# Abstract

In this thesis we study a discrete-time stochastic process modeling the collection of data in a multi-hop sensor network that we call the *Data Collection Process*. We assume data is generated at the sensor nodes as a discrete-time Bernoulli process. All nodes in the network maintain a queue and relay data, which is finally collected by a designated sink. We prove that the resulting multi-dimensional Markov chain representing the queue size of nodes has two behavior regimes depending on the data generation rate. In particular, we show a non-trivial critical value of data rate below which the chain is ergodic and converges geometrically to a stationary distribution and above which it is non-ergodic, i.e., the queues at the nodes grow in an unbounded manner.

We then extend the Data Collection Process to a multi-commodity setting to provide a family of bounds for the rate at which the functions of many inputs can be computed in-network on general topologies. Going beyond simple symmetric functions where the output is invariant to the permutation of the operands, e.g., average, parity, we describe an algorithm that is analyzed to provide throughput lower bounds for the general functions, particularly the ones with a binary tree schema. Our lower bounds depend on the underlying graph and schema parameters, and we show that these are optimal for the complete network topology, the star topology, and the hypercube topology.

We also present an interesting connection between the Data Collection Process and Laplacian systems of equations. In particular, we show that the steady-state equation of the stochastic process is equivalent to the special class of *one-sink Laplacian systems* of the form $L\boldsymbol{x} = \boldsymbol{b}$ where exactly one of the coordinates of $\boldsymbol{b}$ is negative. As an application of this connection, we first present a completely new approach to solve Laplacian systems using queueing networks, which marks a significant departure from the existing techniques mostly based on graph-theoretic constructions and sampling. Our Laplacian solver can be used to adapt the approach by Kelner and Mądry (2009) to give the first distributed algorithm to compute approximate random spanning trees efficiently. We also extend our algorithm for solving the Laplacian system of equations on strongly connected directed graphs. This is the first distributed algorithm for this problem that can be analyzed in terms of the underlying graph's parameters.

As a second application of the Laplacian connection of the Data Collection Process, we propose a group-to-group version of random walk betweenness centrality. In particular, we relate our proposed metric to the steady-state equation of the Data Collection Process, which allows us to use a Laplacian solver to compute it efficiently. Empirical evaluation on real-world networks proves that our proposed metric is more effective than all natural baselines for controlling the spread of infection from an infected group to a vulnerable group.