

**Accurate Recovery of Internet Traffic Data:A Sequential Tensor Completion Approach**

**Abstract:**

The inference of traffic volume of the whole network from partial traffic measurements becomes increasingly critical for various network engineering tasks, such as capacity planning and anomaly detection. Previous studies indicate that the matrix completion is a possible solution for this problem. However, as a 2-D matrix cannot sufficiently capture the spatial-temporal features of traffic data, these approaches fail to work when the data missing ratio is high. To fully exploit hidden spatial-temporal structures of the traffic data, this paper models the traffic data as a 3-way traffic tensor and formulates the traffic data recovery problem as a low-rank tensor completion problem. However, the high computation complexity incurred by the conventional tensor completion algorithms prevents its practical application for the traffic data recovery. To reduce the computation cost, we propose a novel sequential tensor completion algorithm, which can efficiently exploit the tensor decomposition result based on the previous traffic data to derive the tensor decomposition upon arriving of new data. Furthermore, to better capture the changes of data correlation over time, we propose a dynamic sequential tensor completion algorithm. To the best of our knowledge, we are the first to propose sequential tensor completion algorithms to significantly speed up the traffic data recovery process. This facilitates the modeling of Internet traffic with the tensor to well exploit the hidden structures of traffic data for more accurate missing data inference. We have done extensive simulations with the real traffic trace as the input. The simulation results demonstrate that our algorithms can achieve significantly better performance compared with the literature tensor and matrix completion algorithms even when the data missing ratio is high.

**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, we propose a Sequential Tensor Completion algorithm (STC) so that the tensor can be decomposed for the current data based on the tensor decomposition result of the previous traffic data. To more accurately recover the data exploiting the feature of the dynamic data, we further propose a Dynamic Sequential Tensor Completion algorithm (DSTC) based on STC. Both algorithms do not need to involve a complete tensor decomposition procedure for the current data, so the computation cost can be significantly reduced.

To evaluate the performance of our proposed algorithms, we have performed extensive simulations based on real traffic trace. Compared with existing tensor or matrix completion schemes, our algorithms can achieve significantly better performance in terms of several metrics, including the ratio of the recovery error, the ratio of the successful recovery, recovery loss, MAE, RMSE, and the computation time.