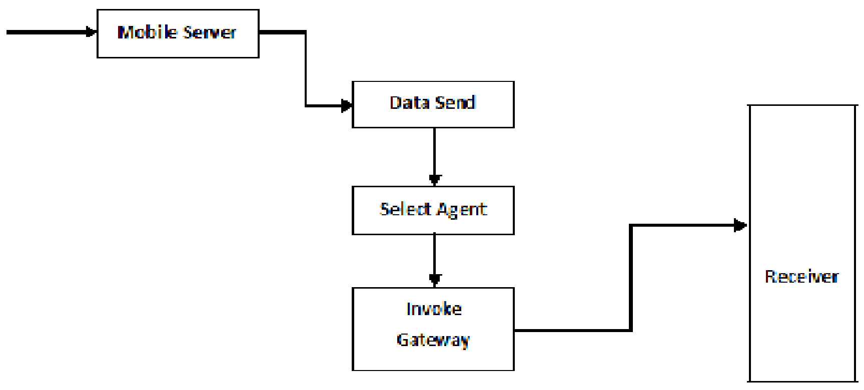
IV SYSTEM IMPLEMENTATION

Implementation phase plays a vital role in implementing the design requirements into the reality. Final result can be found in the implementation phase. Therefore, successful new system can be achieved in this key stage. Careful planning and controlling are required in this implementation phase.

Major decisions have to be taken care, before implementing any project. They are as follows:

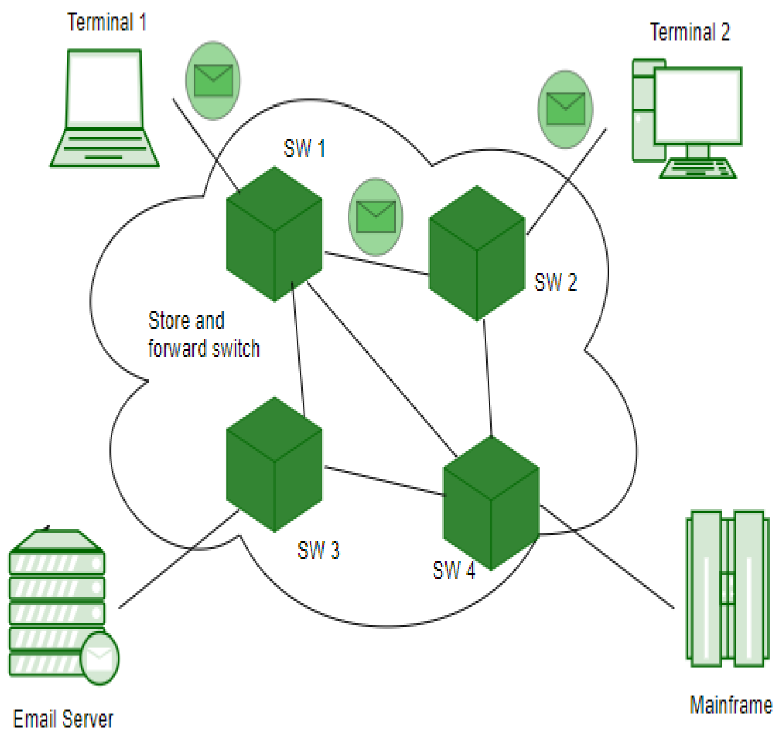
- Selecting the right programming language for the development of the project
- Selecting the right platform
- Coding conventions have to be followed properly

Data Flow

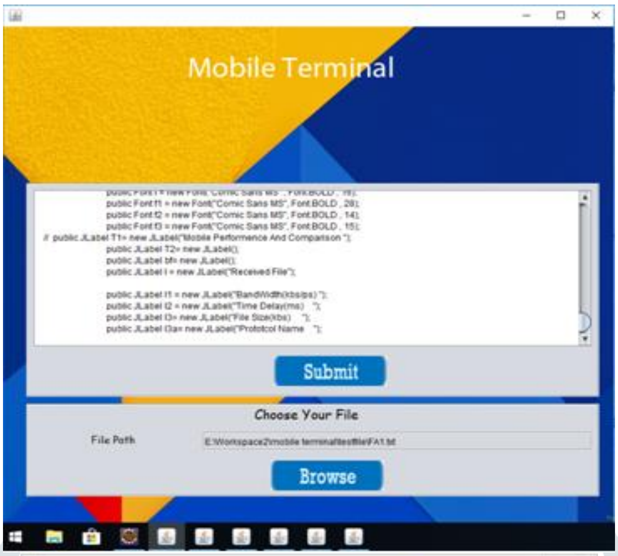


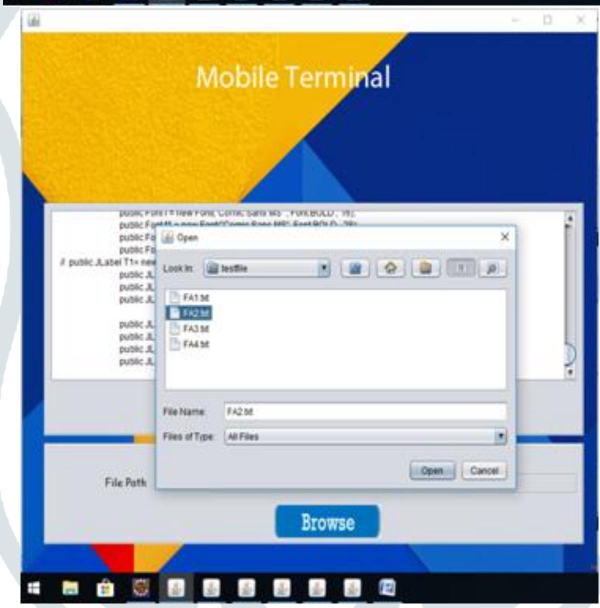
Direct Modeling of the traffic data using 3-way tensor with each mode corresponding respectively to the origin, destination and the total number of time intervals to consider. Then reuse the previous results of the tensor factorization, the derived tensor factorization should be able to capture the main features of both historical data and the current data based only on past data and partially observed new data, and the solution needs to be simple for implementation for online monitoring.

Architecture

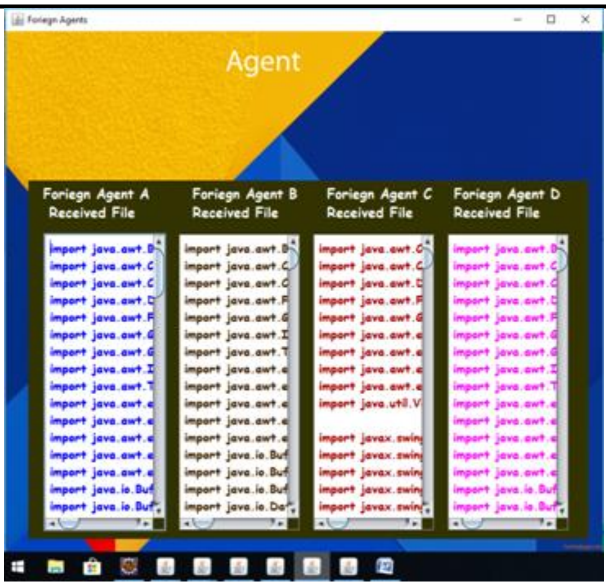


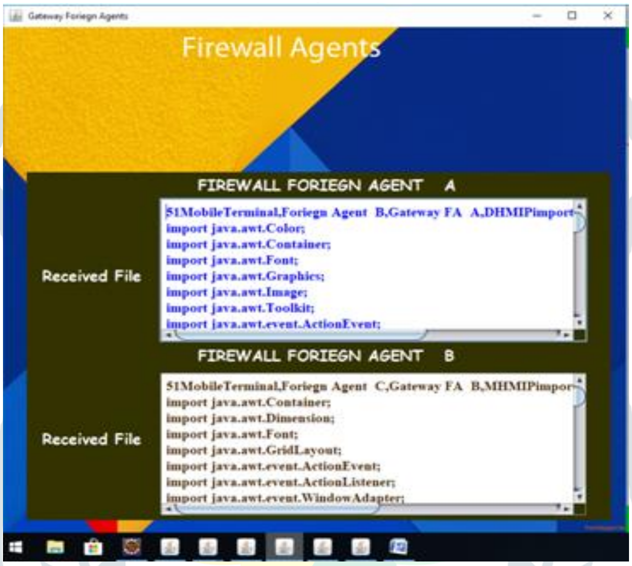
V RESULTS

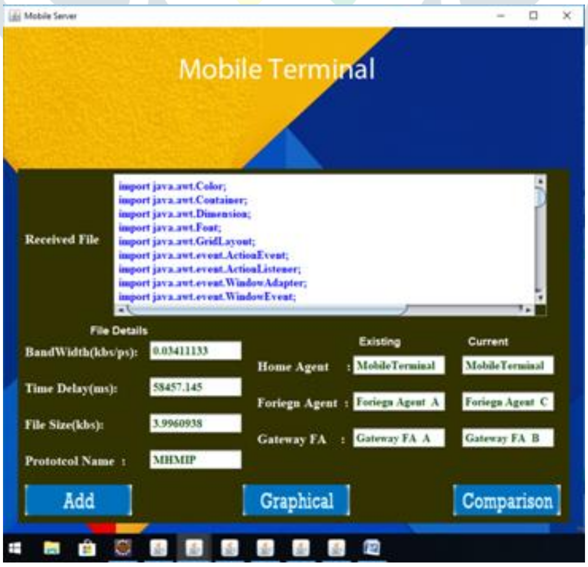

Mobile Terminal
Submit
Choose Your File
File Path
Browse

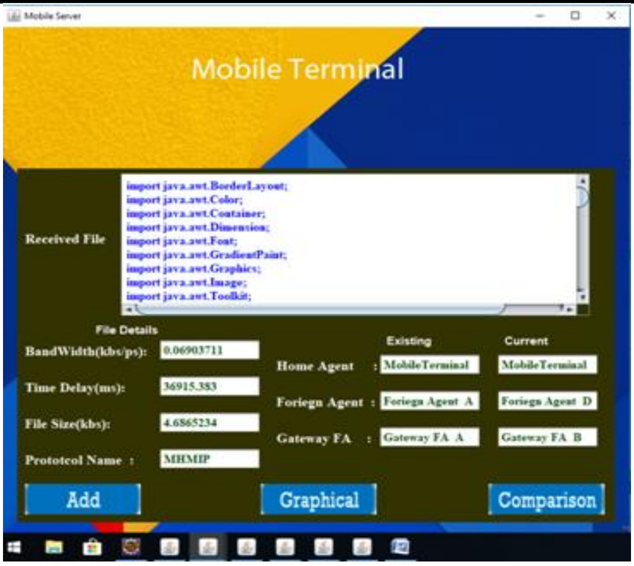

Mobile Terminal
Open
File Path
Browse

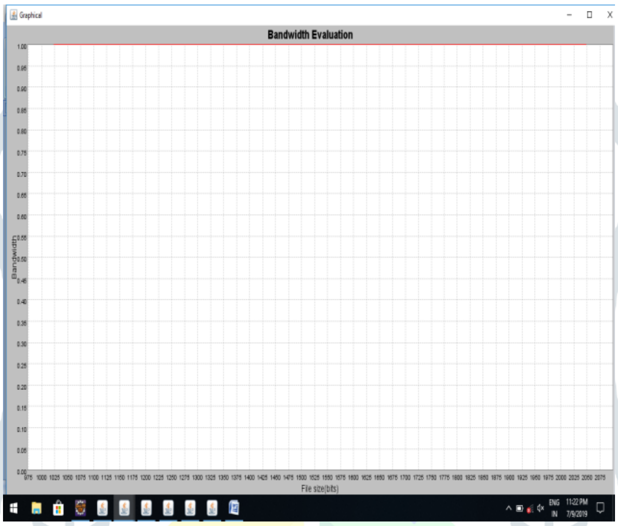

Mobile Terminal
Received File
File Details
Existing
Current
BandWidth(kbs/ps):
Time Delay(ms):
File Size(kbs):
Prototcol Name :
MIP
Home Agent :
MobileTerminal
Foriegn Agent :
Foriegn Agent A
Gateway FA :
Gateway FA A
Add
Graphical
Comparison

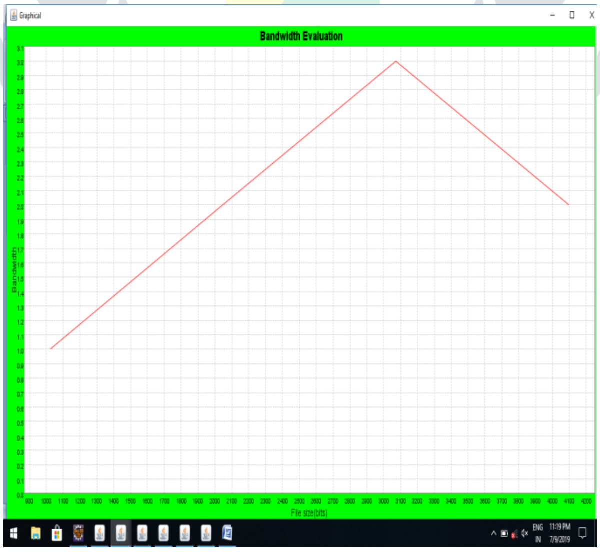

Agent
Foriegn Agent A Received File
Foriegn Agent B Received File
Foriegn Agent C Received File
Foriegn Agent D Received File

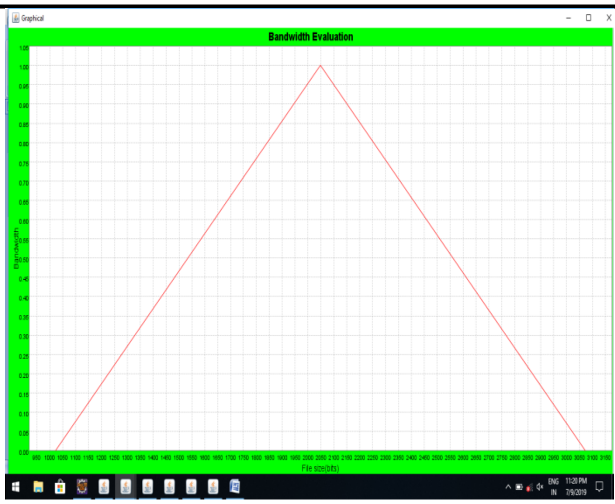

Firewall Agents
FIREWALL FORIEGN AGENT A
Received File
FIREWALL FORIEGN AGENT B
Received File

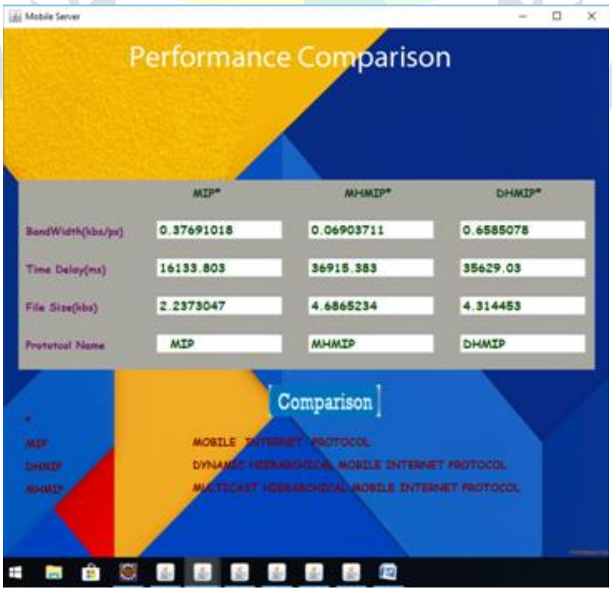

Mobile Terminal
Received File
File Details
BandWidth(kbs/ps):
0.03411133
Time Delay(ms):
58457.145
File Size(kbs):
3.9960938
Prototcol Name :
MHMIP
Existing
Current
Home Agent :
MobileTerminal
MobileTerminal
Foriegn Agent :
Foriegn Agent A
Foriegn Agent C
Gateway FA :
Gateway FA A
Gateway FA B
Add
Graphical
Comparison


Mobile Server
Mobile Terminal
Received File
import java.awt.BorderLayout;
import java.awt.Color;
import java.awt.Container;
import java.awt.Dimension;
import java.awt.Font;
import java.awt.GradientPaint;
import java.awt.Graphics;
import java.awt.Image;
import java.awt.Toolkit;
File Details
BandWidth(kbs/ps): 0.06903711
Time Delay(ms): 36915.383
File Size(kbs): 4.6865234
Protocol Name : MHMIP
Existing
Current
Home Agent : MobileTerminal MobileTerminal
Foriegn Agent : Foriegn Agent A Foriegn Agent D
Gateway FA : Gateway FA A Gateway FA B
Add
Graphical
Comparison


Bandwidth Evaluation
Bandwidth
File size(bits)


Bandwidth Evaluation
Bandwidth
File size(bits)





Conclusion

To well capture the spatial-temporal features inherent in the traffic data, we first analyze a large trace of real traffic data, and our studies reveal that the traffic data have the features of the temporal stability, the spatial correlation, and the periodicity. To fully exploit theses hidden structures for the data recovery, we model the traffic data as a traffic tensor which can combine and utilize the multi-mode correlations.

References

[1] K. Xie et al., "Accurate recovery of Internet traffic data: A tensor completion approach," in Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM), Apr. 2016, pp. 1–9.

[2] M. Roughan, M. Thorup, and Y. Zhang, "Traffic engineering with estimated traffic matrices," in Proc. ACM IMC, 2003, pp. 248–258.

[3] L. Fu, Y. Qin, X. Wang, and X. Liu, "Throughput and delay analysis for convergecast with mimo in wireless networks," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 4, pp. 768–775, Apr. 2012.

[4] K. Xie, X. Wang, J. Wen, and J. Cao, "Cooperative routing with relay assignment in multiradio multihop wireless networks," IEEE/ACM Trans. Netw., vol. 24, no. 2, pp. 859–872, Apr. 2016.

[5] K. Xie, X. Wang, X. Liu, J. Wen, and J. Cao, "Interference-aware cooperative communication in multi-radio multi-channel wireless networks," IEEE Trans. Comput., vol. 65, no. 5, pp. 1528–1542, May 2016.

[6] X. Deng, Q. Peng, L. He, and T. He, "Interference-aware QoS routing for neighborhood area network in smart grid," IET Commun., vol. 11, no. 5, pp. 756–764, Dec. 2016.

[7] A. Lakhina et al., "Structural analysis of network traffic flows," ACM SIGMETRICS, vol. 32, no. 1, pp. 61–72, 2004.

[8] Y. Zhang, M. Roughan, C. Lund, and D. L. Donoho, "Estimating point-to-point and point-to-multipoint traffic matrices: an informationtheoretic approach," IEEE/ACM Trans. Netw., vol. 13, no. 5, pp. 947–960, Oct. 2005

[9] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM), vol. 34. Oct. 2004, pp. 219–230.

[10] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," J. Amer. Stat. Assoc., vol. 91, no. 433, pp. 365–377, 1996.