International Journal of Mobile Computing and Multimedia Communications
Volume 12 • Issue 2 • April-June 2021

A Sensor Data Stream Collection Scheme Considering Phase Differences for Load Balancing

Tomoya Kawakami, Graduate School of Engineering, University of Fukui, Japan

Tomoki Yoshihisa, Cybermedia Center, Osaka University, Japan

Yuuichi Teranishi, National Institute of Information and Communications Technology, Japan

ABSTRACT

In the internet of things (IoT), various devices (things) including sensors generate data and publish them via the internet. The authors define continuous sensor data with difference cycles as a sensor data stream and have proposed methods to collect distributed sensor data streams. In this paper, the authors describe a skip graph-based collection scheme for sensor data streams considering phase differences. In the proposed scheme considering phase differences, the collection time is balanced within each collection cycle by the phase differences, and the probability of load concentration to the specific time or node is decreased. The simulation results show that the proposed scheme can equalize the loads of nodes even if the distribution of collection cycles is not uniform.

KEYWORDS

Collection Cycle, DAT, Data Aggregation, Distributed Aggregation Tree, Load Distribution, Phase Shifting, Skip Graphs, Structured Overlay Network, TBPS, Topic-Based Pub/Sub

INTRODUCTION

The Internet of Things (IoT) (Hodges et al., 2013) has attracted greater interest and attention with the spread of network-connected small devices such as sensors, smartphones, and wearable devices. In the data science field, stream data generated from IoT devices are analyzed to get various information. A larger amount of data can lead to higher-quality information such a faster stream data collection is one of the main techniques in the data science field and various schemes have been proposed. To enable IoT applications for data collection, pub/sub messaging (Eugster et al., 2003) is considered to be a promising event delivery method that can achieve the asynchronous dissemination and collection of information in real-time in a loosely-coupled environment. For example, the sensor devices correspond to publishers, and the IoT application corresponds to a subscriber. Topic-Based Pub/Sub (TBPS) protocols are already widely utilized by many IoT applications (Teranishi et al., 2015; Teranishi et al., 2017). These systems have a broker server for managing topics. The broker gathers all the published messages and forwards them to the corresponding subscribers.

DOI: 10.4018/IJMCMC.20210401.oa1

In IoT and Big Data applications, collecting all of the raw (unfiltered) sensor data is important for conducting various forms of analysis (Bessis et al., 2014). In this case, the larger the number of sensors treated on the application for analysis, the larger the number of messages that need to be received per time unit on the broker and subscribers in TBPS. For example, when the publishers correspond to a certain kind of sensor which publishes sensor data every 10 seconds and the number of target sensors in an application is 10,000, the broker must receive 1,000 messages per second on average. Thus, the number of messages tends to explode on the broker and the subscribers in IoT and Big Data applications. In general, the number of sent/received messages per unit of time affects the network process load because tasks such as adding/removing headers and serializing/deserializing payloads are required for each message. Therefore, even though the size of each sensor data is small, the increase in the number of publishers can cause network process overloads on the broker and subscribers. This leads to the loss of data or unusual increases in delivery latency, problems which have harm on IoT and Big Data applications.

Many existing studies tackle the problem of scalability in TBPS systems. The approach of these studies is based on distributed brokers, in which brokers are run as peers in a peer-to-peer system. The brokers construct an overlay network among themselves. For example, there are distributed hash table (DHT)-based approaches (Castro et al., 2002; Ratnasamy et al., 2001), hybrid overlay approaches (Rahimian et al., 2011), and Skip Graph-based (Aspnes et al., 2007; Shao et al., 2015; Banno & Fujio et al., 2015; Banno et al., 2020) approaches (Banno & Takeuchi et al., 2015; Teranishi et al., 2015). These approaches can keep the number of connections that each broker needs to accept small by multi-hop message forwarding on overlays. However, these existing methods aim to deliver messages from one publisher to multiple subscribers in a scalable manner. Thus, they are unable to avoid network process overloads caused by the collection, such as when messages are received from a large number of publishers. In addition, the existing techniques do not assume the different intervals at the same time to periodically collect data from the publishers.

Therefore, we define continuous sensor data with different intervals (cycles) as a sensor data stream and have proposed collection methods for distributed sensor data streams as a topic-based pub/sub (TBPS) system (Teranishi et al., 2017; Kawakami et al., 2017). Especially in Teranishi et al., 2017, we have proposed a message forwarding scheme on overlays called “Collective Store and Forwarding.” The scheme can reduce loads of network processes dramatically when there are a large number of publishers (nodes), maintaining the delivery time constraints given for the messages. In addition, we have also proposed a flexible tree construction method called “Adaptive Data Collection Tree” on Chord# (Schütt et al., 2008) that can adjust the maximum load of network processes on distributed brokers to avoid overloads caused by Collective Store and Forwarding. Moreover, we have proposed an expanded method assigning phase differences to balance the collection time among the same or specific collection cycle nodes (Kawakami et al., 2018; Kawakami et al., 2019). We call this novel approach “phase shifting (PS).” The PS approach enables the SG method to decrease the probability of load concentration to the specific time or node. Assigning phase differences at random to the nodes, the collection times are distributed even if there are the same collection cycle nodes. We have evaluated our proposed methods in simulation. Our experiment results show that our proposed method can reduce loads of nodes and realize highly scalable systems to periodically collect distributed sensor data. The scalability of the data collection systems is significantly important to accommodate a huge number of objects and encourage the growth of the data science field.

In the following, the problems addressed are defined in the second section. The data collection scheme considering phase differences is described in the third section. We describe the discussion and related work in the fourth and fifth sections, respectively. Finally, the conclusion of the chapter is presented in the sixth section.