CHAPTER I
INTRODUCTION

1.1 Background

The traffic matrix of network is essential to carry out better network management. These data consists of Traffic Matrix (TM), which represents the volumes of traffic between Origin and Destination (OD) pairs in the network [1]. However, direct measurement of the OD pair traffic is usually not feasible. Even good traffic measurement systems can suffer from errors and called missing data. So obtaining the OD traffic precisely is a challenge. Existing completion methods often perform poorly for network traffic estimation. Their recovery accuracy tends to be significantly worse when the data loss rate is high. Taking into account network traffic lower dimensional latent structure and traffic hidden characteristic, a vector is introduced to model a time series of pure spatial traffic matrix in this paper. To recover the missing entries in vectors of traffic matrix, a novel spatial vector completion method has been proposed. This approach not only takes advantage of vector decomposition and its lower-dimensional representation but also well takes into account traffic spatial properties [1]. Compressive sensing is a generic methodology for dealing with missing data that leverage the presence of certain types of structure and redundancy in data from many real world systems. Estimation of missing data in traffic matrix can significantly skew throughput calculations, including generating value greater than 100%.

In previous research in the field of communications systems, the others by Matthew Roughan of the University of Adelaide in Australia, Y.Zhang and L.Qiu from the University of Texas at Austin, USA, W.Willinger from AT&TLab-Research in Florham Park USA, where the title of the research is *Spatio-Temporal Compressive Sensing and Internet Traffic Matrix (Extended Version)* [7], the proposed interpolation techniques to accurately reconstruct missing data in TM based on partial and indirect measurements. If the communication is mediated by several links in series with different bit rates, the maximum throughput of the overall link is lower than or equal to the lowest bit rate. The lowest value link in the series is referred to as the bottleneck. Network data can often be arranged in the form of multidimensional arrays or called matrix. A key challenge that lies at the heart of many of these problems is how to cope with missing data that frequently arise in real world TM. The author proposes interpolation techniques to accurately reconstruct missing data in TM based on partial and indirect measurement [7]. In the process, an author provides a unified approach to several common tasks involving measurement and analysis of traffic matrix using estimation of traffic matrix.

The approach uses the first truly spatial and temporal model of TM with Sparsity Regularized Matrix Factorization (SRMF). Where borrows ideas from the active area of compressive sensing and exploits domain knowledge regarding TM that has accumulated over the years. In practice, it is challenging to reliably measure TM for large networks. Besides above research work, there are other studies which formulate data being processed as a form of the vector to estimate the missing data. Several effective heuristics have been proposed to exploit the sparse or low-rank nature of data [5]. Meanwhile, the mathematical theory of compressive sensing has also advanced to the point where the optimality of many of these heuristics has been proven under certain technical conditions on the matrix of interest. The results of analysis the algorithms used, author are able to do reconstruction to 90% better than other methods commonly used in the interpolation process. The algorithm is able to do the estimation of missing data in order to generate sufficient reconstruction value approaches the true value. But in this method has some drawbacks takes longer in the process of its computing as well as the process of determining the value of k is done in a random selection, where the value of k can minimize of performance classification.

In this research, model network traffic matrix as the vector pattern. Inspired by spatial compressive sensing [5], the proposed a novel spatial vector completion method to recover the missing entries in vectors of traffic matrix. First, in many networks, the TM is not directly observable, and can only be estimated through link load measurements. Such measurements, while linearly related to the TM itself, are not sufficient to unambiguously identify the true TM. Typically, the problem was posed as an underconstrained linear inverse problem, where the solution relied on a prior model of the TM [3]. Second, although many networks now collector sampled flow level measurements for at least part of their network, there are still serious impediments to reliable large-scale collection of TM data collection systems can fail, flow collectors often use an unreliable transport protocol, and legacy network components may not support flow collection or be resource challenged. Third, scalability requirements may mean that flow level collection does not occur at the edge of a network where we would wish it for true TM recovery [4] but often only on some subset of the routers. Recovery of the actual ingress-egress TM from such data is non-trivial. Finally, when finding an anomaly in a set of TM, often need to know the non-anomaly related traffic either for other network tasks or just so that we can infer the cause of the anomaly. The result is that any large set of TM measurements has some, and quite often, a significant number of missing data. Since many network engineering tasks that require TM are either intolerant or highly sensitive to missing data, it is important to accurately reconstruct missing data based on partial or indirect TM measurements. The main contributions of this paper include, a vector is introduced to model a time series of pure spatial traffic matrix, which preserve the multi-way nature of the network traffic matrix and extract the latent structure of traffic via vector factorization. By taking advantage of vector decomposition, which projects instances into a lower-dimensional latent space, and spatial information with model regularization, we propose a novel spatial vector completion method to estimate the missing traffic matrix. And through extensive experiments with real-world traffic trace data, the evaluations show that our method can accurately recover the missing data with very low estimation error.

1.2 Problems

Issues formulated in this thesis is traffic matrix estimation because of the many networks that need traffic matrix intolerant or very sensitive to missing data. It is important to accurately reconstruct the lost data based on measurements of a partial matrix or indirect traffic. The traffic matrix is not directly observable and can only be estimated through link load measurements. Such measurements, while linearly related to the traffic matrix are not sufficient to unambiguously identify the true traffic matrix. Typically, the problem was posed as an underconstrained linear inverse problem, where the solution relied on a prior model of the traffic matrix.

1.3 Objectives

The purpose of this research is to estimate the missing data on internet traffic matrix using compressive sensing method with Spatial Regularized SVD (SRSVD) Algorithm. For evaluation the method is based on Normalized Mean Squared Error (NMSE) to reconstruct the true value.

1.4 Scope of Work

Limitations of the research problems are:

1. The measurement is done using a network dataset by the previous research because of the difficulty of conducting measurements directly on the backbone network.
2. The estimation of missing data from various measurement, scenarios are already made. Estimation algorithm using Sparsity Regularized SVD (SRSVD) and will be designed

using computational software applications, such as the Optimization Toolbox in Matlab.

3. Dataset will be used is network Dataset Abilene from Matthew Roughan's, Professor of The University of Adelaide in the School of Mathematical Sciences.

1.5 Hypothesis

Compressive sensing is a generic methodology for dealing with missing values that leverage the presence of certain types of structure and redundancy in data from many real world systems. Missing traffic data estimation can significantly skew throughput calculations, including generating values greater than 100%. In previous research in the field of communications systems, the proposed interpolation techniques to accurately reconstruct missing values in TM based on partial and indirect measurements. The results of analysis the Sparsity Regularized Matrix Factorization (SRMF) algorithms used, author are able to do reconstruction to 98% better than other methods commonly used in the interpolation process. But in this method has some drawbacks takes longer in the process of its computing as well as the process of determining the value of k is done in a random selection, where the value of k can minimize of performance classification. In this research, despite much recent progress in the area of compressive sensing, with developing Sparsity Regularized SVD (SRSVD) using ℓ_2-optimization norm technique, which finds low-rank approximations of TMs that account for spatial properties of real TMs. Based that can be used to find solutions of SPL is consistent and best solutions to approach the SPL is inconsistent and SRSVD can be used to find the pseudo inverse and rank of a matrix, and the results provide a promising new direction for estimating such matrices. Expected by using this algorithm can perform a reconstruction of missing data with the amount of missing data and the error is small.