An Examination of the Internet of Things through the Data Management Perspective

[1] Wei Wang, [2] Woonkian Chong, [3] Dawei Liu, [4] Hai-Ning Liang, [*5] Ka Lok Man,
[3] Yo-Sub Han, [4] Seungmin Rho, [5] Sang-Soo Yeo

[1]Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, China, {wei.wang03, dawei.liu, haining.liang, ka.man}@xjtlu.edu.cn
[2] International Business School Suzhou (IBSS), Xi'an Jiaotong-Liverpool University, China, woonkian.chong@xjtlu.edu.cn
[3] Department of Computer Science, Yonsei University, Republic of Korea, emmous@cs.yonsei.ac.kr
[4] Department of Multimedia, Sungkyul University, Anyang-si, Republic of Korea, smrho@sungkyul.edu
[5] Division of Convergence Computer & Media, Mokwon University, Daejeon, South Korea, sangsooyeo@gmail.com

Abstract

Internet of Things is seen as one of the most exciting initiatives towards the evolution of the current Internet, contributed by research communities and industries of networking, ubiquitous/pervasive computing, distributed/parallel computing, semantic Web, service oriented computing, data mining and management, telecommunication, social science and many others. The scope of its research and development has been continuously extended over the years due to its interdisciplinary nature. Once, we believed that object identification and addressing using the Radio-Frequency Identification tags (a "Things" oriented perspective) should be the pre-requisite for the Internet of Things. Later, we witnessed the research with the "Internet" oriented perspective which focuses on communication and internetworking among Things. After, the "Semantics" and "Service" oriented perspectives emerged and aim to provide an interoperable and scalable infrastructure for the cyber-physical-social applications and, more generally, business applications on the future Internet. The most recent trend is, however, the "Data" oriented perspective, and this deals with data collection, distribution, storage, processing, and finally both query and analytics. The strategic shift of the focus, to a considerable extent, is a result of the recognition of the importance of deriving useful knowledge and creating real values from the tremendous amount of data collected from our surrounding environments and the physical world. To this end, it is also in line with one of the most eye-catching research topics-"big data analytics", which aims to derive values from data with large volume, high velocity and different variety. While a number of surveys on the Internet of Things research are available, none of them appears to have investigated the state-of-the-art from a "data-oriented" perspective. Due to its particular importance, this paper provides a review of the current techniques and methods for data processing and management on the Internet of Things at different levels and point out some of the future research issues. The main objective of the paper is to enable readers to gain a better understanding of this challenging and sophisticated problem, and to further explore and contribute to the related research.

[*] Corresponding Author
Received: Mar. 15, 2014, Revised Apr. 30, Accepted: Jun. 25, 2014

***Keywords**: Internet of things, Data management, storage, and analytics, Business applications, Cyber-physical-social applications, Mobile communications*

1. Introduction

The Internet of Things (IoT) in the broad sense refers to a system of uniquely identifiable objects and their virtual representations in an Internet-like structure—that is, a distributed network of smart objects that can be used standalone, or integrated into complex distributed applications. The term was first coined by Kevin Ashton in 1999 [1] and since then the concept has grown rapidly in popularity and given rise to a number of research initiatives. IoT was first implemented in applications for market analysis and in those early days radio-frequency identification (RFID) was seen as the foundational enabling technology. In the past, the idea was mainly to see how machines (or devices) could be connected and enabled to communicate with each other. Today, the term IoT has a slightly different and broader connotation, and denotes advanced connectivity of devices, systems, and services that goes beyond simple machine-to-machine interoperability. To achieve its current purpose, IoT will need to cover a variety of communication protocols, domains, and applications [2].

In essence, the grand idea of IoT is that if all things (including people) could be equipped with identifies, which are passive or actively emitting signals for other devices to receive, they could be tracked and inventoried automatically by computers [1, 3, 4]. By attaching these identifiers and signal emitters, life as we know it could be transformed in significant ways [3, 5, 6]. One simple example is with business inventories, which will never be out of stock or generate any waste because all parties involved (e.g., suppliers, buyers, and the businesses) will know what is available and in stock, what items are needed, what products are sold and their expiry dates, and so on [6].

In an age witnessing an ever-growing number of items that have an impact on our lives, the potential use of having a well-established IoT is immense. According to market research groups, the devices with the power to be connected to an IoT-like network could be more than 26 billion by 2020 [7, 8]. The bulk of these objects will be composed of mobile devices (e.g., phones or tablets), which in 2014 stand at more than 7.7 billion. Currently, these devices still rely on Wi-Fi or Bluetooth transmission mechanisms, which are power intensive. In the future, other low-power mechanisms will be devised, and this will lead to a further explosion of devices with IoT connection capabilities.

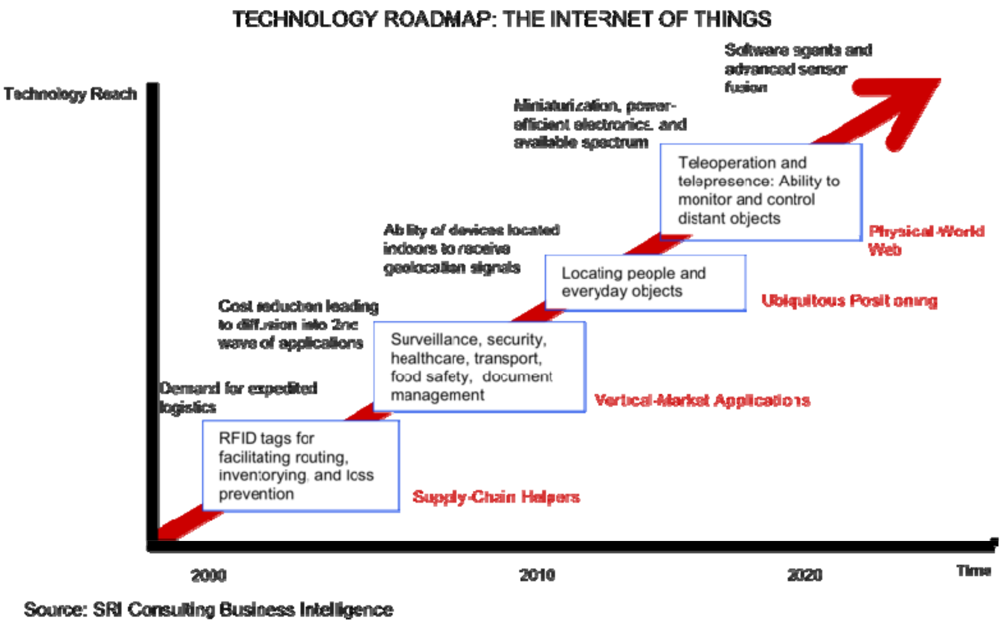

Figure 1. The technology 'roadmap' of the Internet of Things

IoT started with support from existing research communities and industries in networking, ubiquitous/pervasive computing, distributed/parallel computing, semantic Web, service oriented computing, data mining and management, telecommunication, and domains in the social sciences. IoT can be applied to activities beyond those in these areas but to enterprises of increasing societal importance such as waste management, urban planning, environmental sensing, social interaction gadgets, sustainable urban environment, continuous care, emergency response, intelligent shopping, smart product management, smart meters, home automation and smart events. The development and change of the focus of IoT is shown in Figure 1 (source: SRI consulting business intelligence). In its early days, IoT was primarily conceived as a system to facilitate inventory and prevention of loss due to theft or misplacement. The main technology is RFID tags. The main beneficiaries are supply-chain helpers. The next evolutionary phase of IoT includes surveillance, security, healthcare, transport, food safety, and document management, which reflect the demands and needs of society at the time. After, the emphasis is on devices to locate people and every object, as we move on to allow more social connectivity and interaction. The future of IoT is on the miniaturisation of devices and power efficient electronics so that more things can be added to IoT with minimal cost, in addition to allow for monitoring and controlling distant objects. As the implementation of IoT matures, its uses will expand and more application domain will likely take advantage of it.

Underlying the whole structure of IoT is *data*, in immensely huge amounts. This is the focus of this paper: IoT from a data management, data centric perspective. We have this focus in recognition of the importance of knowledge (or insight) derived from the growing amount of data that are collecting. Our intention here is not to survey all the techniques on big data analytics (especially those developed by the machine learning and data mining communities for data visualization, predictive analysis and classification) for any particular application domains. Our survey is based on the management of the data generated by sensors, which is of tremendous scale, highly distributed, time and location-dependent, dynamically changing (high degree of velocity), and often associated with formal semantics. We will review existing techniques and methods for domain or application independent data processing and management. Although data security, privacy, and trust are important aspects of data management, they are not the focus of this paper, and thus not covered here.

The rest of the paper is structured as follows. In the next section, Section 2, we discuss some new developments related to mobile devices as important enabling and communicating tools in the future of IoT. Section 3 is the main bulk of this paper and in it we describe further the need to focus on data management for IoT, what we need to be aware of in order to better utilize the data, and the latest trends in data management. Finally, before concluding, we discuss the future of IoT from a data centric perspective.

2. Current developments on mobile device communication

One of the most important aspects of IoT deals with the network requirements. In this section we provide a brief survey of network requirements and also the advances of mobile devices.

As stated earlier, in the past, the emphasis of IoT had been on object identification and addressing using the Radio-Frequency Identification (RFID) tags, which would allow wireless communication among the tagged objects. We are now experiencing the explosion of cheap mobile devices. Networking mobile devices of different types *will be*, or already *is*, an i

mportant feature of IoT. Mobile devices like RFID and sensors are specially designed to have low battery consumption and small package size. This means limited communication and computation capabilities.

Instead of using the standard OSI 7-layer protocols, IoT platforms typically make use of tailored protocols in communication. NanoStack [9], a protocol stack designed for wireless sensor networks, has included only IEEE 802.15.4, IPv 6LoWPAN, and UDP in the stack, based on the consideration that (i) these protocols are well-known energy-efficient protocols, and (ii) many other protocols are not necessary in IoT communication. The compromise is that the developers would not have many choices in designing the platform for different purposes. In addition to the underlying communication protocols, the application protocols can also be customized. A tailored XML parser for web service is proposed for wireless sensor network application [10]. Since sensor nodes only respond to messages specified in their own WSDL, many web service components that are designed for WWW are no longer needed.

A commonly used computation platform is TinyOS, a flexible application-specific operating system developed for sensor networks [11]. With a pair of AA batteries, a sensor node that operates on the TinyOS can work for at least 15 hours in active mode and up to 5 years in idle mode [12]. The rise of the TinyOS has stimulated many other platforms such as the hardware platform Mica [13], Telos [14], the query processing system TinyDB [15], and the Java Eclipse platform TinyDT [16]. These platforms together have made IoT a technology accessible by ordinary developers. However, this leads us to the question of data transfer and fusion across devices and systems.

Data fusion plays an important role in IoT platforms design. Take smart transportation as an example. Individual GPS data is nothing but a pair of geographical coordinates. For the purpose of navigation, GPS coordinates must be coupled with a digital map. Combined with other sensors' data, GPS data can be used to evaluate a vehicle's condition and make recommendations like "it is time to change brake pads". GPS data can also help traffic analysis. Real-time traffic information of a given road estimated based on the vehicles traveling in that road could help drivers make better decisions. All these applications rely on data fusion of multiple sensors.

A widely used data fusion method for multiple different sensors is the Kalman filter [17]. The Kalman filter can relate the location information obtained from GPS to the traveling distance obtained from an odometer and the direction information obtained from a compass. When GPSs cannot work properly, such as inside a tunnel where GPS signals are completely blocked, the above-mentioned location-related information can help the vehicle location and navigation(help locate and navigate the vehicle). Other commonly used data fusion methods include particle filtering [18] and neural network [19].

Traffic data analysis is a problem involving many objectives, some of which are conflicting. One needs to possess a complete knowledge of every vehicle's location and travelling information in order to give an accurate analysis of the traffic condition. For this purpose, many data mining methods have been developed [20]. On the other hand, individuals' location information should be kept private. Some privacy preserving methods have been proposed to protect the location information from being discovered [21].

A serious problem that has arisen from the pervasive use of IoT is information security. Let's consider again the location services. In early systems, location service providers such as GPS were considered reliable. In an IoT platform, a user can be located in many different ways: sensor nodes in a sensor network can locate each other in a cooperative approach; an indoor cell phone user can be located based on the WiFi signal strength. Location in these platforms is no longer reliable. A sensor node can be malicious [22] and the WiFi signal can be spoofed [23]. How to protect IoT platforms against malicious attacks is still a challenging problem since IoT, unlike other traditional computing platforms that are well-designed beforehand, may not have a reliable infrastructure. Although security is an important aspect of IoT, this will not be the purpose of this paper. The main focus of this paper is on data issues with an IoT, and we will elaborate further on them next.

3. Data management in Internet of Things

Having established the fundamental infrastructures, the research, industry and business communities realize that the data indeed is the most valuable "Thing" in the Internet of Things. The strategic shift of focus is a result of the recognition of the importance of deriving useful knowledge and creating real values from the tremendous amount of data available at our disposal. With the knowledge, businesses can, for example, identify profitable opportunities from potential market segments and design effective marketing campaigns; flood monitoring systems can initiate instantaneous warnings in dangerous situations and suggest the best evacuation plans.

It should be noted that data in many IoT applications undoubtedly could be seen as a type of big data, which is often characterized by its volume, velocity and variety [24]. For example, Boeing jet generates 10 terabytes of data per engine every 30 minutes of flight, and a cross-country flight from New York to Los Angeles on a twin-engine Boeing 737 will generate a massive 240 terabytes of data [25]. However, the intention here is not to survey all the techniques on big data analytics (especially those developed by the machine learning and data mining communities for data visualization, predictive analysis and classification) for any particular application domains. Specifically, we focus our discussion on the management of data generated by sensors, which is of tremendous scale, highly distributed, time and location-dependent, dynamically changing (high degree of velocity), and often associated with formal semantics.

Our objective is to provide a review of the prevailing techniques and methods for domain or application independent data processing and management. These techniques and methods, as part of the middleware family for many IoT applications, are developed to alleviate the problems introduced by the extremely large scale and high velocity of IoT data. They provide distributed functionalities on data fusion, integration, processing and abstraction, and produce valuable and meaningful information granules which could significantly facilitate the analytics tasks (such as statistical analysis or knowledge extraction) in high-level business services and applications.

3.1. Data representation

Data on the IoT, such as sensor data, is not meaningful unless it is associated with certain contextual metadata in the spatial, temporal and thematic dimensions [26]. For example, reading from a temperature sensor only makes sense if we know where and when the value is sampled and for what object the temperature is measured. In its simplest form, IoT data is annotated with time-stamp and stored as time-series data [27] (other metadata might need to be stored in separate locations).

Given the extraordinarily large amount of things, the size of the metadata on the IoT becomes significant and needs to be considered carefully. The Open Geospatial Consortium has a suite of standards that aim to discover and access sensor networks and archived sensor data using standard protocols and representation methods, such as sensorML [28] (based on XML formats). Similarly, 52North's [29] sensor observation service is implemented to query sensor observation data stored in a database or sensor descriptions stored in XML files. However the representation based on XML does not enable semantic reasoning over the sensor data as well as the metadata.

The work by [30, 31] proposes to use semantic technologies (e.g., ontologies) to model things and the data generated on the IoT. Things and data are usually descried according to their spatial, temporal and thematic characteristics using formal languages such as RDF (Resource Description Framework [32]) or OWL (Web Ontology Language [33]) designed for the semantic Web. The semantic descriptions not only provide support for interoperability and enable semantic discovery, but also facilitate creation of machine-interpretable or human-understandable knowledge using data abstraction techniques [34]. Semantic modelling of IoT data is particularly powerful, e.g., to enable more meaningful semantic discovery and search [35], when linked data is created based on the semantic descriptions. In fact many of the data processing models and techniques on different layers of the IoT have their foundation built on semantic technologies. It should be noted, however, that there is a trade-off between the expressiveness of the data semantics and the complexity of reasoning. For this reason, researchers propose to use lightweight semantics for the IoT applications [31, 36]. A notable work by [37] for efficient management of semantic data utilizes a vertical partitioning approach with column-oriented database system in order to provide better performance.

3.2. Data storage

Storage of the IoT data usually depends on the nature and requirements of the applications. In general, three kinds of storage can be identified: persistent, ephemeral and semi-persistent. It is not unusual for many applications to use a combination of these different storage methods. Persistent storage is suitable for historical IoT data which can be retrieved later for analytics and visualization purposes. Stream and event based applications mostly provide transient or semi-persistent storage for IoT data. Once the incoming data streams are processed, they are then discarded and the newly derived streams or events (which represent the higher-level abstraction of the raw data) can be stored in the persistent storage [38, 39]. Some of the existing platforms for stream or event processing are specially designed to handle big data, such as [40, 41]. Semantic repositories provide persistent storage for semantic data, especially linked data for the things; however, they are not suitable for the large-scale, dynamic streaming data. Another way of providing ephemeral data storage is through the services, which expose standard interfaces for clients so that they can access and retrieve data, for example, using traditional Web services or RESTful service.

3.3 Stream processing and reasoning

Stream processing is a technique for information flow processing in which streams of data coming from different sources are processed by a set of rules in a timely manner to produce new data streams [42]. It is a powerful computing paradigm which is able to accommodate the velocity problem of big data analytics and to analyze data in real-time with micro-latency [24]. While traditional database management systems are designed to manage persistent data (i.e., gathering, manipulating, storing and processing large amounts of data), systems for stream data are designed to manage transient data and to process standing or continuous queries [38]. The output from the stream processing systems is continuously updated as new data becomes available. Result of the queries can be exact or approximate depending on whether the system has sufficient storage capabilities to store historical data or not. Some of the examples in IoT applications include stream data access, subscription and query in environmental decision support systems, e.g., flood emergency response [43]; and RFID stream data analysis in inventory management systems to track valid paths of shipments and to capture events [44].

There have been considerable research interests towards integrating the capabilities offered by the current data stream management systems with background knowledge to perform stream reasoning [45]. Much of the efforts in this line come from the semantic Web community, which anticipates that combining stream reasoning with existing streams processing methods enables more complex user queries to be answered. For example, how many times the reading of the temperature sensor installed in a particular geographical location exceeded the threshold in the past hour? More importantly, this enables us to discover previously unknown patterns or insight (e.g., traffic and road conditions in highways) that might be of interests to the user, rather than merely providing integrated sensor readings for the continuous queries. Some of the tools for semantic reasoning over the streaming data have been made available. The streaming SPARQL language [46] extends the logical SPARQL algebra based on the foundation of a temporal relational algebra and enables query and simple reasoning over RDF streams. C-SPARQL [47] is another extension to the original SPARQL languages and supports continuous queries using windows. The authors show how the language can be used to perform reasoning to facilitate decision making

3.4. Complex event processing

The ability to process large amounts of stream data is essential for IoT applications which need continuously updated data, e.g., meteorological monitoring and prediction. However, the extraction of events or complex patterns from the data in motion are usually not included in stream data management systems. This naturally leads to another paradigm in information flow processing, called complex event processing [39], which aims to address the limitation of most data stream processing systems, i.e., the ability to detect complex patterns of incoming items which involve sequencing and ordering relationships [42]. The complex event processing model detects complex events according to the defined event patterns and notifies the interested subscriber applications. A simple but classic example of the complex event processing in IoT is fire detection based on the readings of temperature and smoke sensors.

The need for complex event processing is due to the fact that a lot of information items pass silently back and forth through our information systems as unrelated pieces of communication; if aggregated together and correlated, they become a source of great power and will yield a wealth of information [39]. Nowadays, complex event processing plays an important role in developing enterprise applications which need to immediately react to events signifying opportunities or threats that are critical to businesses. It has been widely used in many enterprise applications as well as the IoT, e.g., smart hospitals using RFID events [48] and sensor event detection in smart spaces [48]. However, most of the event processing techniques are confined within the boundaries of enterprises. The challenge in applying complex event processing into the IoT is how to define and detect the relationships between different events generated at a tremendous scale and in different representation formats.

This leads to semantic complex event processing [49], which aims to enable reasoning over event streams with background knowledge to support event detection in even more complex situations, leveraging the value and power of the open knowledge created on the semantic Web. In the ETALIS open source system, background knowledge can be evaluated on the fly either to capture domains of interest or to prove relationships between different types of events [50]. An application on route planning for goods delivery in the city of Milan is developed to show the effectiveness of the semantic event processing based on an ontology about the city. The system allows users to specify complex event patterns in the Event Processing SPARQL language that are continuously evaluated over the incoming events and the background knowledge [51]. The work in [52] uses ontologies to specify and recognize complex events from heterogeneous sensor networks, which are subsequently translated into native languages in commercial complex event processing systems for execution. Hasan et al. propose to enrich the event data from sensors with linked semantic data to support situational awareness for decision-making in enterprises [53].

3.5. Cloud computing

Cloud computing is a powerful paradigm that provides computing resources as on-demand services to businesses and individual users. The services, including processing power, storage, data analytics and software functionalities, as well as the hardware in the cloud data center can be delivered as utilities with the pay-as-you-go pricing model in a fully virtualized manner [54, 55]. The National Institute of Standards and Technology defines cloud computing as a model for ‘enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction’ [56]. It identifies three service models, i.e., software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS), and different types of clouds, e.g., public, private, community and hybrid. Some of the prominent examples include the [57-59].

No matter what the tasks are, large scale data storage, application deployment and hosting, or data processing and analytics using providers’ applications, clouds can provide the needed functionalities with great flexibility. Companies no longer need to make huge investments in purchasing and maintaining the software licenses and hardware infrastructure. The salient benefits of cloud computing such as rapid elasticity, economies of scale, and cost (capital and spending) reduction, are also attractive to IoT applications. As discussed earlier, data management on the IoT includes a variety of tasks, such as data collection, integration, storage, processing, abstraction and analytics, all of which can potentially be accomplished in a service-oriented way (e.g., SaaS and IaaS) with the clouds.

A cloud centric framework for IoT is envisaged in [60] to realize the full potential of cloud computing and ubiquitous sensing. For example, sensing service providers can offer sensor data through storage clouds, data mining and analytics application developers can provide software which derive useful information or knowledge from the large amount of raw data and can deliver the software as services through the clouds. A case study using the Aneka [61] cloud computing platform, which offers a runtime environment and a set of APIs allowing developers to build customized IoT applications, is performed to show how interactions between private and public clouds such as the Azure [61]. There has already been some work implementing IoT applications based on private clouds. The work presented by [62] applies cloud computing to efficiently store and process data collected from wireless sensors for the purpose of real-time monitoring of the saturated line, impounded water level and the dam deformation. Kovatsch et al. [63] propose the idea of a thin server, which is to decouple application logic and to implement it in the clouds. They argue that the idea can facilitate the design of an IoT infrastructure agnostic to applications and enable developers with different backgrounds to provide applications for the heterogeneous device types on the IoT.

In the past few years, community supported public clouds have been made available for users and businesses to connect their own objects and create their own IoT applications. Xively (formerly known as Pachube and later Cosm) [64], offers tools and services through it PaaS to allow devices, data, places, users and application to be created and interconnected through the ‘Connected Object Cloud’. SensorCloud [65] is another example of public cloud providing management platform for the IoT applications. It supports rapid visualisation and user programmable analysis for data collected from large, distributed sensor networks. ThingSpeak [66] is an open application cloud platform designed to enable meaningful connections between things and people. It supports a variety of IoT applications such as sensor and energy monitoring, device and system connection, geo-location tracking, social network interfacing and RFID transactions.

3.6. Fog computing and intelligence at the edge

A relatively new computing paradigm, fog computing, has been recently proposed by [67], which aims to extend the cloud computing paradigm to the edge of the IoT, e.g., wireless networks, gateways or routers. Fog computing provides localisation and enables low latency and context awareness; while the cloud computing provides global centralisation. The two computing paradigms complement each other to provide the desired functionalities and requirements for many IoT applications. For example, at the lowest level, fogs perform simple processing of the data collected from wireless sensor networks and issue commands to actuators. They usually provide ephemeral or semi-permanent data storage. At the same time, they will send abstracted data to the high level, e.g., clouds, for visualisation, reporting and analytics purposes, which provide the global coverage and are used as permanent data repository for processing and analytics purposes.

The idea of fog computing in fact coincides with the concept of intelligence at the edge [68, 69], initially proposed in the Internet and network research. Distributing the intelligence at the edge or edge cloud provides great flexibility to serve nomadic users appearing at the edge rather than asking the users to adapt to the network. [70] stated that edge computing 'pushes the frontier of computing away from the centralised nodes to the logical extremes of a network', which enables analytics and intelligence creation near the data source. Both ideas are seemingly familiar to the IoT researchers and practitioners, who have see many smart applications (e.g., smart object, smart office and etc). As indicated by [71], the future Internet could include 'public/private infrastructures and dynamically extended and improved by edge points created by the "things" connecting to one another'. In this sense, we can expect that fog computing or edge computing will become even more popular and important in IoT development.

4. Implications to business opportunities

IoT is a novel paradigm that would have a wide range of business opportunities in the future such as in the areas of transportation and logistics, health care, smart environments (in homes, offices, and public places), and personal/social relations management [72, 73]. Since IoT was first proposed, small numbers of enterprises take the view that IoT can improve and enhance the extent of their business activities. Recent studies [1, 72, 73] indicate that enterprises have started to respond positively to the changes brought about by IoT as it provides a favorable environment for businesses to lower their operating costs, to have better opportunities to promote their products/services, and to enrich their overall marketing strategies. The ability of IoT to increase market efficiency aggregated over the entire economy makes the provision of business services an important platform with potentially large rewards. From the business perspective, the benefits of IoT as reported by many academics and practitioners include:

- Reducing search costs by facilitating comparison of price, products, and services [74],
- Improving production and supply capability [75],
- Improving personalisation and customisation of product offerings [76],
- Enhancing customer relationships [76],
- Reducing marketing costs compared to traditional marketing media [77],
- Reducing the number of staff [78],
- Facilitating global presence [72, 73], and
- Interacting more effectively in terms of services business communication [75].

Due to the above benefits, the integration of customer relationship management (CRM) and IoT have been studied by business practitioners, academics and researchers (e.g., [74-76]). IoT will be likely the main CRM resource for enterprises for customer relationships development [72, 73]. Generally, enterprises are the main element to investigate the extent of usage in terms of the attitude and perception to IoT. CRM has encouraged the growth of markets and IoT-based solutions could be touted as one of the primary drivers of the business models. In order to achieve this wider perspective, there is a need to review the benefits of the integration with IoT for businesses, and the improved businesses' ability and results. It is also crucial to explore best practices for enterprises to capitalize on the opportunities provided by IoT by the integration of current CRM systems with their existing marketing channels and practices to increase consumer satisfaction, loyalty and profitability.

Furthermore, the integration of supply chain management (SCM) with IoT has already been explicitly recognized and deliberately included in business studies [79, 80]. Enterprises have realized that IoT is emerging as a platform that is highly integrated with SCM technology, and can thus enhance the performance of their existing operational practices. The adoption level of IoT will be high, *only* if the understanding of the benefits for businesses to achieve realistic expectations of SCM technology to fit their supply chain networks. Among the benefits of developing supply chain efficiencies using IoT are (1) improved demand forecasts, (2) efficient logistics information management, and (3) reduced cycle times [79, 80]. Traditionally, supply chain networks do not achieve this without high overhead cost. Even in the case where each party shows evident willingness to cooperate, businesses are still subject to the problem of the case that one link fails and this can lead to the entire chain being defunct. In order to make SCM more resistant to unexpected failures, it is useful to use IoT as a substitute platform for each participating business.

To fully exploit the potential of business opportunities derived from IoT, future studies will need to address the issues of integration, interoperability, and verification. The infrastructures required to support IoT can also be applicable to enterprises; because of the needed interoperability amongst communicating devices of great diversity and properties, IoT can be a platform which can allow for open business and trading environments. An important aspect of future research is to provide a suitable specification framework as a guideline for businesses. Although the breadth of activities pursued in the field is limited at present, the continued growth of IoT-related business opportunities will enable enterprises to engage in currently under-utilized applications. Enterprises need to perceive that the benefits of IoT outweigh the costs of doing business within an IoT-like structure. As stated by Uckelmann and Harrison [79], IoT provides various incentives to help enterprises in lowering the costs when they want to extend their geographic reach, in enhancing customer relationships, and in improving sales and gains. In order to reap the benefits, businesses must fully embrace it by take advantage of the potentials of the IoT.

5. Discussion

Data management for IoT application is an extremely complex task, which needs cooperative efforts from many fields, such as wireless sensor networks, database, cloud computing, distributed computing, data mining, machine learning and big data analytics. Our bird view of the relevant research and development in this field shows that no single data processing technique is versatile for all IoT applications; different methods and techniques co-exist and complement each other to realize the requirements of different applications, e.g., the size of the business and number of customers served. In a smart office or smart home scenario operated locally, a centralized data storage and processing system would be sufficient; however, if the system is to be monitored and controlled remotely, then transferring the storage and processing to a cloud would be plausible.

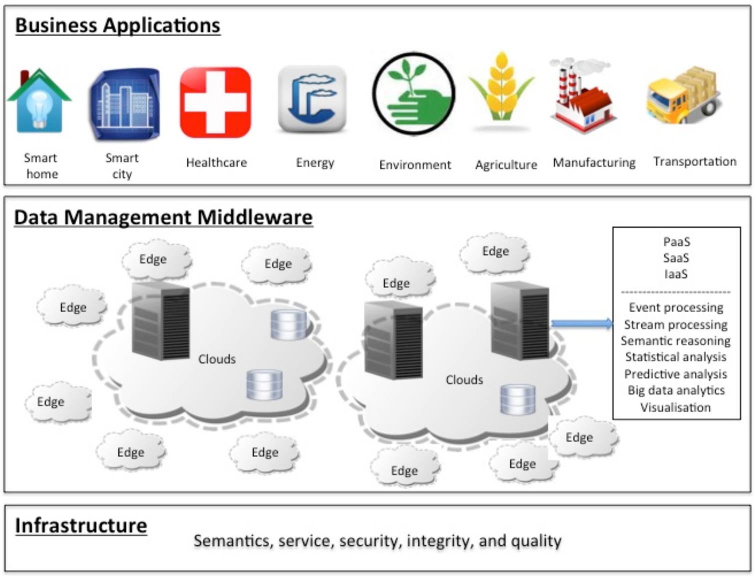

Figure 2. The data management middleware for the IoT applications

The nature of the IoT data is very different from the ones associated with many conventional data types, e.g., large quantity and scale, heterogeneous formats and types, high degree of distribution, time and location dependency, varying quality, considerable velocity, etc. As we can see, some of these characteristics match the ones that define the big data. The objective of data management on the IoT is also in line with objective of big data analytics, which is to derive true value for businesses from the raw data with high volume, velocity and variety. In IoT, this can be interpreted as creating useful information granules at different abstraction levels, such as a processed data stream, a complex event, a piece of knowledge deduced from semantic reasoning, which can facilitate decision making using various statistical or predictive analysis tools. We envision a highly distributed, hybrid while flexible architecture for IoT data management as shown in the following diagram.

The infrastructure of the IoT data management is constructed with the existing technologies around the four perspectives for IoT, i.e., Things, Internet, Semantics and Services, which have been well investigated in previous studies. In IoT data management, security, integrity and quality of the data are also important issues; interested readers can refer to [36, 72, 81] for more information related to those areas.

The fifth perspective, data, has been considered as the most significant dimension in IoT research. Ultimately, deriving values from data and making use of it in creating more intelligent applications are essential in realizing the vision of the IoT. We believe apparatus developed in distributed computing and big data analytics (e.g., cloud computing [55], Hadoop [82] and MapReduce [83]) are most prominent in addressing the challenges in future IoT data management. As Figure 2 shows, raw data collected directly from physical devices can be cleaned and integrated at the edge nodes (or the fogs); the summarized data is then sent to the clouds for further processing. It should be noted that, however, not all the data will be integrated at the edges, especially when data is scarce. This step not only helps reducing the size of the data and the network traffic, but also provides certain level of quality assurance. The data sent to the clouds can be processed using different methods and techniques, such as stream and event processing, statistical analysis, data mining, machine learning, pattern recognition and etc, often offered as PaaS or SaaS. Software developed using the MapReduce framework [83] is particularly efficient for distributed processing of the IoT data, as the nature of the computations involve linearly computable statistical functions over the elements of the data stream [81]. Finally, the processed data, in the form of valuable information or knowledge at different granules, is either stored in databases or offered through standard services interfaces for different business applications and opportunities.

There can be a myriad of business applications within IoT, and businesses can rip many benefits directly or indirectly from it. There were 2.5 billion devices (e.g. smart phones, desktop personal computers, laptops, tablets, etc.) connected to the internet in 2009 with unique IP addresses; this number will reach 30 billion in the next decade with most of these requiring or providing products or services [84]. Gartner [8] predicts that by 2020 the total business value for IoT will be US$1.9 trillion, benefiting a wide range of industries. Similarly, it estimates that by 2020 the revenue contribution from IoT products/services such as hardware (e.g. wireless sensor, RFID devices), software, and information communications services that are associated with making uniquely identifiable "objects' or "things" will reach US$309 billion [8]. Many practitioners have seen the transformation from niche implementations to a more coherent, widespread and integrated IoT that can provide global accessibility, a convenient way to update in real-time information services, and unique ways to customize and personalize capabilities. Therefore, it is crucial for IoT to work with the rest of the businesses to understand the impacts and potential opportunities arising from this evolution. From our perspective, the most crucial aspect of making the best from IoT is for business to be able to store and manage the rich amount and variety of data that we can obtain from a well-established IoT. Proper storage and management of data will then allow for higher order use of the data, that is, for data mining and analysis to find useful patterns, trends, and relationships that may prove of significant value to business.

6. Summary and Conclusions

In this paper, our focus is on exploring Internet of Things (IoT) from the data management perspective. IoT is one of the most exciting developments with a wide of implications in transforming how we do business and businesses interact with each other. It is comparable to how the Internet becomes the enabling technology that has led to a revolution in the way we share and produce information, and with it significantly changing the entire business and research landscape. IoT is predicted to bring about a similar revolution.

The idea of IoT is gaining momentum but its implementation and realization is still lagging behind. As with revolution is other domains, there is need to have a catalyst. The catalyst in the context of the Internet is cheap computers that masses of people can purchase and use. The equivalent catalyst for IoT could be mobile devices whose number is growing exponentially and whose price is dropping rapidly. By mobile device is not simply meant smartphones and tablets, but other devices of low cost, with low power processors and other communication components that it practical to produce in large quantities and embed these low cost devices into objects and infrastructures in ways that were not possible in the past. A network of integrated devices communicating with other at all times will bring about efficiency and productivity in unprecedented ways. One of the greatest challenges that it also brings is data management (e.g., storage and retrieval) and the subsequent desire to analyse the vast amounts of information generated to find patterns, trends, and anything that is useful to help us understand how people and business function and based on this analysis to improve things further.

It is therefore for this reason that we need to examine IoT from the data management perspective. From a survey of a large number of papers, we have attempted to provide a big picture of the current developments in IoT. We first provide a summary of current developments in mobile device communication to set the foundations on which we can discuss in more details data management requirements for IoT. In the subsequent sections we discuss the data management issues related to IoT, including data representation, storage, processing, and distribution. In particular, we want to bring to the foreground the latest development in data processing, as it is becoming evident that it is challenging to process the immense amounts of data we have and will have. We mention the stream and complex event processing and reasoning, and some of the new techniques that have been developed. We conclude the section with a short description of cloud computing and fog computing, two new paradigms for distributed data storage and processing, and how they can be integrated into the IoT infrastructure. In the next section we pay attention to the potential of IoT in creating business opportunities, and in particular how current business practices can be enhanced.

The main objective of this paper is to provide our readers with a broad and extensive survey of the current technological developments in data management in the context of IoT, which would be able to encourage further research into this exciting initiative. The creation of a feasible IoT is only possible if researchers and practitioners from diverse areas, such as networking, ubiquitous/pervasive computing, distributed/parallel computing, semantic Web, service oriented computing, data mining and management, telecommunication, and the social sciences, can come together and bring their expertise together. Similar, there is need for governments to join in this enterprise, something akin to what has happened when countries started investing heavily into infrastructures that have allowed people to have access to the Internet and get connected to virtually all places on earth. The Internet has made the world a lot *smaller*; the IoT can make the world a lot *smarter*.

7. References

[1] K. Ashton, “That 'Internet of Things' Thing, in the real world things matter more than ideas,” *RFID Journal*, June 2009.

[2] J. Höller, V. Tsiatsis, C. Mulligan, S. Karnouskos, S. Avesand, and D. Boyle, “From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence,” in *Academic Press*, 1st ed. 2014.

[3] R. van Kranenburg, E. Anzelmo, A. Bassi, D. Caprio, S. Dodson, and M. Ratto, “The Internet of Things,” in Proc. of the 1st Berlin Symposium on Internet and Society, Berlin, Germany, Oct. 2011.

[4] Commission of the European Communities, “*Internet of Things–An action plan for Europe*,” 2009.

[5] P. Magrassi, A. Panarella, N. Deighton, G. Johnson, “Computers to Acquire Control of the Physical World,” *Gartner Research Report T-14-0301*, Sept. 2001.

[6] C. Associati, “*The Evolution of Internet of Things*,” 2011.

[7] ABI Research. (2013). More Than 30 Billion Devices Will Wirelessly Connect to the Internet of Everything in 2020. [Online]. Available: https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne

[8] Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020. (2013). *Gartner* [Online]. Available: http://www.gartner.com/newsroom/id/2636073

[9] U. Sarwar, G. Rao, Z. Suryady, and R. Khoshdelniat, “Comparative study on available IPv6 platforms for wireless sensor network,” *World Academy of Science, Engineering and Technology*, pp. 889–892, 2010

[10] B. Priyantha, A. Kansal, M. Goraczko, and F. Zhao, “Tiny web services: design and implementation of interoperable and evolvable sensor networks,” In Proc. of the ACM conference on Embedded network sensor system (SenSys), Raleigh, NC, USA, Nov. 2008, pp. 253–266.

[11] P. Levis and D. Gay, “TinyOS Programming,” in *Cambridge University Press*, 2009.

[12] P. Levis and S. Madden. “TinyOS: An operating system for sensor networks,” *Ambient intelligence*, pp. 115–148, 2005.

[13] D. Culler and J. Hil, “Mica: The commercialization of microsensor motes,” *Sensors Magazine*, pp. 40–48. 2002.

[14] J. Polastre, R. Szewczyk, and D. Culler. “Telos: enabling ultra-low power wireless research,” in Proc. of the *ACM IPSN*., 2005.

[15] S. Madden, M. Franklin, et al. “TinyDB: an acquisitional query processing system for sensor networks,” *ACM Transactions on database systems*, pp. 122–173, 2005.

[16] N. Burri, R. Schuler, and R. Wattenhofer, “YETI: A TinyOS plug-in for Eclipse,” in Proc. of the *ACM Workshop on Real-World Wireless Sensor Networks*. 2006.

[17] F. Caron, E Duflos, et al. “GPS/IMU data fusion using multi sensor Kalman filtering: introduction of contextual aspects,” *Information Fusion*, pp. 221–230. 2006.

[18] J. Georgy, A. Noureldin, et al. “Low-cost three-dimensional navigation solution for RISS/GPS integration using mixture particle filter,” *IEEE Transactions on Vehicular Technology*, Vol. 59, No. 2, pp. 599–615, 2010.

[19] K. Chiang, H. Chang, et al. (2009). An artificial neural network embedded position and orientation determination algorithm for low cost MEMS INS/GPS integrated sensors. *Sensors*, Vol. 9, No. 4, pp. 2586–2610.

[20] Y. Bu, L. Chen et al. (2009). Efficient anomaly monitoring over moving object trajectory streams. In *Proceedings of ACM SIGKDD*.
[21] R. Cheng, Y. Zhang, et al. (2006). Preserving user location privacy in mobile data management infrastructures. *Privacy Enhancing Technologies*, pp. 393–412.
[22] D. Liu, M.C. Lee, and D. Wu. (2010). A Node-to-Node Location Verification Method. *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 5, pp. 1526–1537.
[23] N. Tippenhauer, K. Rasmussen, and C. Popper. (2009). Attacks on Public WLAN-based Positioning Systems. *Proc. of ACM MobiSys*, pp. 29–40.
[24] P. Zikopoulos and C. Eaton. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data (1st Ed.)*. McGraw-Hill Osborne Media.
[25] S. Higginbotham. (2010). Sensor Networks Top Social Networks for Big Data. [Online]. Available: http://gigaom.com/2010/09/13/sensor-networks-top-social-networks-for-big-data-2/ [Retrieved: 07 Mar 2014].
[26] M. Nagarajan, K. Gomadam, A.P. Sheth, A. Ranabahu, R. Mutharaju, and A. Jadhav. (2009). Spatio-Temporal-Thematic Analysis of Citizen Sensor Data: Challenges and Experiences. In G. Vossen. D. Long and Jeffrey Xu Yu, ed., 'WISE', pp. 539–553. Springer.
[27] TempoDB. https://tempo-db.com/.
[28] SensorML, OGC Specification (2007). "OpenGIS Sensor Model Language (SensorML) Implementation Specification".
[29] 52North, Sensor Observation Service, http://52north.org/
[30] S. De, T. Elsaleh, P. Barnaghi, and S. Meissner. (2012). An Internet of Things Platform for Real-World and Digital Objects. *Journal of Scalable Computing: Practice and Experience*, Vol. 13, No. 1, 2012.
[31] P. Barnaghi, W. Wang, L. Dong, C. Wang. (2013). A Linked-data Model for Semantic Sensor Streams. In *the 2013 IEEE International Conference on Internet of Things (iThings)*, Beijing, China, 2013.
[32] RDF, Resource Description Framework, http://www.w3.org/RDF/
[33] OWL, Web Ontology Language, http://www.w3.org/2004/OWL/
[34] C. Henson, A. Sheth, and K. Thirunarayan. (2012). Semantic Perception: Converting Sensory Observations to Abstractions. *IEEE Internet Computing*, Vol. 16, No. 2, pp. 26–34.
[35] W. Wang, S. De, G. Cassar, K. and Moessner. (2012). Knowledge Representation in the Internet of Things: Semantic Modelling and its Applications. In *Journal for Control, Measurement, Electronics, Computing and Communications* (special issue on Knowledge Acquisition and Management in the Internet of Things, Automatika), 2012.
[36] P. Barnaghi, W. Wang, C. Henson, and K. Taylor. (2012), Semantics for the Internet of Things: early progress and back to the future. *International Journal on Semantic Web and Information Systems* (special issue on sensor networks, Internet of Things and smart devices), pp.1–21.
[37] D.J. Abadi, A. Marcus, S.R. Madden, and K. Hollenbach. (2007). Scalable Semantic Web Data Management using vertical partitioning. In Proceedings of *VLDB Conference*, 2007.
[38] S. Babu, and J. Widom. (2001). Continuous queries over data streams. In *SIGMOD Rec*., Vol. 30, No. 3, pp. 109–120.
[39] D.C. Luckham. (2001). *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
[40] SQLstream, http://www.sqlstream.com/
[41] StreamBase, http://www.streambase.com/
[42] G. Cugola, and A. Margara. (2012). Processing flows of information: From data stream to complex event processing. *ACM Comput. Surveys*, Vol. 44, No. 3, Article 15.
[43] A. Gray, J. Sadler, O. Kit, K. Kyzirakos, M. Karpathiotakis, J. Calbimonte, K. Page, R. Garcia-Castro, A. Frazer, and I Galpin. (2011). A semantic sensor web for environmental decision support applications. *Sensors*, Vol. 11, No. 9, pp. 8855–8887.
[44] F. Wang, and P. Liu. (2005). Temporal management of RFID data. *VLDB Endowment*, 1128–1139.
[45] E. Della Valle, S. Ceri, F. van Harmelen, and D. Fensel. (2009). It's a Streaming World! Reasoning upon Rapidly Changing Information. *IEEE Intelligent Systems*, Vol. 24, No. 6, pp. 83–89.

[46] A. Bolles, M. Grawunder, and J. Jacobi. (2008)/ Streaming SPARQL - Extending SPARQL to Process Data Streams. In *ESWC2008*, pp. 448–462. Springer-Verlag.
[47] D.F. Barbieri, D. Braga, S. Ceri, E. Della Valle and M. Grossniklaus. (2009). C-SPARQL: SPARQL for continuous querying. In *WWW*, pp. 1061–1062.
[48] W. Yao, C.-H. Chu, and Z. Li. (2011). Leveraging complex event processing for smart hospitals using RFID. *Journal of Network and Computer Applications*, Vol. 34, No. 3, pp. 799–810.
[49] N. Stojanovic, L. Stojanovic, D. Anicic, J. Ma, S. Sen, and R. Stühmer. (2011). Semantic Complex Event Reasoning - Beyond Complex Event Processing. In Dieter Fensel (ed.), *Foundations for the Web of Information and Services*, pp. 253–279. Springer.
[50] D. Anicic, S. Rudolph, P. Fodor, and N. Stojanovic. (2012). Stream reasoning and complex event processing in ETALIS. *Semantic Web*, Vol. 3, No. 4, pp. 397–407.
[51] D. Anicic, P. Fodor, S. Rudolph, and N. Stojanovic. (2011). Ep-sparql: a unified language for event processing and stream reasoning In *WWW2011*, pp. 635–644.
[52] K. Taylor, and L. Leidinger. (2011), Ontology-Driven Complex Event Processing in Heterogeneous Sensor Networks. In *ESWC2011*, pp. 285–299. Springer-Verlag.
[53] S. Hasan, E. Curry, M. Banduk, and S. ORiain. (2011). Toward Situation Awareness for the Semantic Sensor Web: Complex Event Processing with Dynamic Linked Data Enrichment. In *4th International Workshop on Semantic Sensor Networks* (SSN11), page 60–72, 2011.
[54] M. Armbrust, A. Fox, R. Griffith, D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, et al. (2010). A view of cloud computing. *Communications of ACM*, Vol. 53, No. 4 (April 2010), pp. 50–58.
[55] M. Voorsluys, J. Broberg, and R. Buyya. (2011). Introduction to Cloud Computing. In Buyya, R; Broberg, J; and Goscinski (eds.), A Cloud Computing Principles and Paradigms. Wiley Publishing.
[56] P. Mell, and T. Grance. (2011). 'The NIST Definition of Cloud Computing' (800-145). Technical report, National Institute of Standards and Technology (NIST), Gaithersburg, MD.
[57] Amazon EC2, Amazon Elastic Compute Cloud, http://aws.amazon.com/ec2/
[58] Google Cloud Platform, https://cloud.google.com/
[59] Windows Azure, Microsoft Cloud Platform, https://www.windowsazure.com/en-us/
[60] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Comp. Syst.*, Vol. 29, No. 7, pp. 1645–1660.
[61] Y. Wei, K. Sukumar, C. Vecchiola, D. Karunamoorthy, and R. Buyya. (2011); Aneka Cloud Application Platform and Its Integration with Windows Azure. In R. Ranjan, J. Chen, B. Benatallah, L. Wang (eds.*), Cloud Computing: Methodology, Systems, and Applications, 1st ed*. CRC Press, Boca Raton.
[62] E. Sun, X. Zhang, and Z. Li. (2012). The internet of things (IOT) and cloud computing (CC) based tailings dam monitoring and pre-alarm system in mines. *Safety Science*, Vol. 50, Issue 4, April 2012, pp. 811–815.
[63] M. Kovatsch, S. Mayer, and B, Ostermaier. (2012). Moving Application Logic from the Firmware to the Cloud: Towards the Thin Server Architecture for the Internet of Things. In Ilsun You; Leonard Barolli; Antonio Gentile; Hae-Duck Joshua Jeong; Marek R. Ogiela & Fatos Xhafa (ed.), IMIS, pp. 751–756. IEEE Press.
[64] Xively, Platform as a Service, https://xively.com/
[65] SensorCloud, http://www.sensorcloud.com/
[66] ThingSpeak, https://thingspeak.com/
[67] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli. (2012). Fog computing and its role in the internet of things. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing (MCC '12)*. New York, NY, USA, pp. 13–16. ACM Press.
[68] L. Kleinrock. (2003). An Internet vision: the invisible global infrastructure. *Ad Hoc Networks*, Vol. 1, No. 1, pp. 3–11.
[69] S. Islam, and J.-C. Grégoire. (2010). Network Edge Intelligence for the Emerging Next-Generation Internet. *Future Internet*, Vol. 2, No. 4, pp. 603–623.
[70] R. LaMothe. (2013). Edge Computing, Pacific Northwest National Laboratory. [Online]. Available: http://vis.pnnl.gov/pdf/fliers/EdgeComputing.pdf [Retrieved: March 2014].

[71] O. Vermesan, et al. (2011). Internet of Things Strategic Research Roadmap. [Online]. Available: http://internet-of-things-research.eu/pdf/IoT_Cluster_Strategic_Research_Agenda_2011.pdf [Retrieved: April 2011]
[72] L. Atzori, A. Iera, and G. Morabito. (2010). The internet of things: a survey. *Computer Networks*, Vol. 54, pp. 2787–2805.
[73] J.W. Cao, Y. Wan, G. Tu, S. Zhang, A. Xia, and X. Liu. (2013). Information system architecture for smart grids. *Chinese Journal of Computers*, Vol. 36, pp. 144–167.
[74] D. Giusto, A. Iera, G. Morabito, and L. Atzori. (2010). *The Internet of Things*. Springer.
[75] Y. Chang, X. Dong, and W. Sun. (2014). Influence of characteristics of the internet of things on consumer purchase intention. *Social Behavior and Personality*, Vol. 42, No. 2, pp. 321–330.
[76] G. Song, and N. Zhang. (2009). Innovation 2.0: Innovation democratization in the Knowledge Society. *China Soft Science*, Vol. 10, pp. 60–66.
[77] C. Meyer and A. Schwager. (2007). Understanding customer experience. *Harvard Business Review*, Vol. 85, pp. 116–126.
[78] Y. Xu. (2010). Innovation 2.0 promotes healthy IOT development. *Computer World*, Vol. 12, pp. 1–2.
[79] D. Uckelmann, and M. Harrison. (2010). Integrated billing mechanisms in the internet of things to support information sharing and enable new business opportunities. International Journal of RF Technologies: Research and Applications, Vol. 2, No. 2, pp. 73–90.
[80] R. Yan. (2010). Cooperative advertising, pricing strategy and firm performance in the e-marketing age. *Journal of the Academic of Marketing Science*, Vol. 38, No. 4, pp. 510–519.
[81] C.C. Aggarwal, N. Ashish, and A.P. Sheth. (2013). The Internet of Things: A Survey from the Data-Centric Perspective. In Charu C. Aggarwal (ed.), *Managing and Mining Sensor Data*, pp. 383–428. Springer.
[82] D. Borthakur. (2007). The Hadoop Distributed File System: Architecture and Design. The Apache Software Foundation.
[83] J. Dean, and S. Ghemawat. (2004). MapReduce: Simplified Data Processing on Large Clusters. In *'OSDI ' 04: Sixth Symposium on Operating System Design and Implementation*. San Francisco, CA.
[84] K. Steenstrup. (2013). The internet of tings is moving to the mainstream, Keeping good companies. *Business Magazines*, December 2013.