Information Fusion 36 (2017) 68–79

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

A delay-aware schedule method for distributed information fusion with elastic and inelastic traffic

Bo Shen[a,*], Seungmin Rho[b], Xingshe Zhou[a], Ru Wang[a]

[a] School of Computer Science and Engineering, Northwestern Polytechnical University, PR China
[b] Department of Media Software, Sungkyul University, South Korea

ARTICLE INFO

Article history:
Received 15 July 2016
Revised 28 October 2016
Accepted 10 November 2016
Available online 11 November 2016

Keywords:
Distributed information fusion
Online scheduling
Big data
Real-time
Internet of things

ABSTRACT

Information fusion is an efficient way to detect the specified events and extract useful information, especially in the context of big data. As a large-scale data-gathering system, Internet of Things (IoT) has the traffic with the mixed timing characteristics. The real-time observations with various delay constraints and the non-real-time observations are needed in information fusion. In order to guarantee the performance of Distributed Information Fusion (DIF), the paper focuses on the communication mechanism from the perspective of real-time delivery of sensing data. An online scheduling algorithm and its distributed implementation, named Delay-Guaranteed CSMA, are proposed. Both the timing constraints and the historical transmission statistics of sensors are taking into consideration. The simulation results have shown that the proposed policy achieves good delay-guaranteed satisfaction. The goal of real-time data delivery for distributed information fusion is achieved.

1. Introduction

Due to the rapid advancements of mobile computing, wireless communication and Internet, the past few years have been witnessed an unprecedented proliferation of data space. A plenty of applications involve the big data analysis and the data-intensive computing. By gathering huge-amount of information, i.e. so-called big data, various monitoring purposes can be achieved. One key source that is responsible for big data is Internet of Things (IoT). IoT paradigm interconnects many objects in the physical world [1–3]. The fundamental technologies of IoT include big data [4], signal processing [5], Near Field Communication (NFC), Radio Frequency IDentification (RFID), wireless communication [6,7], information diffusion [8], etc. IoT is attracting extensive attentions from both academia and industry since its birth, and has become more prevalent [9].

IoT connects the physical space and the digital space [3]. The interaction of IoT involves the discrete dynamics of digital domain and the continuous dynamics of physical domain. Various functional sensors, mobile devices and cameras are deployed spatially to detect the targets within its own region of observation. As the amount of generated data is very large, data fusion or information fusion is used as an efficient method to gather and obtain the improved data [10–12]. Performing information fusion has several benefits. Useful information is extracted while redundant data and incorrect data is eliminated. The sensing capacity of single sensor is limited by its spatial and temporal coverage. Information fusion can extend the data-gathering coverage in both spatial and temporal dimensions. Decisions can be made based on more available data. Data integrity is also enhanced during information fusion.

One main goal of IoT is to monitor the specified scenarios. These scenarios include car traffic in the transportation systems, production process in the manufacturing, energy flows in the smart grid, public order in the emergency management, medical parameters in the healthcare, etc. [2,13]. These cases require real-time monitoring and fast decision-making. Accordingly, distributed information fusion is a necessary step to accomplish the application objectives.

In the scenario where application objectives and optimization problem of information fusion cannot be distinguished, communication systems play an important role [14]. They provide the ability to acquire the interested data over large-scale areas. However, the communication performance may affect the fusion results, especially in distributed information fusion. The stochastic nature of communication imposes uncertainties on the process of gathering data. The fusion becomes more challenge due to non-deterministic delay and traffic loss during the data delivery. From the point of application level, communication induces additional complexity in remote monitoring. The timing properties of communication is related to the Quality of Service(QoS) in most content delivery net-

* Corresponding author.
E-mail address: shenbo@outlook.com (B. Shen).

http://dx.doi.org/10.1016/j.inffus.2016.11.008
1566-2535/

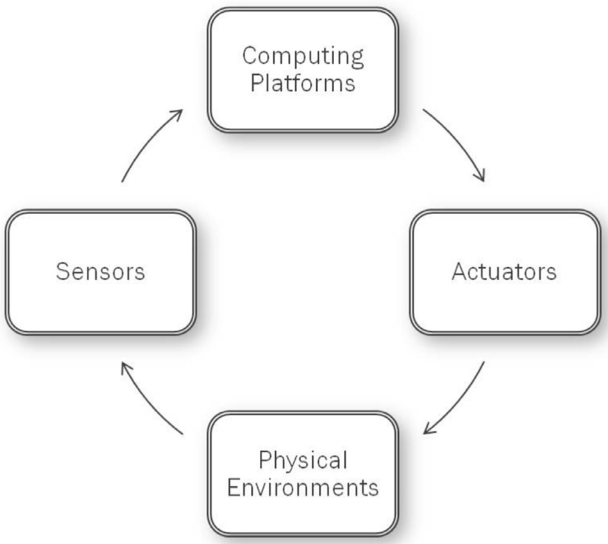

Fig. 1. The data flows in IoT.

works. The delay of delivering traffic will not affect or change the results of applications. Meanwhile, some IoT applications focus on the real-time monitoring on physical objects in the dynamic environments. Temporal correctness is quite important for fusion result. In the distributed environment, the correct information fusion also depends on the satisfaction with many timing constraints imposed on the data delivery. To gather the data and perform the information fusion in a real-time and reliable manner is vital to the monitoring performance.

The paper tries to deal with the delivery of sensing data for distributed information fusion. In IoT different functional sensors are spatially scattered. The interested data is sampled by the sensors and delivered to the fusion center. Then the actuating demands are generated and assigned to the actuators. The data flows are described in Fig. 1. The amount of sensing data is enormous in IoT. A very large number of data is transmitted towards the fusion center. As the broadcasting nature of wireless communication will cause interference to adjacent fusion links, scheduling algorithms are needed. Meanwhile, sensors generate two kinds of data. One kind is characterized by real-time and has hard delay constraints. This kind of data is usually generated because of the occurrence of a specified event. Sensors must deliver the data to the fusion center successfully before its deadline. This data is inelastic traffic. The other kind is non-real-time traffic. It is generated by regular detecting tasks. It has no timing constraints. The case that sensor executes a periodic detecting-task and reports the result to the fusion center generates this kind of traffic. It is elastic traffic. To allocate the communication resource under the heterogeneous timing constraints in a distributed manner is very challenging [15]. This issue is not well addressed currently. Most of the related researches focus on information fusion itself. However, information fusion can be done only when required information is successfully received by the fusion center. Therefore, attentions should paid to data delivery involving in information fusion, especially when timing constraints are imposed by high-level applications.

The paper consider the scheduling problem from the MAC layer. The scheduling problem among multiple sensors is similar to the scheduling issue in a single-hop input-queued switch. They both deal with M inputs and N outputs. Some exclusive constraints exist. The problem in input-queued switch has some encouraged results. It is demonstrated that based on good algorithms, the switch can achieve almost 100% throughout. However, distributed information fusion aims to deliver the sensing data and detect the event quickly rather than achieve high throughput. A significant difference between the two kinds of scheduling problems is that the issue in distributed information fusion is characterized by real-time nature [16]. Instead of the timing analysis and guarantee, previously published studies dealing with switch scheduling focus on achieving high throughput. The groundbreaking work introduces the queue-backlog-based strategy, which has been shown to be throughput-optimal [17]. However, the algorithm is too complicated. It is hard to implement even in the centralized manner. Meanwhile, the aim of distributed information fusion is to detect the specified events in physical environments, especially the aperiodic events. Sensors have no prior knowledge of the generated data, such as the inter-arrival time and the delay constraint. An online scheduling algorithm, only having the information of the current link status and timing constraints, is needed.

In order to provide real-time communication mechanism to support distributed information fusion, an online scheduling policy for traffic with elastic and inelastic features is proposed. A distributed implementation, named delay-guaranteed CSMA, is presented. A metric called urgent degree is defined, which includes the timing features of traffic and the historical transmission statistics of sensors. The detected decision for real-time information fusion is discard if it is not successfully received before reaching its maximum-allowable delay. In order to guarantee the performance of non-real-time information fusion, virtual delay constraint is defined for the detected decision to avoid a long waiting time. However, the traffic will not be discarded when the waiting time reaches to upper bound of the virtual delay.

The policy provides per-packet delay guarantee for real-time information fusion while avoids the problem of locking the scheduling set. The formation of scheduling set in the policy is not Markovian. Central to the policy is that the data for real-time fusion has a higher access priority. The fusion link with a lower served rate in past period has a higher priority among the fusion links with the same timing constraints. Meanwhile, the waiting time of the data for non-real-time fusion is also considered. A tradeoff is made between the timing parameters and the historical statistics of each fusion link. Therefore, the algorithm avoids link starvation problem. Simulation results have shown that the proposed policy can achieve better delay-guaranteed satisfaction. Therefore, real-time delivery of sensing data for distributed information fusion is guaranteed.

The remainder of this paper is organized as follows. Section 2 reviews the state of the art of related works. Section 3 describes the information fusion process, system model and notations used in the paper. Based on these, the scheduling algorithm for distributed information fusion is presented in Section 4. Section 5 evaluates the performance of the proposed model. Section 6 concludes the paper.

2. Related work

The work in [14] reviews information fusion techniques for WSN. A later work surveys the state-of-the-art of distributed data aggregation algorithms in information fusion [18]. Most works in communication issue in the context of information fusion focus on routing topology or communication pattern. Meanwhile, Varshney et al. focus on the fusion of decisions over fading channel in wireless sensor network [19,20].

Resource allocation in wireless network is a hot research field [21,22]. As the Maximum Weight Scheduling (MWS) is hard to implement, many sub-optimal algorithms are proposed, even if they only achieve a fraction of the maximum throughput region. The Longest-Queue-First (LQF) algorithm is an alternative with low