Summary of the doctoral dissertation
"Probabilistic methods for analysis of complex networks"

Karol Krzysztof Gotfryd

Thesis supervisor: Prof. dr hab. Jacek Cichoń

Algorithmics of distributed systems relies heavily on mathematical methods. They are used for designing of efficient procedures, as well as for theoretical analysis of algorithms, incl. proofs of correctness, analysis of their complexity and the precision of results. An important role play randomized algorithms, which can be used for symmetry breaking or efficient processing of large datasets. In many applications obtaining the exact results is computationally expensive and it often suffices to calculate some approximation. The execution of randomized algorithms can be modelled by stochastic processes. Their properties such as time or energy complexity are formally described by some random variables. The approximation errors of calculated results are also of a probabilistic nature.

In the doctoral dissertation we present the examples of applications of probabilistic methods and tools from graph theory and information theory for designing and formal analysis of distributed algorithms for complex networks, consisting of multiple small devices with limited resources. We pay particular attention to the correctness of the protocols, their efficiency and the precision of outcomes.

In the thesis we study two classes of algorithmic problems, incl. distributed data aggregation and the problem of location hiding for mobile agents. For the formal description of networks we use the theoretical models known from the literature. The networks are represented by simple, connected, undirected graphs.

In the first chapter we introduce the basic notation and recall some definitions and facts from probability theory and information theory. We discuss in more details the extrema propagation technique and the concept of probabilistic counters, which are the basis of our algorithm for distributed data aggregation.

The second chapter is devoted to the problem of distributed data aggregation. We focus on computing an average of the values stored by the nodes in the network. We define an estimator, which can be easily calculated in large networks in a fully distributed manner. The algorithm proposed by us is based on building an approximate histogram of observed data using probabilistic counters. In this approach we use counters, that are random variates from an inverse gamma distribution. We perform a detailed theoretical analysis of our algorithm. We start with introducing the definition of approximate histograms. Then we prove an auxiliary lemma characterizing some properties of the probabilistic counters. Based on the classical tools from probability theory and some analytical methods, incl. delta method and Laplace's method, we analyse asymptotic properties of the proposed estimator. We show that it is asymptotically unbiased and derive a formula for its variance for the case of a histogram with two bins. Generalization of this result is left as an hypothesis. We analyse the estimator's precision, taking into account two types of error. We define the measure of precision and show its basic properties. Then we identify the distribution of data that maximizes the variance of the estimator. We also analyse the expected message complexity of our averaging procedure. We discuss the results of numerical simulations of the algorithm.

In the third chapter we present an example of using the tools from probability theory, information theory and graph theory for analysis of the problem of location hiding of mobile agents in dynamic systems. We begin with the formal definition of a theoretical model, incl. the model of an adversary. Based on the notion of uncertainty coefficient we define the measure of the amount of revealed information and introduce the notion of hiding algorithms. We study different scenarios, depending on the resources available to the agents (knowledge of the graph of a network, operational memory, access to random bits). We propose some algorithms and perform their formal analysis using probabilistic methods, incl. the theorems that characterize the properties of random walks and the entropy of random variables. For some scenarios we prove lower bounds on the number of rounds required for hiding the information of the agents' initial location. We also show, that there is no deterministic hiding algorithm for an oblivious agent.

For each studied problem we present an overview of the most important results known from the literature. We also discuss some open problems and possible directions for further research.

The doctoral dissertation is based on the results from three author's publications. The analysis of the distributed averaging algorithm is complemented by the proofs of theorems on the asymptotic properties of the proposed estimator. In the discussion on the problem of hiding the agent's locations, we modified the original definitions of hiding algorithms. The new approach seems to be more natural. That required adjustments of the statements and proofs of some theorems. In some cases we use new, more general arguments. We have also clarified some details of the theoretical models and eliminated minor issues.

04.12.2020r. Karol Gotfryd