論文の内容の要旨

Framework for Fault Tolerance in Dense Data Centric Wireless Sensor Networks

(高密度データ無線センサネットワークのための耐故障性フレームワークの提案　)

ヴァレンティナ　バリャク

Primary task of wireless sensor networks is to gather and process observations about an entity or phenomena of interest. Reliability of the network depends entirely on the accuracy of the inference, which in turn is closely related to the accuracy of sensory readings. Over the time, different types of faults develop in the network, affecting the accuracy of sensory readings. Since reliable and independent functioning of the network is one of the key demands for the quality of service, it is crucial to minimize the number of erroneous readings that could affect the final result and functionality of the backend application.

Wireless sensor networks consist of numerous relatively inexpensive devices that are capable of creating rather complex systems. Often these networks are deployed in harsh environments and they have to meet demanding requirements. Sensors and actuators, vital parts of sensor node, are constantly exposed to the environment so the inherent vulnerability to failures is constantly present. Faults are a common occurrence rather than an exception. Development of faults is an inevitable event, even if nodes and network are perfectly calibrated in the beginning. Cumulative effects of these faults can shorten the effective lifetime of a network, the period in which the network provides accurate and reliable data. Quick recovery is the key for maintaining autonomous functioning over the prolonged period.

Faults in wireless sensor networks can be a result of various causes, such as malfunction or environmental noise and they can be transient or lasting. A fault can occur at any level of a network stack, but the effects can propagate and cause failures and errors even at some distance form the fault's site. Fault tolerance is a complex task, and it can focus on different aspects of the functionality in the network, for example, fault tolerant routing protocols or fault tolerant event detection. The focus of this work is on faulty readings and the detection, classification and correction of faults by observing the behavior of data.

The effects different types of faults produce in the readings also fall into a wide range of categories. Some types of faults can render data unusable, while some can still provide additional information about the system and possibly even point out to the cause of the fault. However, the main result is mostly related to the accumulation of faults, which can lead to the progressive decrease of reliability and accuracy of sensor readings and shorter effective lifetime of the network.

In this work we maintain the focus on faults and types of faults in sensory readings. The main question is whether a fault can be detected through observing the trace it leaves in collected data, independent of the underlying cause. This means that we can observe a fault through the pattern it leaves in collected data. In terms of duration, persistence and observable and learnable pattern, faults can be detected and classified.

The proposed method requires certain redundancy of readings, including the spatial and temporal correlation, while maintaining the assumption of uncorrelated development of faults in different nodes. Because of this, the best applicability is for data-centric, densely deployed wireless sensor networks. In this type of network, multiple sensors collect data in the same area at the same time, which preserves mentioned spatial and temporal correlation amongst readings and provides sufficient information for statistical methods to work.

Some of the key requirements for the complete fault tolerance scheme are quick self-recovery, flexibility and low cost. Full cycle consists of detection, diagnosis, resiliency mechanisms and fault models. To address these issues, this thesis proposes a complete and consistent *fault classification* based on two criteria, frequency and continuity of the occurrence and observable and learnable pattern in the data. This classification is used in the detection and diagnosis phase, and it enables handling of more than one type of fault at the same time. Further, we propose *model learning* based on the given classification. Model of the fault can be learned from the observed pattern as a function, and applied to the correction of the fault. It is important to note that in this system model of fault is expected to be applicable only for the specific sensor node. Finally, to complete the cycle, elements are integrated into a flexible *fault correction framework*. This framework uses learned models to handle faulty sensors in an appropriate manner, incorporating those models into improved assumptions about the network's behavior. The method gives the flexibility to address each faulty node individually without the big increase in overall communication cost. Even though the proposed method is centralized, it maintains some benefits of in-node solutions for fault correction.

The framework itself can be implemented with a different combination of algorithms for each stage of the process. In the current evaluation, methods for each phase have been chosen based on the applicability and simplicity. Detection phase relies on statistical features of data, analysis of time series and a classic neighborhood vote technique. Considering that a fault can belong to a one of a small set of classes, for the classification phase we use statistical pattern recognition, more specifically, decision trees. Finally, in model learning phase, we use regression. However, this is one limitation of the system, since currently it can only learn a model of a polynomial form. For the complete evaluation, we have used analysis of dataset, simulation and finally a small testbed deployment. Results so far maintain that the method is capable of satisfactory detection and classification of faults, as well as of learning the polynomial form of the fault model. Correctly classified faults in readings are either adequately corrected or their negative effects are eliminated from the data.