# Visualizing Anomalies in Sensor Networks

Qi Liao*, Lei Shi†, Yuan He‡, Rui Li§, Zhong Su†, Aaron Striegel*, Yunhao Liu◇

*University of Notre Dame, USA. E-mail: {qliao,striegel}@nd.edu.
†IBM Research, China. E-mail:{shllsh,suzhong}@cn.ibm.com.
‡Hong Kong University of Science and Technology. E-mail: heyuan@cse.ust.hk.
§Xi'an Jiao Tong University, China. E-mail:rli@mail.xjtu.edu.cn.
◇Tsinghua University & Hong Kong University of Science and Technology. E-mail: liu@cse.ust.hk.

## ABSTRACT

Diagnosing a large-scale sensor network is a crucial but challenging task due to the spatiotemporally dynamic network behaviors of sensor nodes. In this demo, we present Sensor Anomaly Visualization Engine (SAVE), an integrated system that tackles the sensor network diagnosis problem using both visualization and anomaly detection analytics to guide the user quickly and accurately diagnose sensor network failures. Temporal expansion model, correlation graphs and dynamic projection views are proposed to effectively interpret the topological, correlational and dimensional sensor data dynamics and their anomalies. Through a real-world large-scale wireless sensor network deployment (GreenOrbs), we demonstrate that SAVE is able to help better locate the problem and further identify the root cause of major sensor network failures.

## Categories and Subject Descriptors

C.2.3 [**Computer Systems Organization**]: Computer-Communication Networks—*Network Management*; H.5.2 [**Information Interfaces and Presentations**]: User Interfaces

## General Terms

Management

## Keywords

Wireless sensor networks, diagnosing, anomaly detection and analysis, visualization

## 1. INTRODUCTION

Sensor networks play a pivotal role in numerous modern and future industries such as smart grid, logistics and healthcare. However, diagnosing sensor networks can be tremendously challenging when faced with the issues in full scale deployments: First, sensor networks consisting of numerous low-end embedded devices are resource-constrained, especially its non-rechargeable batteries, making it hard to deploy the approaches of traditional networks (e.g. SNMP). Second, most sensor networks are deployed in outdoor or even hostile environments, thus greatly affects the performance and reliability of sensor networks. For example, the changes in temperature and humidity, the wind and rain, and the physical damages by human and wild animals, may all cause performance degradation or even system failures in sensor networks. Third, the low-power wireless communication of sensor networks is likely to
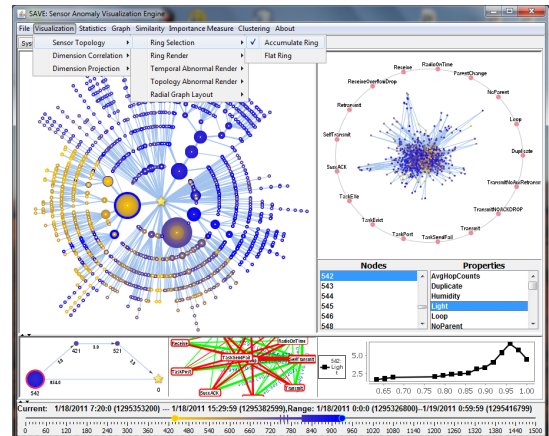
Figure 1: SAVE: Analyzing spatiotemporal anomalies for wireless sensor networks.

be lossy in certain environments and hence cannot guarantee perfect data collection.

The current algorithms to diagnose sensor networks from the collected data mainly rely on certain inference models which link symptoms to the underlying root causes. They depend on the evidence-based sensor data fault taxonomy [1] to characterize the symptom. However, due to the resource constraints and the existence of salient failures, such inference models could only focus on a portion of symptoms, restricting their applicability and efficiency in real usage.

In this work, we present Sensor Anomaly Visualization Engine (SAVE) (Figure 1), an integrated system that tackles the sensor network diagnosis problem using both visual analytics technologies. Compared to simple algorithmic approaches, our method has the following advantages:

- The visualization of sensor network data from multiple perspectives, at large scale and in long term helps the administrator collect evidence in a more comprehensive manner. In our system, we introduce several novel visualizations to illustrate all the three time-varying high-dimensional sensor data facets – routing topology, sensor networking status and physical sensor readings.

- Rather than derive hypothesis directly with the inference model, visual analytics solution focuses on collecting better evidence for the human to improve the decision-making process. This extra robustness makes it a better choice in diagnosing salient sensor network failures, whose symptoms and root causes are previously unknown to network operators and researchers.

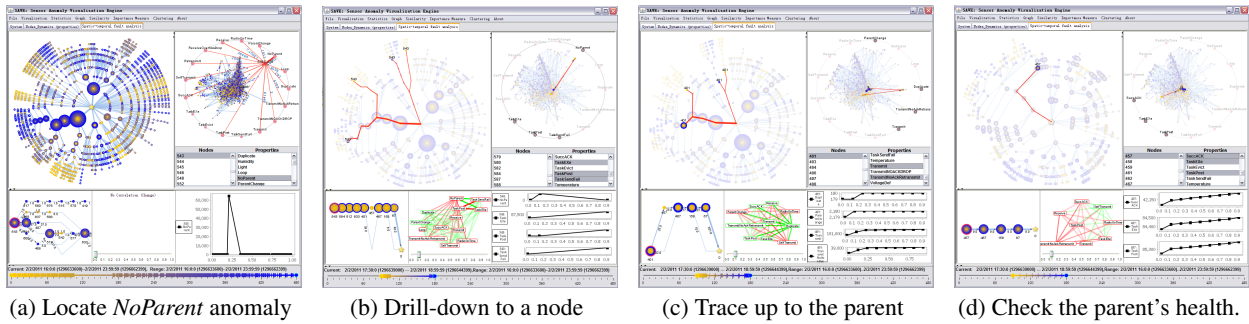| (a) Locate *NoParent* anomaly | (b) Drill-down to a node | (c) Trace up to the parent | (d) Check the parent's health. |

Figure 2: Case studies on sensor network link failure diagnosis.

- The visual interface provides more intuitive and interactive display of data. It is also easier to incorporate human domain knowledge, and furthermore, save significant amount of time by working on the machine-generated insights directly, rather than executing the algorithms and putting them together manually.

## 2. SYSTEM

### 2.1 GreenOrbs and Data Collection

GreenOrbs [2] is a long-term large-scale wireless sensor network system in the forest located on the campus of Zhejiang Forestry University. The system realizes all-year-round forest ecological surveillance and supports various forestry applications. The current deployment includes approximately 500 nodes in total. A portion of the sensor nodes have been in continuous operation for over one year. Each sensor node in GreenOrbs collects four categories of data: the first is sensor readings such as temperature, humidity, etc. The second is the routing path to the sink. The third is link status, including the RSSI/LQI/ETX value of all the neighbors. The fourth is a large collection of statistical information on each sensor node, including the cumulative time of radio power on, the cumulative number of packets (received, transmitted, dropped, etc.).

### 2.2 SAVE System Framework

SAVE system is designed as a three-stage pipeline: data pre-processing, anomaly detection and multi-view visualization. In GreenOrbs case, the data collected are streamed to a central database in real time. In the first stage, the multi-facet raw data is preprocessed, including cleansing, structuring and normalization. Next, anomalies are detected online by the analytics components of SAVE. Results such as outliers and correlation changes are computed and cached. Then as a core stage, all these data facets and the prepared analytics results are brought together and consumed by the visualization components integrated with SAVE.

### 2.3 Anomaly Detection and Visualization

By nature, the sensor routing topology is a large-scale time-varying graph, analytics-unfriendly due to the extremely random routing changes over time. Such a graph is quite messy drawn in normal methods either with geographical or logical layouts. We propose the Temporal Expansion Model (TEM) graph in this work to prepare a more intuitive graph for the following visualization stage. TEM leverages the key feature of the sensor network studied here – all the sensor nodes only send packets to the central sink node for information fusion. The basic idea is to split one physical sensor node into multiple logical nodes according to the separate routing paths to the sink. A snapshot of the TEM graph is shown in the top-left part of Figure 1. To visualize this packet sending pattern, a temporal ring is composed for each node. The color of each ring is selected by interpolating between two boundary colors using the normalized value, i.e., orange to indicate the earliest time and blue to indicate the latest time. The advantages of TEM are two-fold: first, the graphs generated are directed trees, much better for visualization and navigation; second, temporal changes to the network are surfaced to the graph, providing input for further analytics.

Another temporal dynamic occurs at the various dimensions of sensor data. One question is to ask how one dimension changes in relation to another. For example, should the number of packet in transmission increase in the same proportion of sensor voltage decrease? We introduce the correlation graph to address such questions to the data. Two time-series data sets of property value vectors (i.e., one for sensor readings and one for sensor counters) are extracted from the raw reported sensor data in real time based on the selection of the node, start and end timestamps in the system. SAVE computes the correlation scores according to the Pearson's product-moment coefficient and construct a graph representing the correlations among dimensions, An example is given in the bottom-center of Figure 1.

The data reported from sensor nodes is also high-dimensional (up to 30 dimensions). Since the wireless sensors deployed in wild environment are quite dynamic, how to detect and analyze the anomalies in high-dimensional space becomes a challenging task. Specifically, we want to know both the spatial and temporal anomalies of high-dimensional sensor nodes. To achieve such goal, we introduce a dynamic projection graph to represent the distribution of multi-dimensional data in 2D space, as shown in the top-right of Figure 1. Each high-dimensional sensor node measurement in a given time is mapped as a data point onto the space using a concept similar to Star Coordinates. Figure 2 shows a series of interactive visual exploration steps guided by built-in spatiotemporal anomaly analytic algorithms. With SAVE, the underlying causes for these hard-to-detect silent failures in the complex yet dynamic sensor networks may be efficiently identified.

## 3. REFERENCES

[1] K. Ni, N. Ramanathan, M. N. H. Chehade, S. Nair, S. Z. E. Kohler, G. Pottie, M. Hansen, and M. Srivastava, "Sensor network data fault types," *ACM Transactions on Sensor Networks (TOSN)*, vol. 5, pp. 25:1–29, May 2009.

[2] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, and X.-Y. Li, "Does wireless sensor network scale? A measurement study on GreenOrbs," *IEEE INFOCOM*, April 10-15 2011.