Distributed Storage in Wireless Sensor Networks with Network Coding

Jan-Willem van Bloem, Richard J. Boucherie, Jasper Goseling, Geert Heijenk, Jan-Kees van Ommeren, and Roel Schiphorst

University of Twente

1 Project description

As part of the Wireless and Sensor Systems (WiSe): Real Time Enterprise project we provide a description of our ongoing research in the area of wireless sensor networks. In general, wireless sensor networks are prone to node failures, unreliable transmission links, interference and buffer overflows. In our project we intend to set up a reliable multi-hop communication for such harsh environments, where the main focus is on data persistence. In order to achieve these goals data redundancy in the network is required. The approach of adding redundancy in wireless networks corresponds to distributed networked storage which is relevant for wireless sensor networks where node failures are not uncommon. For this purpose we consider network coding to store the sensed data in a robust and efficient way, while at the same time minimizing communication costs in the network. Network coding is a technique invented to achieve efficient multicast communication by allowing coding inside the network which means that packets are mixed at the intermediate nodes. This can be viewed as applying encoding operations on the node's received data from the incoming channels whereas the resulting encoded information is sent over the outgoing channel. At the sink nodes the multicast information is retrieved by decoding the received information.

Hence by using distributed storage with network coding we aim to sense the data at the highest possible rates and to reconstruct the information by querying *any arbitrary subset nodes of K storage nodes in a network of N storage nodes*, where $K \leq N$. Note that storage nodes can operate as relay nodes as well. Furthermore we require the datacollector to reconstruct all stored data at the K storage nodes at once with high probability, as long as the sensing rates are inside the network-layer capacity region.

A possible application of distributed information storage in wireless sensor networks is environment monitoring and data persistence in a logistical environment.

2 State of the art

In [1], Ahlswede *et al.* show that network coding can achieve multicast capacity in a network which outperforms traditional routing techniques. A description of decentralized network coding is given by Ho *et al.* where the coefficients are chosen locally from a finite field in a randomized fashion [3].

The main research on network coding in a distributed storage context is introduced by Dimakis *et al.* [2]. The key idea in their work is that the code repair problem for distributed storage can be mapped to a multicasting problem, which enables them to solve the storage minimization problem. In [4], they study a scheduling algorithm called backpressure for multicast communication with network coding. Here, the sensed data is routed over the network in a multi-hop fashion to the receivers in the multicast session where routing decisions are made based on the local queue lengths according to the backpressure policy.

The most common schemes for data persistence in networks such as Reed-Solomon, LDPC codes and more recently fountain codes offer fast encoding and decoding times, but are centralized solutions.

3 Specific Contribution

In our work we elaborate on the idea proposed by Dimakis *et al.* formulating the distributed storage problem as a multicast problem for acyclic directed graphs with queue stability for all nodes in the network as the performance metric. Our contribution concerning distributed storage is as follows:

- Formulation of the (network-layer) capacity region for the distributed storage case, by defining the flow constraints concerning distributed storage and network coding. In other words we seek the set of feasible data sensing rates for which we can multicast to an arbitrary subset of K storage nodes whereas these rates should be supported by the network given a particular time-varying channel state model.
- We observed that the subset of K storage nodes with the lowest minimum cut determines the maximum flow at which the data can be offloaded to the datacollector. Hence, we derived the maximum rate at which a datacollector can reconstruct the data for any arbitrary subset of K storage nodes.
- In order to achieve all points in the (network layer) capacity region an algorithm is required. A formulation of the dual problem is obtained by Lagrange relaxation which leads to a decentralized solution. It turns out that solving the dual optimization problem by means of an iterative algorithm corresponds to the backpressure algorithm, which works in conjunction with network coding [4]. Hence we opt for random linear network coding in order to operate in a complete decentralized manner [3].

4 Future interest

The plan is to extend the distributed storage scheme by incorporating node failures into the model. Hence, in case of a node failure we can not assume independent channel models per link, wherefore a new stochastic model is required. Moreover, the correlative nature of a wireless sensor network lends itself for distributed source coding, which we consider beneficial in combination with distributed storage.

Besides the (network-layer) capacity region for different channel models can be observed, taking into account fading and interference. Furthermore, the model can be built out considering node mobility. From a practical point of view it would be interesting to verify the theoretical model with simulations to show the actual performance of the backpressure algorithm with respect to queue stability for distributed storage with network coding. Another option is to simulate for different time-varying channel state models with particular interest for incorporating the node failure model. Finally a protocol implementation of the model can be considered.

References

1. Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.
2. Alexandros G. Dimakis, Vinod Prabhakaran, and Kannan Ramchandran. Decentralized erasure codes for distributed networked storage. *IEEE/ACM Trans. Netw.*, 14(SI):2809–2816, 2006.
3. Tracey Ho, Muriel Médard, Ralf Koetter, David R. Karger, Michelle Effros, Jun Shi, and Ben Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, 2006.
4. Tracey Ho and Harish Viswanathan. Dynamic algorithms for multicast with intra-session network coding. *IEEE Transactions on Information Theory*, 55(2):797–815, 2009.