論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A study on abnormal-node detection and root cause analysis for dependable IoT sensor networks
著者(和文)	ベルジャブ ニスリーン
Author(English)	Nesrine Berjab
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12298号, 授与年月日:2022年12月31日, 学位の種別:課程博士, 審査員:横田 治夫,宮崎 純,DEFAGO XAVIER,渡部 卓雄,吉瀬 謙二
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12298号, Conferred date:2022/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報工学 系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 (学術) Doctor of
学生氏名： Student's Name	Nesrine Berjab	指導教員（主）： Academic Supervisor(main)	横田 治夫
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 800 語程度）

Thesis Summary (approx.800 English Words)

The Internet of things (IoT) is a distributed, networked system composed of many embedded sensor devices. Unfortunately, these devices are resource constrained and susceptible to many abnormalities, leading to unreliable IoT systems. Indeed, most collected data streams include missing or abnormal values caused by failure, malfunction, environmental events, or malicious attacks (i.e., FDIAs: false data injection attacks). Therefore, reliable methods for recovering the missing data, detecting the abnormal ones, and analyzing the root cause are critical to improving the reliability of IoT. In this thesis, I present solutions to these challenges to ensure a dependable IoT sensor network.

First, the dissertation introduces FuzHD++, a new distributed method to recover missing sensor data and detect abnormal nodes jointly rather than independently. Both elements, data recovery, and abnormal node detection rely on the observed temporal correlation of sensor data to effectively achieve reliable recovery estimation and detection performance. In the data recovery process, the system adopts a matrix profile to extract the top-k repeated patterns from different sensor nodes. Furthermore, it utilizes the k-nearest neighbor estimator to recover the missing data based on the extracted pattern information of multiple neighbor nodes. During the abnormal node detection process, the system adopts a refined fuzzy rule-based detection method. The refined fuzzy rule-based inference system integrates the expert rules and the rules obtained from sensor data analysis to treat ambiguity in the decision-making process. I validated the performance of FuzHD++ by comparing it with existing methods using two real-world datasets. The results show that the proposed missing sensor data recovery method TkRP improved the root mean square error by more than 0.25 compared to the most accurate existing methods. Furthermore, FuzHD++ achieves an accuracy of at least 0.89 and 0.92 with Intel Lab data and Yokota Lab data, respectively.

Then, the dissertation proposes Multi-criteria Xcorr, a new centralized cross-correlation anomaly detection method based on spatial correlation. It conducts a more in-depth analysis and extracts the sensor relationships. The cross-correlation is extracted in both space and time by conducting shape-based logical sub-clustering and two-phase analysis methods. In the first analysis phase, the system uses a variable-size sliding window and a median absolute deviation (MAD) measure. In the second analysis phase, the system performs both tumbling-window and sliding-window analyses to extract multicriteria cross-correlation measures. Finally, all the extracted sensor time-series features will be fed to the shape-based clustering to generate a sensor similarity-like graph. The latter reflects the similarity degree of the sensor with the other nodes. The nodes with a low similarity degree below the threshold will be identified as abnormal nodes. The experiments demonstrate that the proposed method achieves an accuracy of at least 0.90 and 0.93 with Intel Lab data and Yokota Lab data, respectively.

The dissertation also proposes T-LHDM, a sensor state trust model based on a Markov chain, to represent sensor state transitions in normal operating conditions. The correlations between the temporal data and the physical distance between the sensor nodes are taken into consideration to assess the sensor's trustworthiness. To demonstrate the model, I first observed sensed data in a hierarchical decision-making process by intentionally exploiting the sensor data's temporal and spatial correlations. Then, the consistency between the precomputed set of sensor state transitions and the observed ones was used to detect corruptions and assess the sensor's trustworthiness. The experiments using two real-world datasets demonstrated that the proposed model achieves an accuracy of at least 0.95 with Intel Lab data. Furthermore, I conducted a comparative evaluation against a contemporary temporal correlation-based method and three well-known correlation tests. The results showed that T-LHDM outperformed all the competing methods.

Finally, the dissertation also proposes A2S2, an Anomaly Analysis on Semantic Sensor ontology, designed to model and describe the collected sensor data and its associated multivariate-attributes heterogeneous information. The objective is to integrate the heterogeneous sensor information and background knowledge as an interlinked RDF Knowledge Graph. The idea behind capturing both is that analyzing the heterogeneous sensor data and its associated information at a particular period and location will lead to insights into understanding the underlying root cause of an observed anomaly. The approach is solely based on semantics techniques such as SPARQL queries. Starting from the extended SOSA/SSN description, I first derive the physical process model from capturing the correlations among the heterogeneous sensors. This allows me to derive diagnosis rules and, finally, to obtain the identification results. The experimental results show that the identification accuracy exceeds 0.94 in the fire scenario, 0.83 in the FDIAs scenario, and 0.79 in the malfunction errors scenario.

The works in this dissertation contribute directly to designing a dependable IoT sensor network that supports reliable methods for recovering missing data, detecting anomalies, and analyzing the root cause. Specifically, the proposed methods are applicable to handle different anomalies, including FDIAs, which have been neglected in previous studies.