Anomaly Stream Detection Based on Parallel and Distributed Computing

Bakhtiar Amen, Grigoris Antoniou, Violeta Holmes and Ilias Tachmazidis
Department of Computer Science and Engineering, University of Huddersfield, UK
Email: bakhtiar.amen@hud.ac.uk

Abstract— The size of data is growing rapidly at a speed of millions per second, whereas, only the Internet of Things data is estimated to have 50 billion sensors connected around the world by 2020. Therefore, integrating, processing, and mining large data volume requires an effective and efficient framework as well as algorithm to extract knowledge or predict an accurate result. Predicting anomaly detection at a high speed of a data stream is a critical and difficult task, due to the dynamically evolving nature of data streams. Most of the current and traditional anomaly detection methods heavily depend on stationary data which it takes hours or even days for the centralised algorithm to compute and detect accurate results. To address these issues, this paper proposes a different methodology: a new distributed Internet of Things anomaly detection framework and an implemented parallel regression algorithm for concept drift adaption and for detecting data stream changes. A primarily advantage of our approach and results from real world sensors data stream demonstrates that distributed anomaly detection is more effective to detect unusual behavior of large data streams.

Research Contribution

Our contributions are summarized in the following four aspects:

1. First, in this paper, we designed a unique, scalable Big IoT distributed anomaly detection architecture for real-time integration, process and analysis large scale of IoT stream as shows in Fig. 2.
2. Second, we have designed new distributed synchronization channel between apache Storm and Zookeeper to handle network failures. Network failure is a major issue in P2P networking [18] [19]. Therefore, data stream is infinite and arrives at high speed, it is necessary to design a dynamic synchronizing channel to guarantee latency as well as replace failure nodes within cluster nodes.
3. Third, in this paper, we introduced new anomaly detection method called contextual IoT anomaly detection with our theoretic study through a real world experiment, while most of existing IoT detection either considering point or collective anomaly regardless of the data context. Additionally, we grouped IoT sensors based on their context of the data sensors. This form of anomaly examples, definitions, and parameters are described in section III.
4. Fourth, the experiment of real IoT sensor results highlighted the efficiency and low latency of our new distributed three module of architecture as (sensors layer, processing layer, and detection layer), this technique can be implemented in other real-time detection application domains like (e.g., road traffic monitoring, credit card fraud, network monitoring). In addition, our method has the ability to recover both concept drift and changes in the data stream.