Correlation-Based Data Dissemination in Traffic Monitoring Sensor Networks

Antonios Skordylis
Birkbeck College
University of London
adonis@dcs.bbk.ac.uk

Alexandre Guitton
Birkbeck College
University of London
alexandre@dcs.bbk.ac.uk

Niki Trigoni
Birkbeck College
University of London
niki@dcs.bbk.ac.uk

ABSTRACT
In this work, we investigate the nature of spatio-temporal correlations in an urban traffic scenario, and how they can be exploited to reduce the cost of sensor data propagation to the gateway nodes, where users are connected. We conduct experimental analysis of our proposed algorithms using a real dataset of road traffic data generated in the city of Cambridge.

Keywords
Sensor Networks, Lossy Compression, In-network Data Reduction, Spatio-temporal correlations, Traffic Monitoring

1. INTRODUCTION
Road congestion in the UK costs of the order of £20bn per annum. Urban areas would considerably benefit from a sensor network infrastructure able to detect vehicle flow, speed and occupancy at high spatial and temporal resolutions. In this work, we investigate how to use spatio-temporal correlations inherent to traffic data to reduce the maintenance costs of the monitoring infrastructure. We compare the communication savings and computation costs incurred by two compression techniques, Wavelet and Fourier transform, which we run locally at sensor nodes with limited communication and processing capabilities. Our results reflect realistic urban traffic scenarios: they use real traffic flow data and run on a real sensor platform. We examine the compression rates of typical traffic time series given user-defined requirements for data accuracy. We propose distributed algorithms to identify spatio-temporal correlations, and exploit them to further reduce the cost of data propagation from the sensors to the gateway nodes. Wavelets and spatio-temporal correlations have been considered in previous work [3, 4, 2] albeit not in conjunction.

2. SCENARIO
We are investigating a sensor network architecture that includes multiple sensor nodes with limited communication, computation, storage and energy resources, which will operate in low duty cycles and will be primarily targeted at sensing tasks, i.e. measuring traffic flow or street occupancy.

Our traffic sensor deployment is located in the city of Cambridge, UK. It consists of 112 nodes that have been gathering readings since February 2006. It spans an area of roughly 3.5km × 8km. The sensors generate traffic flow (i.e. cars/min) and road occupancy readings every 5 minutes. We are focusing on minimizing the cost of disseminating traffic data from the sensor network to an off-site database.

3. MODEL
3.1 FFT and Wavelet Signal Compression
Each node in our deployment locally produces a large time series of traffic data that, if propagated in an uncompressed form, would require a substantial amount of energy to be spent. In the current work we study two techniques, namely the Fast Fourier Transform (FFT) and the Wavelet Transform (Daub4 and Daub8), in an attempt to evaluate their efficiency in compressing the road traffic dataset.

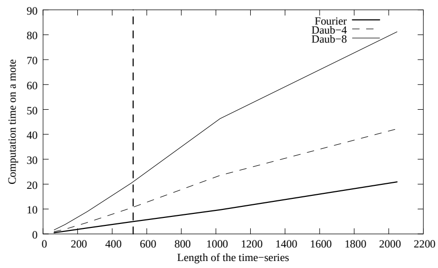

Figure 1. Computation time on a real sensor node.

Both the Fourier and the Wavelet transforms have been implemented for the TinyOS platform and tested on real sensor nodes. Figure 1 shows the computation time for both algorithms while varying the input, measured on the tmote sky [1] platform. Although the complexity of FFT is higher than that of the Wavelet transform ($O(n \log n)$ for FFT vs. $O(n)$ for Wavelet), FFT computes a time series approximation twice as fast as the Wavelet Transform using Daub-4 wavelet functions due to the fact that the latter requires mainly float operations, which are much slower on our sensor platform than operations on integers.

FFT is an attractive choice because it can easily be computed online in an incremental way, i.e. the arrival of a new sensor value

CoNext'06, December 4-7, 2006, Lisboa, Portugal.

does not force the system to recompute the whole transform, and it also facilitates the computation of linear correlations between two time series of traffic data.

The framework under development in the current work is modular, i.e. any compression technique can be used instead of the Fourier and Wavelet transforms.

3.2 Extent of Spatial and Temporal Correlations in Traffic Data

We explore the potential of exploiting spatial and temporal correlations in road traffic to reduce the volume of data propagated through the network. Let us assume that sensors forward their readings to one of the gateway nodes at regular intervals. The interval duration is configurable and it depends on the requirements of the application for data freshness.

Temporal Correlations: In traffic data, we observed very strong *temporal correlations* in the time series of a sensor node. During our initial experiments we observed that for more than 80% of the sensors, their daily readings are highly correlated with the readings of the same sensor in one of the previous seven days. This demonstrates the strength of temporal correlations in traffic data, and reveals a unique opportunity for exploiting these correlations to achieve communication savings.

Spatial Correlations: Another question that arises is whether we can equally rely on *spatial correlations* between the time series of different sensors, and whether the strength of spatial correlations depends on the physical distance between them. In order to determine whether two time series $u(t)$ and $v(t)$ generated by sensors u and v respectively are correlated, we allow a time-shift between the time series. Experiments on the traffic dataset revealed that node proximity is not indicative of high spatial correlation.

4. CORRELATION-AWARE DISSEMINATION ALGORITHMS

In this section we explore a class of dissemination algorithms that exploit relaxed user requirements for data accuracy to reduce the communication cost. Figure 2 shows a comparison.

Fourier Compression (FC): This algorithm, which is used as a basis for comparison, propagates the Fourier coefficients that constitute the compressed version of a node's time series to the closest gateway through the shortest path. Each node identifies the least number of Fourier coefficients k that can be used to reconstruct the time series with an error less than ε, without performing any further in-network reduction.

Fourier Compression and Temporal Correlations (FC-Temporal): This dissemination algorithm exploits temporal correlations to reduce the cost of time series propagation. Each node iterates through the time series of the previous days, computing the correlation factor between them and the current time series. As soon as it identifies a correlation factor that will lead to an error below ε, it selects the corresponding day, let us call it p, and evaluates the regression parameters of the linear function that evaluates today's times series based on day p's readings. Subsequently the node only forwards the regression parameters of the linear fit and p to the gateway.

Fourier Compression and Spatial Correlations (FC-Spatial): The next step is to exploit spatial correlations, i.e. correlations between the time series of different nodes on the same day. In this algorithm, the intermediate node receives approximate time series from its descendants in the tree, annotated with error values that show how the approximate versions deviate from the real ones. It subsequently attempts to correlate these time series, to reduce their size. For example, if the time series of two descendant nodes are strongly correlated (within a certain error), FC-Spatial sends Fourier coefficients to represent one of them, and regression parameters to represent the other. The coefficient of aggressiveness α specifies how the available error budget is divided between Fourier compression and linear correlation.

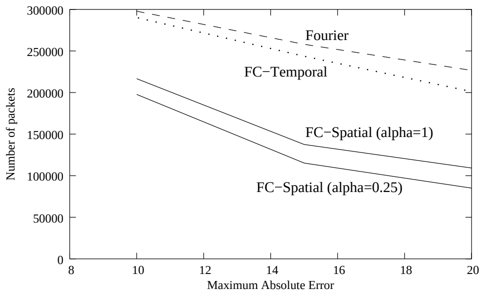

Figure 2. Communication cost vs. max absolute error.

5. FUTURE WORK

Apart from determining an efficient data reduction algorithm tailored to a traffic monitoring dataset, future work on this framework will investigate the problem of gateway placement, with the goal of further reducing the cost of data propagation from the sensors to the gateway nodes. Taking into account spatio-temporal correlations, we can place gateways at critical positions in the network where most data traffic is expected. Sensor placement is also under investigation, with the goal of reducing the installation cost of the sensor network. This can be done by identifying spatial correlations that hold consistently over time and removing nodes whose behavior can be accurately predicted.

6. REFERENCES

[1] http://www.moteiv.com

[2] Chu, D., Deshpande, A., Hellerstein, J.M., Hong, W. Approximate data collection in sensor networks using probabilistic models. In *Proceedings of the 22nd International Conference on Data Engineering (ICDE'06)*. IEEE Computer Society, 2006.

[3] Ganesan, D., Greenstein, B., Perelyubskiy, D., Estrin, D., Heidemann, J. An evaluation of multi-resolution storage for sensor networks. In *Proceedings of the first international conference on Embedded networked sensor systems (SenSys '03)* ACM Press, New York, NY, 2003, 89-102.

[4] Jain, A., Chang, E.Y., Wang, Y.F. Adaptive stream resource management using Kalman filters. In *Proceedings of the International Conference on Management of Data (ACM SIGMOD '04)*. ACM, 2004, 11-22.