MARSA: A Marketplace for Realtime Human Sensing Data

TIEN-DUNG CAO, Tan Tao University
TRAN-VU PHAM, HCMC University of Technology
QUANG-HIEU VU, Khalifa University
HONG-LINH TRUONG, DUC-HUNG LE, and SCHAHRAM DUSTDAR, TU Wien

This article introduces a dynamic cloud-based marketplace of near-realtime human sensing data (MARSA) for different stakeholders to sell and buy near-realtime data. MARSA is designed for environments where information technology (IT) infrastructures are not well developed but the need to gather and sell near-realtime data is great. To this end, we present techniques for selecting data types and managing data contracts based on different cost models, quality of data, and data rights. We design our MARSA platform by leveraging different data transferring solutions to enable an open and scalable communication mechanism between sellers (data providers) and buyers (data consumers). To evaluate MARSA, we carry out several experiments with the near-realtime transportation data provided by people in Ho Chi Minh City, Vietnam, and simulated scenarios in multicloud environments.

Categories and Subject Descriptors: K.4.4 Applied computing [**Electronic Commerce**]: Data Interchange

General Terms: Data Marketplace, Data as a Service, Realtime Data, Human Sensing Data

Additional Key Words and Phrases: Internet of Things, platform, data contract, cost model

ACM Reference Format:
Tien-Dung Cao, Tran-Vu Pham, Quang-Hieu Vu, Hong-Linh Truong, Duc-Hung Le, and Schahram Dustdar. 2016. MARSA: A marketplace for realtime human sensing data. ACM Trans. Internet Technol. 16, 3, Article 16 (May 2016), 21 pages.
DOI: http://dx.doi.org/10.1145/2883611

1. INTRODUCTION

With the advances in Internet of Things (IoT), realtime sensing data is becoming increasingly important to many realtime applications, such as smarter city [Jin et al. 2014] and city traffic management [Chuang et al. 2013; Panichpapiboon 2010]. Realtime data might come from specifically configured sensors in well-designed infrastructures for specialized applications [Jin et al. 2014] or collected from the mass of human participation, such as GPS signals from mobile devices. This type of human sensing data is readily available in the public and also crucial for solving many real-life problems. However, it is challenging to gather or share human sensing data owned by people who have different habits, knowledge/perception, income, benefits from society, social responsibilities, and so forth. Apart from technical solutions, such as data

Authors' addresses: T.-D. Cao, School of Engineering, Tan Tao University, E-city Tan Duc, Duc Hoa District, Long An Province, Vietnam; email: dung.cao@ttu.edu.vn; T.-V. Pham, Faculty of Computer Science and Engineering, HCMC University of Technology, 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam; email: t.v.pham@cse.hcmut.edu.vn; Q.-H. Vu, EBTIC, Khalifa University, Abu Dhabi, UAE, POBox 127788; email: quang.vu@kustar.ac.ae; H.-L. Truong, D.-H. Le, and S. Dustdar, Distributed Systems Group, TU Wien, Argentinierstrasse 8/184-1, A-1040 Vienna, Austria; emails: {truong, d.le, dustdar}@dsg.tuwien.ac.at.

DOI: http://dx.doi.org/10.1145/2883611

storage and delivery, the platform that enables such gathering and sharing activities needs to have mechanisms to motivate the data owners. Previous studies showed that benefits could be the incentives for data owners to contribute their sensing data [Lee and Hoh 2010; Yang et al. 2012]. Therefore, marketplaces in which data owners are allowed to trade their sensing data for benefits are believed to be the right solution to address the challenge [Lee and Hoh 2010; Yang et al. 2012].

Several platforms have been introduced for sharing and trading different types of data, but none of them is suitable to motivate owners of realtime human sensing data. Examples are Factual [2016], Amazon Data Sets [Amazon 2016], Gnip [2016], Azure Marketplace [Microsoft 2016], and Xignite [2016]. These platforms either exist in the form of data marketplaces or data exchange services. The existing marketplace platforms are commonly designed for trading discrete data packages, such as Web-based information [Möller and Dodds 2012], graph data for structuring human knowledge [Bollacker et al. 2008], or collected data from the physical world [De et al. 2012]. These available marketplaces, however, do not support realtime data [Munjin and Morin 2012]. On the other hand, although data exchange service platforms (e.g., COMPOSE [2016], EveryAware [2016], Ubicon [2016]), and IoT platforms (e.g., Etherios [2016], ThingSpeak [2016], and Xively [2016]) allow the sharing of realtime or near-realtime data, they do not have functions of a marketplace, such as features for selling, marketing, buying, and paying for data.

Our work in this article is motivated by the lack of a platform that could both enable the sharing and provide the incentives for owners to share realtime human sensing data. We present a new marketplace design that addresses the two challenges mentioned earlier: (1) the ability to handle realtime human sensing data and (2) offering mechanisms to motivate data owners of different backgrounds to actively contribute their data to the community. In addition, the new marketplace architecture is designed to interact with existing IoT platforms. To develop such a marketplace, we first study closely a real-life scenario in which realtime human sensing data (i.e., GPS data from mobile devices) is needed to address a practical traffic problem. A set of requirements derived from this motivating scenario is incorporated into the design of the marketplace architecture. A proof of concept system (i.e., a prototype) is built to demonstrate and evaluate capabilities of our novel marketplace.

The remainder of this article is organized as follows. Section 2 presents a working scenario of our marketplace. Section 3 analyzes requirements, proposes the architecture, and discusses several components in detail. The prototype and its experiments for a real-life case study are shown in Section 4. Related works are discussed in Section 5. Finally, Section 6 concludes the article and discusses open issues for future work.

2. A WORKING SCENARIO

In this section, we first present an overview of a traffic monitoring system. Then we analyze features of the system and show how our marketplace can serve as a data exchange platform for the system.

2.1. A Traffic Monitoring System

Traffic congestion is a typical problem of many big cities around the world. A traffic monitoring system is a system that continuously computes the state of traffic (e.g., fast, normal, slow, congested) at any location in a city. Once obtained, the traffic states can be utilized by different stakeholders, from police traffic departments to government agencies and the public, in realtime, to design solutions for the congestion problem. Generally, the traffic states can be derived by analyzing live traffic information collected from different types of sources, such as traffic and surveillance cameras; signals from GPS-enabled devices on cars, buses, and mobile phones; and through road pressure sensors [Chuang et al. 2013; Picone et al. 2012; Panichpapiboon 2010]. In this work, we chose to study requirements of a traffic monitoring system for Ho Chi Minh (HCM) City, one of the most crowded cities in Vietnam, as an example. This city currently suffers from severe traffic problems and urgently needs innovative solutions to address these problems. In addition, we chose this city for two other reasons:

—Traffic in HCM City is not only crowded but also mainly composed of motorcycles whose drivers' behavior are hard to predict. When an traffic incident occurs, it will soon become worse if traffic police cannot take prompt action. These distinct characteristics make HCM City very different from other cities in developed countries, and hence solutions that are proven to be effective in cities of developed countries cannot be applied in this city.
—Realtime data is essential to solving traffic problems. However, HCM City currently does not have a systematic method to collect traffic data, even though various sources of realtime data that can be utilized for solving traffic problems, such as GPS signals from buses, taxis, and mobile phones, and videos from surveillance cameras, are available. Therefore, it is reasonable to study closely the traffic system in this city to see how a realtime data marketplace can be employed to improve the collection of traffic data.

2.2. A Marketplace for Traffic Data Exchange

Effective solutions to traffic problems require the integration of many sources of data. However, with the current condition of the economy in Vietnam, it is almost infeasible for the government to buy expensive traffic data collection systems for its major cities. Fortunately, in HCM City, data that can be utilized for solving traffic problems currently exist in different forms and are owned by different parties. For example, the Voice of Vietnam, a broadcasting agency, has many video cameras installed at major intersections to monitor the city traffic, and it broadcasts the information through radio. The traffic police department also maintains a number of surveillance cameras on streets to detect traffic violations. Major taxi, truck, and bus operators in HCM City have GPS-enabled devices installed in their cars, as required by the government to keep track of their fleets. Additionally, mobile users with GPS-enabled smartphones are also valuable sources of GPS data.

Even though many different valuable sources of data are currently available, the use of these data sources are limited within the scope of their designated applications. Thus, the challenge is how to motivate owners of these data sources to share and make them available to those solving traffic issues. Data owners need to have something to compensate for their investments in data collection equipment, and human resources are needed to keep the equipment operating. A data marketplace where owners have the incentive to trade their data for benefits could be an appropriate solution in this case [Lee and Hoh 2010; Yang et al. 2012].

The integration of data sources in the preceding scenario can be carried out via data exchange in our marketplace, as shown in Figure 1, in which users can trade their traffic data for some benefits. In the marketplace, the trading is usually achieved via intermediate values, such as credits or money (commonly referred to as credits in later discussions). In particular, if a data provider has some data to contribute (e.g., a mobile user with live GPS data), the provider can trade the data for some credits with data consumers. These credits can later be used to exchange integrated data or services from others (e.g., processed information from traffic application providers). With the marketplace, data providers will no longer be limited within the scope of designated applications, where data has to be directly exchanged for services, as in the case of existing traffic applications. Instead, users can freely exchange their data with other participants in the market. As a result, owners of data sources will be motivated to contribute their data. The marketplace also encourages third-party data processors

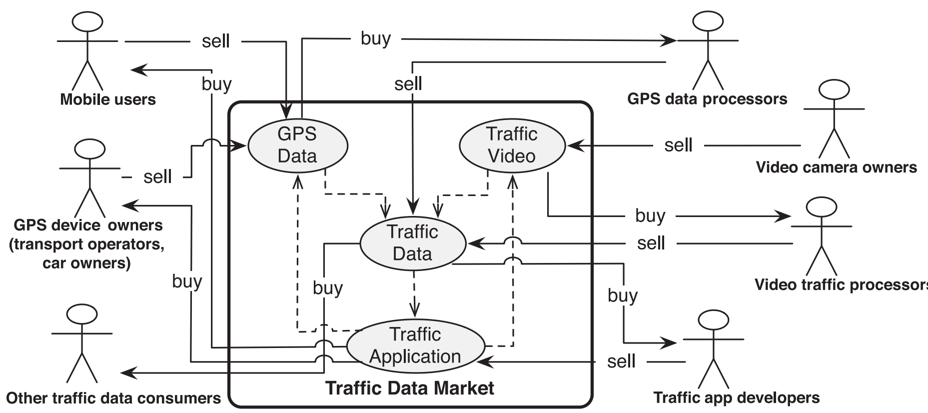


Fig. 1. Data, stakeholders, and their interactions in a market-oriented view of traffic scenarios in HCM City.

Table I. Costs and Benefits of Parties Involved in the Traffic Scenario

Parties	Costs of Collecting Raw Data	Benefits from Processed Traffic Data
Bus, taxi, and truck operators	GPS devices, Internet and mobile network subscription fees, acquiring and maintaining data on servers	Able to track status of their buses, knowledge of current traffic conditions to better provide services to commuters
Private car owners	GPS devices, mobile network subscription fees	Knowledge of current traffic conditions to better navigate cities
Mobile device owners	Mobile devices (e.g., smartphones, tablets), mobile network subscription fees and device battery time	Knowledge of current traffic conditions to better navigate cities
Video camera owners	Video cameras and network connections to video cameras	Selling of video data and traffic information
Data processors	Cost of raw data, infrastructures for collecting and processing raw data	Selling traffic data
Traffic data users	Buying traffic data	Knowledge of current traffic conditions to better navigate cities

(e.g., GPS data processors), who can use their knowledge and tools to buy raw data (e.g., GPS signals), integrate and convert the data into more valuable data (e.g., traffic data), and sell it back to the market for profit.

To summarize, Table I shows benefits from processed traffic data and costs of maintaining the collection of raw data of various data owners. For each owner, the cost and the benefit are not always balanced. If the cost is greater than the benefit, the owner may hesitate to directly exchange raw data for processed traffic data. However, in the market context, by exchanging via intermediate values, the cost and benefit of each data owner can be balanced. This could motivate owners to contribute their data to the market.

3. REALTIME HUMAN SENSING DATA MARKETPLACE

In this section, we first analyze the characteristics of near-realtime human sensing data and the needs from various participants in a marketplace in Section 3.1. Then we present our design for the marketplace in Section 3.2.

3.1. Requirements of Near-Realtime Human Sensing Data Marketplaces

Generally, there are two primary types of participants involved in a data marketplace: data providers and data consumers, who sell and buy the data, respectively. There could be another type of participant who buys data from the marketplace, processes data to add value to it, and resells the processed data to others. In the context of a realtime data marketplace, this type of participant is referred to as an intermediate data processors. Depending on activities that intermediate data processors perform in the marketplace, they can expose themselves as data providers or data consumers.

A data marketplace generally must support buying and selling activities of data providers and data consumers, such as the listing and discovering of data, pricing the data, negotiating and managing the data contract, making payments, and transferring data from providers to consumers. However, near-realtime human sensing data and their owners have some distinct characteristics that a data marketplace designed for near-realtime data needs to support:

—*Near realtime*: The sensing data has its highest value when it is delivered from consumers to buyers immediately. The total time from the creation of data at sources to the point when the data is delivered to consumers must be reasonably small so that it can be processed and timely decisions to react to the current situations can be made.
—*Streaming*: The data is usually delivered continuously from sellers to buyers in forms of streams of events.
—*Heterogeneity of data providers/consumers*: Providers of near-realtime data can be organizations or businesses, such as bus or taxi operators (big/commercial providers). They can also be individuals, such as people with mobile phones, who have limited computing infrastructures (small/personal providers). For example, in the case of the working scenario for HCM City illustrated in Section 2, whereas the data from bus and taxi operators only cover main streets, the data from small providers, such as individuals with mobile phones, have much larger coverage. Therefore, the contribution of small data providers plays a significant role in the success of the system. In addition, because of the heterogeneity of the providers, the volumes and formats of the data streams are also varied. Similarly, data consumers are also diverse.
—*Heterogeneity of data quality*: Because of the heterogeneity of data providers, the quality of data is also varied. Data quality of the same provider may also be varied through time, as it depends on other factors, such as network infrastructure and habits of data owners.

These characteristics have a strong influence in the design of the data marketplace. In particular, the marketplace needs to address the following requirements:

—*Efficient mechanisms for providers to distribute near-realtime data to consumers immediately*: Different from marketplaces for discrete data packages where the data packages from providers can be stored statically on servers for consumers to download, a marketplace for near-realtime sensing data needs to deliver data immediately from providers to consumers. Near-realtime data will lose its value if it is stored for a long time on servers.
—*Data stream transferred from providers to consumers in both push and pull modes*: In the *push* mode, the provider actively makes a data connection to the consumer to push the data through this connection as a stream. The *pull* mode happens in an opposite way, in which the consumer initiates the data transfer by making a connection to the data provider to get the data. For small data providers, such as individuals, the *push* mode is more suitable. However, for big data providers, such as taxi operators, who usually have sufficient computing infrastructure to collect raw GPS data from their fleet and provide the data to the market, the *pull* mode may be more appropriate.

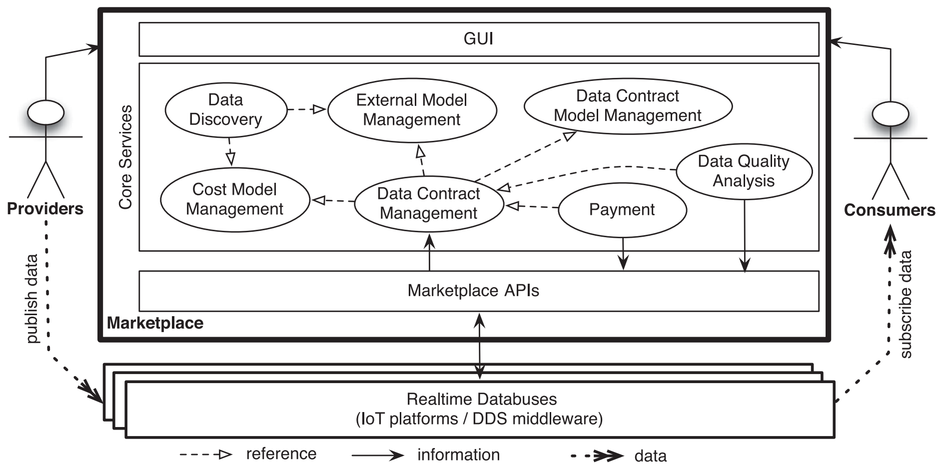


Fig. 2. General design of MARSA.

—*Different mechanisms for listing data, pricing the data, negotiating contracts, and making payments*: Big data providers may make data available through streaming servers and list information about the streaming servers, together with information about the data, prices, and so forth in the marketplace. A consumer, once agreeing to the listed prices, can make a data contract and connect to the server to stream the data. Similarly, big data consumers may also set up servers to receive data, then list the data they want to buy, prices, and so forth on the market. If a data provider wants to sell data with the listed prices, he or she can connect to the server of the data provider to transfer the data.

—*Mechanisms for monitoring and assuring data quality*: As near-realtime data is usually delivered in the form of data streams over a period of time through the Internet, a selling-buying transaction cannot be completed at the time of buying but has to last for a period of time. Different from other kinds of goods, the quality of near-realtime data (the goods) cannot be verified at the time the provider and consumer agree to the deal. The quality of data may also fluctuate over time. Hence, there is the need for a mechanism to monitor the quality of data during the data transfer process.

—*API for interactions with data providers and data consumers*: Near-realtime data is usually not consumed directly by humans. Instead, it is fed to applications on the consumer side for further processing. Therefore, there is also a need for an API or a set of services so that applications of end users can be easily integrated with the data marketplace.

3.2. The MARSA Architecture

Based on the requirements identified in Section 3.1, we have developed MARSA, an architecture for the realtime human sensing data marketplace (Figure 2). The core of the marketplace is a set of services that deliver the main functions of a market, including Data Discovery, Cost Model Management, Payment, Data Contract Management, and Data Quality Analysis. In addition, Data Contract Model Management and External Model Management services are also included to support a variety of contract models due to the homogeneity of data providers and consumers. A graphical user interface (GUI) is provided for user interactions with these core services. A set of APIs are also provided so that external programs can interact with the marketplace. In this reference architecture, to address the need for streaming realtime data from providers to consumers, we utilize available realtime databuses provided by the current IoT platforms or data distribution service (DDS) middleware [OMG 2007; Al-Madani et al. 2013]. By using available databuses, existing applications can easily interact with the marketplace through provided APIs with minimal modifications. Main services and the way in which different components of the architecture interact with each other are discussed in detail in the following sections.

3.2.1. Marketplace Orchestration. For the *pull* model, data providers first use cost model management, data contract model management, and external model management to specify the provided cost of and general information with regard to the data contract, such as its quality and data rights, then publish their data description to the Data Discovery service. Next, they publish their data to databuses, where their data will be delivered immediately to data consumers. Data providers can also access the Data Quality Analysis service and the Payment service to get a report about the quality of data and payment status. On the other hand, through the marketplace, data consumers can use the Data Discovery service to find their required data by submitting a set of requirements (e.g., keywords for the data, expecting price of the data, and data quality). If their required data are found, they can subscribe (i.e., choosing the payment model, quality of the data model, and accepting the data rights) to use one or several data services. At this step, a data contract/agreement based on the matched properties among service description and customer requirements will be generated and managed by the Data Contract Management service. While data transferring process between provider and consumer is executed, through the *APIs*, the Payment service and Data Quality Analysis service work as the intermediates to calculate data value to produce the bill and monitor the quality of data. For the *push* model, the role of providers and consumers are inverted.

3.2.2. Data Discovery. The function of the Data Discovery service is to allow users (i.e., providers and consumers) to publish and search for desired data streams. Metadata is commonly used to enable automatic data/service lookup [Spillner and Schill 2013; Segura et al. 2014]. However, existing data/service lookup approaches consider datasets and data services as discrete and static entities. Therefore, their metadata sets are not appropriate for near-realtime data delivered in streams. Thus, in this work, we leverage DEMODS [Vu et al. 2012] to describe information about data and services for near-realtime human sensing data.

From the analysis of data owners (Table I), we observe that there exist two main levels of data and service description: (1) the *data service* level, in which the general information of a group of data streams of a provider (or an intermediate provider) is specified, and (2) the *data stream* level, in which data information of individual devices is specified. To accommodate these descriptions, we extended DEMODS with (i) adding a *device* field that links to an external model to describe the devices; (ii) adding a *data origin* field to distinguish the raw data of devices and the processing data of reseller; (iii) introducing *time* properties, such as data rate, latency, and time series; (iv) redefining *cost models* and *contract models*, which are discussed more in the next section; (v) replacing the *data* field definition with a *data type* model that supports more data types such package-based data and streaming-based data; and (vi) in both levels (i.e., data service and data stream), the databus is also specified to publish/subscribe data. Figure 3 defines structures for necessary items of two main levels: data service and data stream. Some of items (e.g., categories, quality of service (QoS), data types, or devices) refer to external models, whereas the cost and service contract refer to models defined in Section 3.2.3. Table II describes these items in detail.

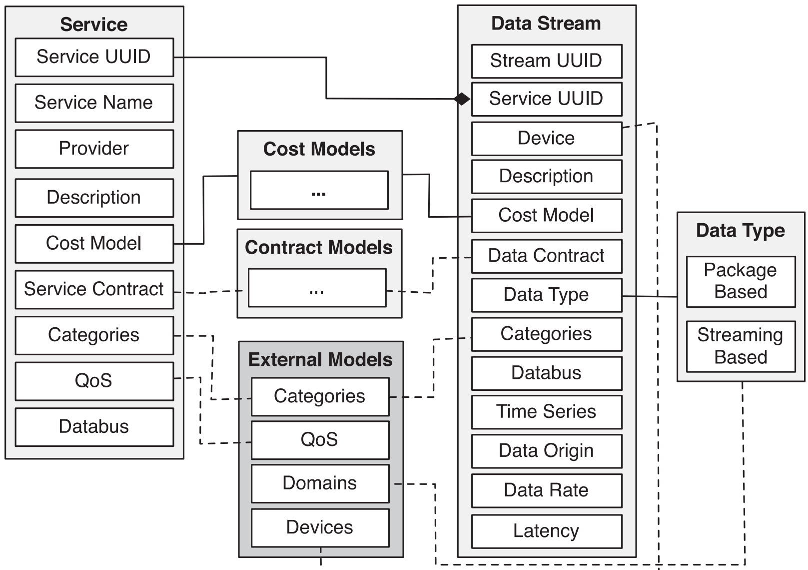


Fig. 3. An overview of general description model for data and services in MARSA.

3.2.3. Data Contract Management. The Data Contract Management service is designed to handle contracts agreed between providers and consumers. Contracts can be in different forms. Each contract usually consists of terms and conditions that govern the quality of data and the associated costs. As shown in Figure 2, the Data Contract Management service refers to Cost Model Management, Data Contract Model Management, and External Model Management services to manage various components of a contract. Section 3.3 discusses the various contract models, quality models and cost models that the data marketplace is designed to support.

3.2.4. Data Quality Analysis. As the quality of realtime streams might vary through time, the Data Quality Analysis service is designed to continuously monitor and analyze the quality of data streams. Reports produced by this service are the basis for verifying data quality specified in contracts.

3.2.5. Payment. The Payment service continuously monitors the data streams in the databuses. Together with the terms and conditions in data contracts on which providers and consumers have previously agreed, the Payment service produces data bills and provides facilities for making transactions between providers and consumers.

3.3. Data Contract and Related Components

A data contract often includes five basic components: data rights, quality of data, regulatory compliance, cost model, and control and relationship [Truong et al. 2012]. The most important component in a data contract for the data marketplace is the cost model, which represents a generic way in which the cost, time, data size, and number of transactions are specified.

3.3.1. Contract Models. The marketplace is designed to support the following four basic contract models:

—*Obligation-free contract*: This type of data contract does not require involving parties to have any obligation to conform to terms and conditions specified in the contract.

Table II. Description Model for MARSA

Level	Properties	Description
Service Description	Service UUID	Service Universal Unique Identifier
	Service name	Data service name
	Description	Detailed data service description
	Provider	Data owner
	Cost model	List of cost models where data owners make offers to their consumers; it links to the cost model defined in Section 3.3.3.
	Contract model	List of contract types where data owners make offers to their consumers; it links to the contract model defined in Section 3.3.1.
	Categories	Linking to a category model defined outside the platform that supports service/stream discovery
	QoS	Description of quality of service; it also links to a QoS model defined outside the platform
	Databus	Place/address where consumers can subscribe to all data of a service
Data Stream	Stream UUID	Stream Universal Unique Identifier
	Service UUID	Data service owner of stream
	Description	Detailed description of stream
	Device	Description of device that generates the data; it links to a device model that supports data service/stream discovery
	Cost model	A stream may offer a different cost model from its data service
	Contract model	A stream may offer a different contract model from its data service
	Data type	Description of the type of data stream; it is classified into object based and nonobject/streaming based and links to a model defined outside the platform
	Categories	A stream may belong to different categories within its data service
	Databus	Place/address where consumers can subscribe to the data of a stream
	Time series	Use if data is object based; the time series represented by a pair $(x, \pm u)$, in which the first one is a positive integer number measured at successive points in time spaced at uniform time intervals (default in second), and the second one is where its uncertainly should be given
	Data origin	Description of from where the data comes (i.e., self-generated, collected, or location of device)
	Data rate	Data rate from the data provider side
	Latency	Maximum time delay from data provider to consumer

For example, contracting parties are not required to follow any restriction on data rights, and they do not need to provide any guarantee on quality of data.

—*User-centric contract*: This type of data contract focuses on requirements of a service that the service provider has to deliver to users. It contains not only elements to determine the quality of data, such as completeness and accuracy, but also elements to specify the support and indemnification of the service provider in cases of failure. For example, a typical user-centric contract for traffic information would have constraints on the latency of information (e.g., less than 5 minutes), the sample rate (e.g., 10 samples per minute), and the accuracy of sample location (e.g., less than 20 meters in geographical distance error).

—*Provider-centric contract*: This type of data contract provides requirements on data rights and regulatory compliance of the data that users have to follow. In particular, it specifies key requirements for using the data (i.e., data rights) and rules on data that are to be obeyed (i.e., regulatory compliance). For example, constraints on the redistribution of traffic information in a provider-centric contract could be no redistribution for commercial purposes and redistribution without modification of the data and with a display of the service provider name for noncommercial purposes.

—*Customizable contract*: This type of contract allows users to modify any of the preceding contract models. On the one hand, a customizable contract can be developed from any contract model by adding constraints to it. It can also be developed by mixing the preceding contract models and even modifying the mixed model after that. On the other hand, it is possible for service providers or users to define their own requirements or constraints to add to the contract model. Since a contract is often expected to include constraints from both parties and both users and providers always have some specific constraints to include, the customizable contract model is the most popular model used in practice.

3.3.2. Quality of Data Models. In a data service, the quality of data has a strong relationship with the cost, because in any market, the cost usually meshes with the quality of data, product, or service. Based on quality of data models that consider the data in many aspects (e.g., accuracy, completeness, and timeliness), data providers can define the appropriate cost for their data service. Actually, linking quality of data to the data cost is important to prevent providers from increasing the data quantity by adding a lot of noise, such as replicated data, out-of-date data, or data already on the protocol layer. Since quality of data models depend on many data properties, we suggest that providers use external models [D'Ambrogio 2006; Dobson et al. 2005] to define and monitor their quality of data.

3.3.3. Cost Models. When a data producer sells data to a data consumer, the data value will be based on cost. Several cost strategies have been summarized by Muschalle et al. [2012] and Schomm et al. [2013]. However, in the context of realtime human sensing data, it is not completely suitable, because at different moments and different locations, a data stream may have different values. To adapt to different business models of different consumers/providers for both the *pull* and *push* models, we support the following payment strategies:

—*Payment on package delivering (API handle)*: Data can be split into separated packages (e.g., messages or images). Consumers are charged every time they successfully receive packages from the marketplace. To describe this cost model, the usage fee of a fixed number of packages has to be included in the API description.
—*Payment on data size*: Consumers are charged on the size of received data. Similar to the previous case, the basic unit charge fee for a data unit (e.g., 1MB, 10MB, 100MB, or 1GB) should be described.
—*Payment on time of subscription*: Providers can split a day or a week into different time units and the corresponding cost for each unit of time (e.g., $1 for 1 hour of a subscription within the hours of 1 pm to 5 pm) is set up. Consumers are then charged on the total time of each unit to which they subscribed. This model is appropriate for streaming data where the data is generated on a duration of time.
—*Payment on data unit*: Providers split their data into different data units and set up the basic unit charge for each. Unlike payment on data size, the data unit here can be split not only by data size but also by time and over a group of data streams. Consumers will pay one time and get the data until reaching the limitation of unit.
—*Payment on plan (fixed payment on a period)*: Consumers subscribe to use data in a subscription period (e.g., a week or a month) and only pay one time for this period with or without maximum limitation of received data. The basic unit charge fee for a unit of time (e.g., 1 hour, 1 day) is also described.
—*Free users*: In a number of cases, consumers can use these services at no charge from providers. There are several reasons to offer a service for free, such as (1) the data comes from the government and the consumer is a public authority funded by tax money, a case which is usually constrained by a data contract, and (2) a person or an organization providing the free data as a social responsibility because the generating data fee is supported by the other organization or the government.

To encourage consumers, for each payment strategy mentioned earlier, we support a mechanism called *Freemium*, in which providers offer a limited access (e.g., the limit of package quantity, the data size, or a duration) at no cost initially. In addition, the provider can define several requirements that allow their consumers to get a discount (e.g., at specific subscription times on the bill or the data size). Finally, for certain types of data or service, the preceding cost models can be expanded to include a free payment option with pop-up advertisements for data consumers (i.e., the data can be provided free of charge in exchange for some advertisements).

3.3.4. Contract Governance. A data contract, once established, is governed via two processes: a data contract monitor and data contract enforcement. A data contract monitor is used to monitor requirements specified in the contract. Our framework provides a basic monitor; however, it is possible to allow a third party to be involved in the monitoring process. For example, the quality of data delivered in a service can be monitored and assessed by a third party, and then the result is reported to our framework. On the other hand, data contract enforcement ensures that the contract is followed strictly by both data providers and data consumers. In cases of violation, different types of penalties can be applied, such as a warning, a temporary suspension of the service, penalty fees, or even termination of the service.

4. IMPLEMENTATION AND EXPERIMENTS

4.1. Marketplace Prototype

For demonstration purposes, a prototype[1] of the reference architecture described in Section 3 has been implemented. In this prototype, we fully implemented two basic services: Data Discovery and Cost Model Management. Whereas the first one was a revision of a previous work [Vu et al. 2012], the second service was a new implementation. A set of APIs have also been built for external databuses, such as Mosquitto[2] and the Xively IoT platform [Xively 2016] to interact with the marketplace via Web services. These APIs are used to enforce the cost model and support latency analytics. Mosquitto and the Xively IoT platform were chosen to illustrate the interactions between the marketplace and external databuses because they implemented a lightweight broker-based publish/subscribe messaging protocol—MQTT.[3] A simple Payment service version was also implemented. This service uses a data log captured by databus via APIs and is cost managed by the Data Contract Management service to produce online bills. To prove the realtime capability of the design, we implemented, as a part of the Data Quality Analysis service, a latency analytics that allows us to measure the near realtime capacity of databuses. An overview of the MARSA prototype is shown in Figure 4. Based on our work, a full version has been also implemented by TMA Research.[4]

Together with core services, cost models used in data contracts discussed in Section 3.3.3 were also implemented. The class diagram in Figure 5 describes the relationship among payment plans used in cost models during the implementation. Each payment plan was implemented as a class, which was a subclass of the cost model. Even though all payment plans inherited a price property from the cost model, the use of this property was slightly different depending on the actual payment plans. For example, with a data size plan, it was used to represent the price of one data unit (e.g., 1Kb, 10Kb, and 1Mb). In the time Plan and subscribe plan, it was the price for one time unit (e.g., 1 hour, 1 day, or 1 week). With the data unit plan, it was the cost of using data streams.

[1]The platform has been released as an open source at http://dungcao.github.io/marsa.
[2]http://mosquitto.org/.
[3]http://public.dhe.ibm.com/software/dw/webservices/ws-mqtt/mqtt-v3r1.html.
[4]http://www.tmaresearch.com.

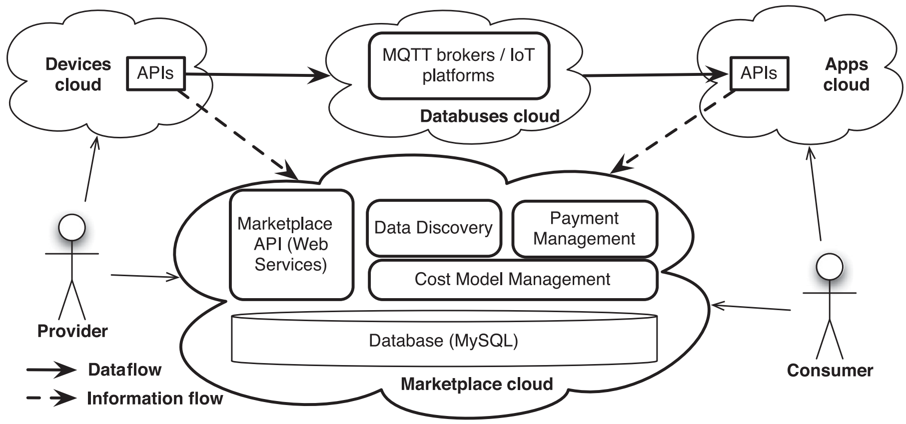


Fig. 4. MARSA prototype.

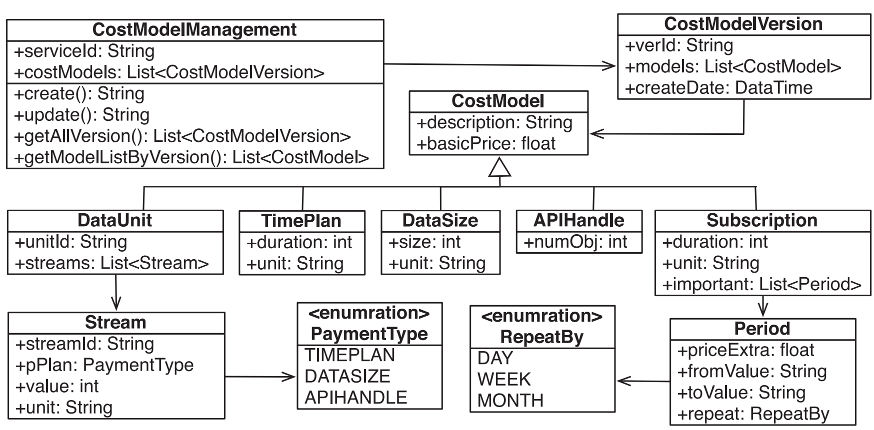


Fig. 5. Class diagram of cost model management.

In the current prototype, once a consumer subscribes to a data service and chooses a particular payment plan, the marketplace enforces the plan by continuously monitoring and analyzing information collected from databuses via the APIs. Depending on the type of payment plan, the necessary information for enforcing the plan may be different. Table III lists a set of provided APIs and the types of information that these APIs can collect for enforcing different payment plans.

4.2. Comparison with Other Platforms

Several data marketplaces exist. In this section, we compare our proposed marketplace with some popular marketplaces to identify the strengths and weaknesses of each marketplace with respect to their support for realtime human sensing data. Features of a realtime data marketplace, including supported data types, data sources, data publishing and delivery method, cost model, automatic data lookup, data contract, and payment management, are used as the main criteria for our comparison. Through this comparison, we want to demonstrate that our solution meets typical requirements of a data marketplace for realtime IoT/human sensing data. The requirements include reusing the IoT platform for data publishing and delivery, providing flexible contract

Table III. List of Marketplace APIs

Methods	Description	Use for Plan
subscribeStart()	Used when consumer starts to subscribe to a stream.	Subscription
subscribeEnd()	Used when consumer unsubscribes from a stream.	Subscription
getTimePlan()	Return either the period if consumer chose the time plan as the payment strategy or null. After calling this function, databus must control the deadline upon the time period it returns.	Time plan/ Data unit
packageCount()	Whenever a package is delivered to consumer, the function is called only one time. In case the current stream touched the limitation (if it is set), this function will return false.	API handle/ Data unit
setDataSize()	Set the transferring data size of objects or data stream within a duration. In case the current stream touched the limitation (if it is set), this function will return false.	Data size/Data unit

Table IV. Comparison of Data Marketplaces

Products	Data Type	Data Source	Publishing/ Delivery	Cost Model	Auto Lookup	Data Contract	Payment
MARSA	Realtime, streaming	IoT devices	MOM,[5] IoT platform	Flexible[6]	Yes	Yes	Online billing
Xignite	Datasets, realtime	Range, finance	Files, API	Asset, delivery	Yes	N/A	N/A
Amazon	Datasets	Range	Files	Free	N/A	N/A	N/A
Azure	Datasets	Range	OData API	Subscription	N/A	Publisher offers terms	N/A
Factual	Datasets	Geography	Files, API	Free/ Subscription	Yes	Terms of service	N/A
Trimble InSphere	Datasets	Geography	Files	Per user/ Device/Data	N/A	License agreement	N/A
Gnip	Realtime, historical	Social network	API	N/A	Yes	N/A	N/A
Sense2Web	Realtime, streaming	IoT devices	MOM, IoT platform	N/A	Yes	N/A	N/A

and cost models for data streams to meet the needs of different data providers and consumers, supporting automatic data lookup services to support different ways of streaming data, and especially the capability to produce online billing in near realtime.

4.3. Application to a Real-Life Case Study

A concrete application, named Traffic Information System[7] (TIS) for HCM City, of the urban traffic systems scenario in Section 2 was analyzed to show the usefulness of the data marketplace. The main functionality of this system is to provide information about the velocity of vehicles on roads in the city in realtime. With this system, users can make a good plan for their travel in the city (e.g., to avoid traffic jams). This system receives raw GPS data (i.e., longitude, latitude, velocity, and timestamp) from GPS devices attached to city buses, taxis, and GPS-enabled mobile phones in realtime. The system then estimates possible travel speeds that vehicles are moving on the roads. The result is displayed on an online city map, in which the possible travel speeds are represented by colored lines. Generally, the system consists of three main

[5]MOM: Message-oriented middleware.
[6]The providers are very flexible to define their data value by data type, data size, or subscription time, and so forth.
[7]http://traffic.hcmut.edu.vn/.

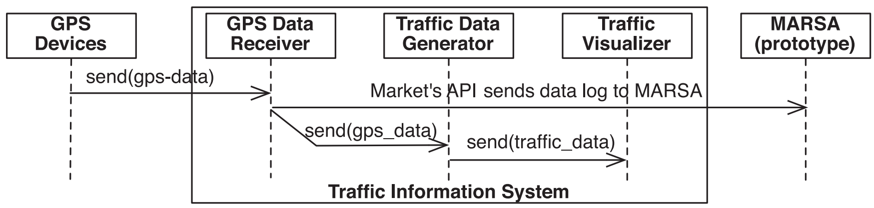


Fig. 6. Interactions between TIS and the data marketplace (i.e., MARSA).

components involved in processing realtime data: *GPS Data Receiver*, *Traffic Data Generator*, and *Traffic Visualizer*. The GPS Data Receiver receives realtime raw GPS data from external sources. The Traffic Data Generator converts raw GPS data into traffic data (i.e., possible traveling speeds of city roads) and provides the traffic data to the Traffic Visualizer. The following description will show how the MARSA prototype can be used to facilitate the exchange of data related to the TIS scenario. The overall interactions are described in Figure 6, in which the GPS Data Receiver is considered a databus of the marketplace. We integrated the marketplace APIs into this component to count the amount of data and send it to the marketplace.

4.3.1. Market Interactions. The TIS plays a dominant role in this interaction scenario, as it is the major consumer of the data. Therefore, the price of GPS data is predefined by the TIS owner using the Cost Model Management service. The information of the databus is also specified by the TIS owner. To sell their raw GPS data, GPS data providers such as bus and taxi operators get into the data marketplace and provide a set of information to the data market as described in Figure 3. After registering with the market, data providers can use the APIs provided by the TIS owner to connect to the GPS Data Receiver to upload their data. It is assumed that the providers have agreed on the terms, conditions, and cost policies set by the TIS owner. The marketplace APIs integrated in the GPS Data Receiver of the TIS will count the amount of data that the system receives from each provider and regularly update the information to the data market. The payment component of the market uses this information for accounting purposes and decides on the cost that the TIS owner has to pay for the data providers. For small data providers such as commuters with mobile devices, a mobile application is built for sending GPS data to the GPS Data Receiver of the TIS. Interactions between the providers and the market, as well as the TIS, such as accounting and billing information, registration, and so forth, are done through this application.

4.3.2. Discussion. The example of the TIS, a practical application of an urban traffic scenario, has demonstrated the use of the dynamic data market for a realistic application. Currently, the TIS receives GPS data from around 4,000 city buses every day. On average, each bus sends about 0.25MB of data to the TIS GPS Data Receiver per day, equivalent to 7.5MB of data per month. For the whole city bus fleet, the total amount of GPS data received per month is about 30GB. We can now use the data free of charge for research purposes. However, if each megabyte of GPS data costs 20 ¢ in the market, the bus operators will receive around $6,000 for the whole bus fleet. This is an accountable amount, and enough for the bus operators to pay for the cost of 3G/GPRS data connections used to send GPS data to data-gathering servers. For mobile users, if each mobile phone sends the same amount of GPS data to the TIS as a bus, the providers will receive approximately $1.50 per month for each GPS-enabled device. Even though this is a relatively small amount, it could be used to pay for half of a 3G data bill for mobiles in the current condition of Vietnam. This illustrated example shows that the data market model can bring some benefits to users. Several studies have shown that benefits can be used as incentives to encourage mobile users to participate in the marketplace and contribute their data [Lee and Hoh 2010; Yang et al. 2012].

Technically, as the marketplace is currently built based on a service-oriented model, it has certain limitations that affect the flexibility of the market. For example, it is hard for small data owners, such as providers with GPS-enabled devices, to sell their data directly in the marketplace. As the realtime data streams must be provided through data services, data providers have to gather their data and then publish the data to some forms of databus so that data consumers can access the data. However, small data owners, such as ordinary mobile users, do not usually have their own computing infrastructure or the necessary computing skills to do so. The specific GPS data collection application developed for mobile devices used in the TIS application is a walk-around solution to help ordinary mobile users easily contribute their data.

4.4. Experiments in Simulated Scenarios

4.4.1. Experimental Scenario and Settings. The main purpose of our experiments is to evaluate the practicability of MARSA according to real-world scenarios.[8] We used Mosquitto, an open source MQTT message broker, as the databus:

—*Infrastructure setting*: As a proof of concept that MARSA works for the scenario in Section 4.3, we deployed one databus in a virtual machine (VM) with 2 CPU AuthenticAMD, 2.80GHz, and 8GB of RAM, running on an Ubuntu Server 12.04.2 LTS 64-bit of our partner cloud (Flexiant[9]). Here, 200 simulated sensors were deployed in four VMs of the same cloud with the databus and 200 simulated consumers were deployed both inside and outside the cloud of the databus—for instance, three servers in Vietnam (Tan Tao University, HCM City University of Technology and Hanoi University of Technology) and our partner clouds around Europe, such as Stratuslab (in France), Distributed Systems Group (DSG lab) at TU Wien (Austria), and in the Flexiant FCO cloud. The simulated sensors published a text file, in which each line was about 100 bytes, to the databus by reading line-by-line every 5 seconds. On the other hand, whenever receiving a message, the simulated consumers saved it into a text file. We also used five personal computers (PCs) located in Vietnam as video streaming sensors. These sensors published video data that was captured by cameras and whose sound was removed. The simulated camera consumers were deployed on other PCs located in the DSG lab. This setting allowed us to verify the near-realtime capability on different environments and the capability of the platform to deal with different data types. Figure 7 depicts our setting across distributed sites.

—*Cost setting*: According to the preceding setting, we have two types of data (i.e., text/message and video streaming). We assume that 200 simulated sensors and five video streaming sensors are owned by a provider, and he or she has set the cost of data as follows. For five video streams, the value of data is evaluated by subscription time (e.g., \$2 /hour/stream). For the text streams, the value of data is evaluated by data size (e.g., \$5 /1GB).

Based on the settings mentioned previously, we analyzed how well the platform measures near-realtime capability and evaluates data values. To support this, the marketplace APIs were integrated into all simulated sensors and simulated consumers to capture a data log and notify the marketplace whenever a data package was sent/received.

[8]In our experiments, MARSA and its APIs (i.e., Web services) were deployed at http://109.231.124.57:8080/marketplace/default and http://109.231.124.57:8080/ws.

[9]http://www.flexiant.com/.

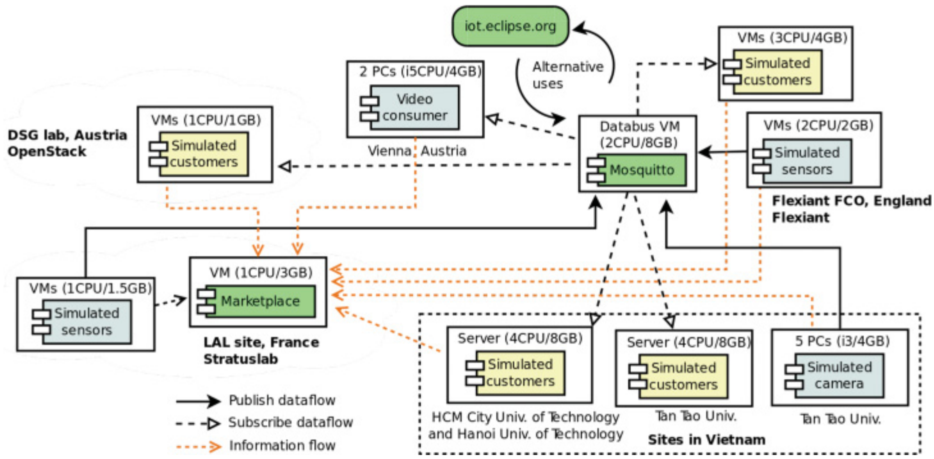
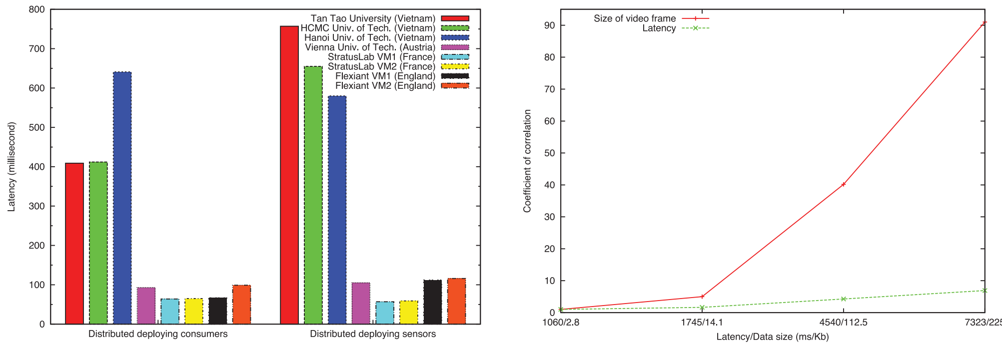


Fig. 7. General resource settings for experiments.

4.4.2. Near-Realtime Capability Measurement. In this experiment, we followed the first setting to test the prototype with two types of realtime data related to the traffic scenario, including structured data (i.e., raw GPS) and unstructured data (i.e., video data). Then we measured the average latency of each type of data caused by the databus for each consumer. The latency was calculated by the difference between two timestamps of a data package recorded by the platform (i.e., MARSA): the time when the source device published notification and the time when the consumer receives notification. This avoided the problem of a time difference between internal clocks, as we had too many source devices and consumers. For raw GPS data, on the left of Figure 8(a), we show the realtime capability of some consumers deploying in different clouds to subscribe for the data published by the simulated sensors deployed in the StratusLab cloud. On the right, we inverted the location of the simulated sensors and the simulated consumers—for instance, 200 simulated consumers were deployed in two VMs of the StratusLab cloud, and their simulated sensors were deployed in the different clouds and the servers. For the video data, we were interested in the coefficient of correlation between the latency and the size of the data package (Figure 8(b)) when we changed the size of video frames. These results show that our marketplace is acceptable for realtime applications.

4.4.3. Data Value Evaluation. In this section, we show how the platform evaluates the value of data in near realtime. There are many consumers who bought the 200 text streams and five video streams discussed earlier; however, we assume that there is one consumer who bought 5 text streams and two video streams in a contract. Based on the data log from the marketplace APIs and data cost defined by the provider (i.e., \$5 /1GB for text and \$2 /hour/stream for video streaming), the marketplace produced an online bill whenever the consumer wanted to check his or her used data (Figure 9). Based on this bill, customers could update their plan, subscribe to more data, or unsubscribe from some data streams.

4.4.4. Discussion. Throughout the preceeding experiment of the simulated scenario, the realtime capability and data heterogeneity are supported by our platform, although the realtime capability of video streaming data is unexpected. However, this is a problem of the databus, and we can solve it by using a specific databus for video streaming, such as the one in Al-Madani et al. [2013]. Our platform also has the capacity to measure the value of data in realtime. Even though we do not focus on the



(a) Latency comparison from different locations around the world

(b) Coefficient of correlation between the latency and the size of video frames

Fig. 8. Realtime capability measurement.

Bill No.: 2015/03-5.1

From date: 2015-03-30 12:39:53 To date: 2015-03-30 18:40:57

Status: Not Payment

Payment on DATA_SIZE (5.0 $ / 1 GB)

List of streams

No.	Stream UUID	Size	Price
1	suuid1427702254973/sid1	0.219 GB	$ 1.1
2	suuid1427702254973/sid2	0.0217 GB	$ 0.11
3	suuid1427702254973/sid3	0.0550 GB	$ 0.28
4	suuid1427702254973/sid4	0.181 GB	$ 0.9
5	suuid1427702254973/sid5	0.205 GB	$ 1.02
			Total price: $ 3.41

Payment on SUBSCRIPTION (2.0 $ / 1 HOUR)

List of streams

No.	Stream UUID	Size	Price	Size Extra	Price Extra	Sum Price
1	suuid1427702254973/sid11	3.67 HOUR	$ 7.34	0	$ 0	$ 7.34
2	suuid1427702254973/sid12	6.02 HOUR	$ 12.04	0	$ 0	$ 12.04
						Total Price: $ 19.38

Total price of contract: $ 22.79

Fig. 9. A screenshot of an online bill.

quality of data analysis in this experiment, by integrating our APIs into sensors and applications of the consumer, the APIs can analyze the quality of data whenever they send/receive a data package and then submit the report to the marketplace.

5. RELATED WORK

5.1. The IoT Platforms

Studying the features of existing IoT platforms and comparing them to a set of requirements for the marketplace are important when developing a data marketplace in the context of IoT applications. Misra and Pal [2013] made a survey on important trends, key requirements, evolving technologies, and emerging solutions for such a platform for IoT and M2M services. They also took a look at some commercial and open source

IoT platforms for data services, device management, and application development in the market. Based on this work and a study of existing IoT platforms (e.g., Arrayent [2016], Axeda [2016], Xively [2016], and ThingWorx [2016]), we recognized that most of them have three basic processes: upload/publish data, storage with a time-series database, and download/query data using time points. Moreover, they allow users to analyze or visualize the data. However, except for Xively, all of them do not consider the issue of data delivery to users in realtime. Nimbits [2016], Sentilo [2016], and Kaa [CyberVision 2016] consider the alert settings in the point properties whenever a new value is recorded into a data point.

From the research community, several middleware platforms have been proposed/developed. Using the Web protocol to establish the connectivity of the things into the Internet, Guinard et al. [2010] defined a resource-oriented architecture for the Web of Things in which data is stored in distributed devices. The applications/consumers can directly access these devices using the HTTP protocol to query data. This architecture is appropriate for the private platform, as the number of queries to devices is limited. However, the challenge of publishing data to consumers with high-performance real-time communication is an open issue due to the limitation of the HTTP protocol. The EVRYTHNG [2016] platform has been using this technology. De et al. [2012] describe the Sense2Web platform for real-world services and smart objects. Starting from the IoT information models detailing entities, resources, and services, the platform makes linkages with tag and location analysis of the existing resources on the Web, then publishes them on the Web within two user interfaces: human-to-machine interaction via Web and M2M interaction via SPARQL endpoints. Valente and Martins [2011] developed a middleware framework in which it receives smart objects from wireless sensor networks and transforms them into Web services. This platform supports a notification mechanism to alert clients whenever a new smart object is available. Tong and Ngai [2012] developed a publish/subscribe system that supports ubiquitous data access from both wireless sensors and mobile phones. This system plays the role of databus in our platform.

Discussion. With respect to the problems that need to be solved for a data market mentioned in Section 2, we recognize that all of these platforms only satisfy a part of the technology infrastructure, such as missing the mechanism to process data in realtime, whereas a set of requirements, such as a business model, service/data broker, or QoS/quality of data, is totally missing. Munjin and Morin [2012] reuse these IoT platforms as the data brokerage platforms to define an architecture for IoT application marketplaces. In this work, the application store is suggested as an element of IoT platforms, which was designed as a registry of networked applications. The authors suggest that this registry should support application description management and authentication. This work, however, focuses on application development instead of selling data.

5.2. Data Services

In recent years, the number of platforms for data marketplace has grown rapidly. Typical examples include Amazon Data Sets [Amazon 2016], Factual [2016], Gnip [2016], Azure Marketplace [Microsoft 2016], and Xignite [2016]. Using these platforms, a registered client can upload his or her data manually or automatically using the supported APIs. However, many types of data are not near realtime and become obsolete for some consumers. Some consumers may not be satisfied with this situation because they need fresh new data or near-realtime data. To have the realtime data for consumers, it is required that the data must be collected by devices and simultaneously uploaded to the platform for prompt delivery to consumers. All of these platforms do not address this issue.

In our previous works, by focusing on the automatic service lookup, data composition, and utilization for several DaaS on the Cloud, we defined a description model for DaaSes, named DEMODS [Vu et al. 2012], which covers all basic information of a DaaS, such the description of service, data assets, APIs, and several linked models: pricing, contract, QoS, and so forth. Additionally focusing on data composition, we developed data contracts [Truong et al. 2012] to support concern-aware data selection and utilization from cloud-based data marketplaces. In an empirical study, Muschalle et al. [2012] presented several pricing strategies for data markets. They also presented attractive research opportunities for the business intelligence community. Li and Miklau [2012] focused on pricing of data market; however, they proposed criteria for interactive pricing instead of analysis of pricing strategies. Using *linked data principles*—for instance, all datasets are represented internally as RDF graphs and each item is identified with a URI, Moller and Dodds [2012] developed a platform as a Web-based information marketplace named Kasabi. A similar work, Bollacker et al [2008], defined a platform for structuring human knowledge using a collaboratively created graph database. However, these works do not consider payment of data. In Carey et al. [2012], starting from a general architecture for data services that can be deployed on top of a data store, the authors reviewed data service concepts and examined approaches to service-enabling data sources, create an integrated data service from multiple sources, and manage data in the cloud. Moreover, they highlighted technical challenges, including updates and transactions, data consistency, or security for data services, and they discussed emerging trends for future research and development.

6. CONCLUSION AND FUTURE WORK

In this article, we presented the design of MARSA, a platform for a human sensing realtime data marketplace. MARSA consists of a set of services interacting with one other to cover data discovery, cost model management, and data contract management. The roles of services addressing payment, data quality analysis, and marketplace APIs were also analyzed. In addition, a prototype has been implemented as a proof of concept of the proposed marketplace platform. By using multiple databuses, with different data types and the flexibility of choosing databuses at providers and consumers to transfer data, users can use a specific databus for a concrete data type to improve performance. Our prototype allows data providers and their consumers to negotiate their payment for each type of data according to five supported cost models. It helps them monitor how their choice of models are executed through the marketplace APIs.

In the future, we plan to introduce a dynamic cost model for the platform. All cost models that we introduced in this article are static. In reality, in some cases, it is desirable to have a dynamic one. Our plan toward this issue is to work for a cost model that can change costs according to the supply and demand of the market. We also plan to continue our work on the implementation of missing services, such as those involving data quality analysis and external model management.

REFERENCES

Basem Al-Madani, Mohammed Al-Saeedi, and Anas A. Al-Roubaiey. 2013. Scalable wireless video streaming over real-time publish subscribe protocol (RTPS). In *Proceedings of the 17th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*. 221–230.

Amazon. 2016. Amazon Data Sets. Retrieved March 14, 2016, from http://aws.amazon.com/publicdatasets/.

Arrayent. 2016. IoT Platform Overview. Retrieved March 14, 2016, from http://www.arrayent.com/platform/overview/.

Axeda. 2016. M2M Connectivity. Retrieved March 14, 2016, from http://www.axeda.com/products/m2m-connectivity.

K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*.

Michael J. Carey, Nicola Onose, and Michalis Petropoulos. 2012. Data services. *Communications of the ACM* 55, 6, 86–97.

Yi-Ta Chuang, Chih-Wei Yi, Yin-Chih Lu, and Pei-Chuan Tsai. 2013. iTraffic: A smartphone-based traffic information system. In *Proceedings of the 42nd International Conference on Parallel Processing*. 917–922.

COMPOSE. 2016. COMPOSE Home Page. Retrieved March 14, 2016, from http://www.compose-project.eu/.

CyberVision. 2016. Kaa Home Page. Retrieved March 14, 2016, from http://www.kaaproject.org.

Andrea D'Ambrogio. 2006. A model-driven WSDL extension for describing the QoS of Web services. In *Proceedings of the IEEE International Conference on Web Services (ICWS'06)*. 789–796.

Suparna De, Tarek Elsaleh, Payam Barnaghi, and Stefan Meissner. 2012. An Internet of Things platform for real-world and digital objects. *Scalable Computing: Practice and Experience* 13, 1, 45–57.

Glen Dobson, Russell Lock, and Ian Sommerville. 2005. QoSOnt: A QoS ontology for service-centric systems. In *Proceedings of the EUROMICRO Conference*. 80–87.

Etherios. 2016. Device Cloud. Retrieved March 14, 2016, from http://www.etherios.com/products/devicecloud/.

EveryAware. 2016. EveryAware Home Page. Retrieved March 14, 2016, from http://www.everyaware.eu/.

EVRYTHNG. 2016. EVRYTHNG Home Page. Retrieved March 14, 2016, from http://www.evrythng.com/.

Factual. 2016. Factual Home Page. Retrieved March 14, 2016, from http://www.factual.com.

Gnip. 2016. Gnip Home Page. Retrieved March 14, 2016, from http://gnip.com/.

Dominique Guinard, Vlad Trifa, and Erik Wilde. 2010. A resource oriented architecture for the Web of Things. In *Proceedings of the IEEE International Conference on the Internet of Things*. 1–8.

J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami. 2014. An information framework of creating a smart city through Internet of Things. *IEEE IoT Journal* 1, 2, 112–121.

Juong-Sik Lee and Baik Hoh. 2010. Sell your experiences: A market mechanism based incentive for participatory sensing. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*. 60–68.

Chao Li and Gerome Miklau. 2012. Pricing aggregate queries in a data marketplace. In *Proceedings of the 15th International Workshop on the Web and Databases*.

Microsoft. 2016. Windows Azure Marketplace. Retrieved March 14, 2016, from http://datamarket.azure.com/.

Balamuralidhar Prateep Misra and Arpan Pal. 2013. Software platforms for Internet of Things and M2M. *Journal of the Indian Institute of Science* 93, 3, 487–497.

Knud Möller and Leigh Dodds. 2012. The Kasabi information marketplace. In *Proceedings of the 21st WWW Conference*.

Dejan Munjin and Jean-Henry Morin. 2012. Toward Internet of Things application markets. In *Proceedings of the IEEE International Conference on Green Computing and Communications*. 156–162.

Alexander Muschalle, Florian Stahl, Alexander Löser, and Gottfried Vossen. 2012. Pricing approaches for data markets. In *Proceedings of the 6th International Workshop on Business Intelligence for the Real Time Enterprise.*

Nimbits. 2016. Nimbits Home Page. Retrieved March 14, 2016, from http://www.nimbits.com/.

OMG. 2007. *Data Distribution Service for Real-Time Systems* (1.2/07-01-01 ed.). Object Management Group.

Sooksan Panichpapiboon. 2010. The future of traffic information systems. *Journal of Information Science and Technology* 1, 2, 70–78.

Marco Picone, Michele Amoretti, and Francesco Zanichelli. 2012. A decentralized smartphone based traffic information system. In *Proceedings of the 2012 IEEE Symposium on Intelligent Vehicles*. 523–528.

Fabian Schomm, Florian Stahl, and Gottfried Vossen. 2013. Marketplace for data: An initial survey. *SIGMOD Record* 42, 1, 15–26.

A. Segura, J. Cuadrado, and J. Lara. 2014. ODaaS: Towards the model-driven engineering of open data applications as data services. In *Proceedings of the IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations*. 335–339.

Sentilo. 2016. Sentilo Home Page. Retrieved March 14, 2016, from http://www.sentilo.io/wordpress/.

J. Spillner and A. Schill. 2013. A versatile and scalable Everything-as-a-Service registry and discovery. In *Proceedings of the 3rd International Conference on Cloud Computing and Services Science (CLOSER'13)*. 175–183.

ThingSpeak. 2016. ThingSpeak Home Page. Retrieved March 14, 2016, from https://www.thingspeak.com/.

ThingWorx. 2016. ThingWorx Home Page. Retrieved March 14, 2016, from http://www.thingworx.com/platform.

Xiaoyu Tong and Edith Ngai. 2012. A ubiquitous publish/subscribe platform for wireless sensor networks with mobile mules. In *Proceedings of the IEEE 8th International Conference on Distributed Computing in Sensor Systems*. 99–108.

H.-L. Truong, M. Comerio, F. De Paoli, G. R Gangadharan, and S. Dustdar. 2012. Data contracts for cloud-based data marketplaces. *International Journal of Computer Science and Engineering* 7, 4, 280–295.

Ubicon. 2016. Ubicon: Smart Ubiquitous and Social Computing. Retrieved March 14, 2016, from http://ubicon.eu/about/.

Bruno Valente and Francisco Martins. 2011. A middleware framework for the Internet of Things. In *Proceedings of the 3rd International Conference on Advances in Future Internet*. 139–144.

Quang-Hieu Vu, Tran-Vu Pham, Hong-Linh Truong, Schahram Dustdar, and Rasool Asal. 2012. DEMODS: A description model for data-as-a-service. In *Proceedings of the IEEE International Conference on Advanced Information Networking and Applications*. 605–612.

Xignite. 2016. Market Data Cloud. Retrieved March 13, 2016, from http://www.xignite.com/market-data/market-data-cloud/.

Xively. 2016. What Is Xively? Retrieved March 14, 2016, from https://xively.com/whats_xively/.

Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang. 2012. Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking*. 173–184.

Received November 2014; revised September 2015; accepted January 2016