# QoS-Aware Scheduling of Services-Oriented Internet of Things

Ling Li, Shancang Li, and Shanshan Zhao

*Abstract*—**The Internet of Things (IoT) contains a large number of different devices and heterogeneous networks, which make it difficult to satisfy different quality of service (QoS) requirements and achieve rapid services composition and deployment. In addition, some services in service-oriented IoT are required to be reconfigurable and composable for QoS-aware services. This paper proposed a three-layer QoS scheduling model for service-oriented IoT. At application layer, the QoS schedule scheme explores optimal QoS-aware services composition by using the knowledge of each component service. At network layer, the model aims at dealing with scheduling of heterogeneous networks environment; at sensing layer, it deals with the information acquisition and resource allocation scheduling for different services. The proposed QoS-aware scheduling for service-oriented IoT architecture is able to optimize the scheduling performance of IoT network and minimize the resource costs.**

*Index Terms*—**Internet of Things (IoT), quality of service (QoS), wireless sensor networks (WSNs).**

## I. Introduction

WE ARE in an age of witnessing the transformation of society from a physical world into a digitalized world [1], where the objects can be interconnected through Internet of Things (IoT) [2]–[4]. Actually, increasing number of interesting services and applications have been developed by exploiting radio-frequency identification (RFID), wireless sensor/mesh networks (WSNs/WMNs), mobile networks, WLAN, and Internet to improve the efficiency and lower the cost of business processes in enterprise information systems [5]–[9]. Each thing that connects to IoT can be seen as an edge-node of IoT, which is able to sense ambient scenarios and bridge the gap between networks and the real world [10]. The huge number of different links and interactions between edge-nodes in IoT makes it a scalable complex system; therefore, brings difficulties for satisfying the dynamic QoS requirements of services [11], [12].

In service-oriented scenarios, the IoT provides a dynamic service creation environment for new applications, means, and tools to monitor the business process or network parameter changes. It suffers from some common obstacles, such as limited energy, real-time adaptive sensing, and fault tolerance to component failures, which can be described in composite quality of services (QoSs) guarantee support. A number of QoS models have been developed for traditional networks [11], [13]–[16]. However, the QoS management in IoT is still poorly studied [14], [17], [18]. The definition of QoS in IoT is still not clear because the definition of service in IoT is not exactly the same, in which a service can be defined as the simple acquisition and processing of information and the decision making process in identification, communication, and so on [19]. The traditional QoS attributes such as throughput, delay, or jitter are evidently inappropriate in IoT [11]–[20]. In IoT, more QoS attributes are concerned, such as *information accuracy* (that is qualified with the probability that an accuracy can be reached), the *network resources* needed, *required energy consumption*, and the *coverage* of IoT. To solve the difficulties mentioned above, a new QoS model for service-oriented IoT is necessary, which must be able to balance the network availability with the quality of information in delivering sensory data [21]–[23]. On the other hand, the information fitness or information accuracy for a service is the key to merging the gap between sensory data and actual world [24], which involves information acquisition, processing, transmission, and energy consumption.

In this paper, we propose a three-layer QoS model for service-oriented IoT, which requires the following features.

1) QoS provisioning in IoT is investigated in terms of decision-making over the application layer, the network layer, and the sensing layer.
2) The QoS model based on the above-mentioned three-layer service-oriented architecture is proposed, in which a top-down decision-making process is proposed for QoS guarantee in IoT through qualitative methods.
3) The QoS analysis framework is given to evaluate services, networks, and sensing devices, which is able to guide the optimization of attributes of QoS in IoT.
4) The decision-making process that includes the optimization algorithms for three layers is provided.

The rest of this paper is organized as follows. Section II summarizes related works. Section III discusses the QoS architecture of IoT. Section IV provides a decision-making model for QoS in service-oriented IoT. Section V verifies the feasibility and effectiveness of proposed mechanisms. Section IV concludes the paper.

## II. Related Works

Although a lot of research results have been reported in the past decade [12]–[25], most research works focus on the QoS support for traditional networks [22]–[26]. Little research efforts

Manuscript received September 22, 2013; revised January 3, 2014 and January 28, 2014; accepted January 29, 2014. Date of publication February 17, 2014; date of current version May 02, 2014. Paper no. TII-13-0655.

L. Li is with the Department of Information Technology/Decision Sciences, Old Dominion University, Norfolk, VA 23529 USA (e-mail: lli@odu.edu).

S. Li is with the Faculty of Engineering, University of Bristol, Bristol BS8 1TR, U.K. (e-mail: shancang.li@bristol.ac.uk).

S. Zhao is with the School of Computing, University of the West of Scotland, Paisley PA1 2BE, U.K. (e-mail: sszhao.uk@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2014.2306782

are found over service-oriented IoT although it provides an exciting and promising vision for seamlessly connecting the virtual world of information to the real world [26].

QoS in service-oriented IoT needs to consider the new features of IoT. The successful implementation of IoT might involve a number of emerging techniques, such as RFID, WSN, mobile technologies, WLAN, and the QoS provisioning on application layer. The WSN is a suitable example of edge-network of IoT and many research works about QoS in WSNs are available. In [16], [17], [27], and [28], the QoS requirements for WSN are split into a number of subtasks at media access control (MAC) layer and network layer. An M-QoMoR protocol was proposed to analyze the QoS performances of group-organized nodes [16], [17]. Actually, it is an extension of traditional QoS framework without considering the resource limitation in IoT. In [18], the dynamic services support in the IoT environment is studied. In [20], Rohokale *et al.* discussed a cooperative QoS scheme for healthcare system, in which the QoS constraints healthcare conditions metrics. In [19], a multidimensional QoS decision model is proposed, where each stakeholder is mapped into a single value that is used to determine the best configuration of interactions.

Most of the existing researches in IoT focus on two aspects: 1) QoS-based design for routing or MAC schemes, i.e., QoS design in the service-oriented WSNs [20]–[22], cross-layer QoS-aware routing [22], and energy-efficient routing approach [23] and 2) QoS supported for multimedia traffic flows, such as the end-to-end *delay, jitter*, and *throughput* [24]. Many QoS schemes have been proposed to solve real-time or best-effort (BE) scheduling tasks. By doing this, the performances can be slightly improved but QoS problems cannot be solved. It is needed to redesign a QoS-aware scheduling scheme in service-oriented IoT according to the new features.

## III. QoS Architecture of IoT

### A. Composite QoS in IoT

The research on architecture is essential in the development of IoT that involves a huge number of heterogonous devices, complex data, and actions such as collection, processing, transmission, and storage.

Recently, many cross-layer approaches and resource limitation in IoT design have been reported that motivate us to reconsider the QoS architectures in traditional networks. In [27], a new concept, sensing as an independent service (*S2aaS*), is proposed, which enables IoT to provide sensing services using mobile devices via cloud computing. In the IoT, the platform and protocol used in heterogeneous networks are significantly depends the requirements of QoS, where the QoS attributes, such as networks availability, coverage, available bandwidth, information accuracy, and so on, span a new design space of QoS. Although a lot of QoS schemes and architecture frameworks have proposed for existing heterogeneous networks, some of the principal operational characteristics of the new wireless networks, such as WSNs, WMNs, RFID, and ad hoc networks are still only poorly understood in IoT. It is still difficult to define the "*quality*" of the services and the "*lifetime*" of networks in IoT. The traditional QoS attributes are evidently inappropriate in IoT, where the services involve many attributes such as "*information accuracy*" and "*services availability*,". Fig. 1 shows a basic architecture of IoT that considers the new features of service-oriented IoT.

The QoS in IoT investigates an integrate treatment of the *information accuracy*, *coverage*, *energy consumption*, *the cost of a network deployment*, and traditional QoS attributes. It is a challenging task to prove the relations among these attributes mentioned above, but characterizing principle behavior is an achievable study objective with great practical consequences.

### B. QoS Architecture in IoT

As a heterogeneous complex system, in IoT, multiple communication techniques might be combined to provide continuous sensory data acquisition, transmission, processing, and storage. An IoT edge-node (e.g., RFID tag or sensor node) collects information and transmits over network layer to the service users/providers, where resource is reserved for services in a proactive manner. By optimizing services quality, network services, and end-node performance, the entire cost of a service cost is minimized.

To achieve the optimal services quality, network services, and edge-node performance is very difficult. It can be modeled with a stochastic programming process by using the randomness of services in *application layer*. The QoS models should be able to schedule the traffic from multiple sources, which depend on QoS requirements over available connections. In our three-layered QoS architecture, we model the QoS optimization problem as a Makov decision process (MDP) in application layer, which targets to maximize the services quality. In network layer, a programming method is used to minimize the connection cost and optimal network attributes. In sensing layer, the optimization object is to obtain the optimal sensing cost and usage of edge-node. This architecture for QoS decision support process is useful to provide flexible and cost-effective services to users/providers, as shown in Fig. 2. More specifically, QoS attributes are scheduled over three layers.

1) At the *application layer*, an application is selected to establish a connection, and the decisions are made by the user and the QoS scheduling engine. In general, QoS module must allocate network resource to services that are selected in the application layer. In some cases, the services may require substantial resources and will affect the coming services and users; therefore, it is important to properly make a decision for services selection and resources allocation.
2) At the *network layer*, the QoS module needs to allocate the network resources to the selected services. The decision-making process at this layer may involve QoS attributes that are used in traditional QoS mechanisms over networks, such as cellular, RFID, WSN, and so on.
3) At the *sensing layer*, the decision-making process involves selection of basic sensing infrastructure based on sensing ability and the required QoS for application/user. The QoS module at sensing layer is responsible for the selection of basic sensing devices. The decision-making process in this layer is expected to reduce the overlapping of sensing and data redundance and lower the energy consumption.

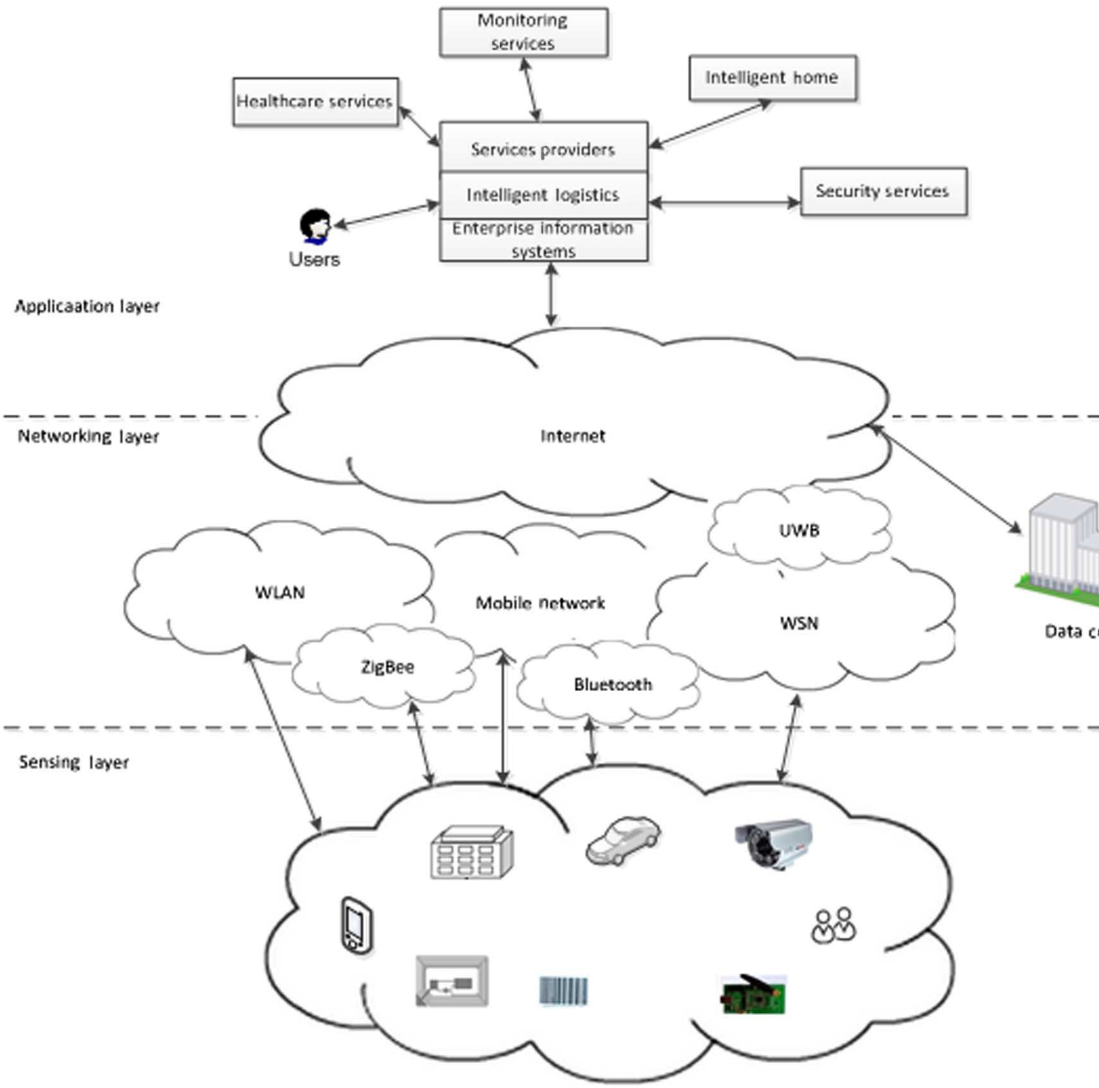



Fig. 1. Basic architecture of IoT [29].

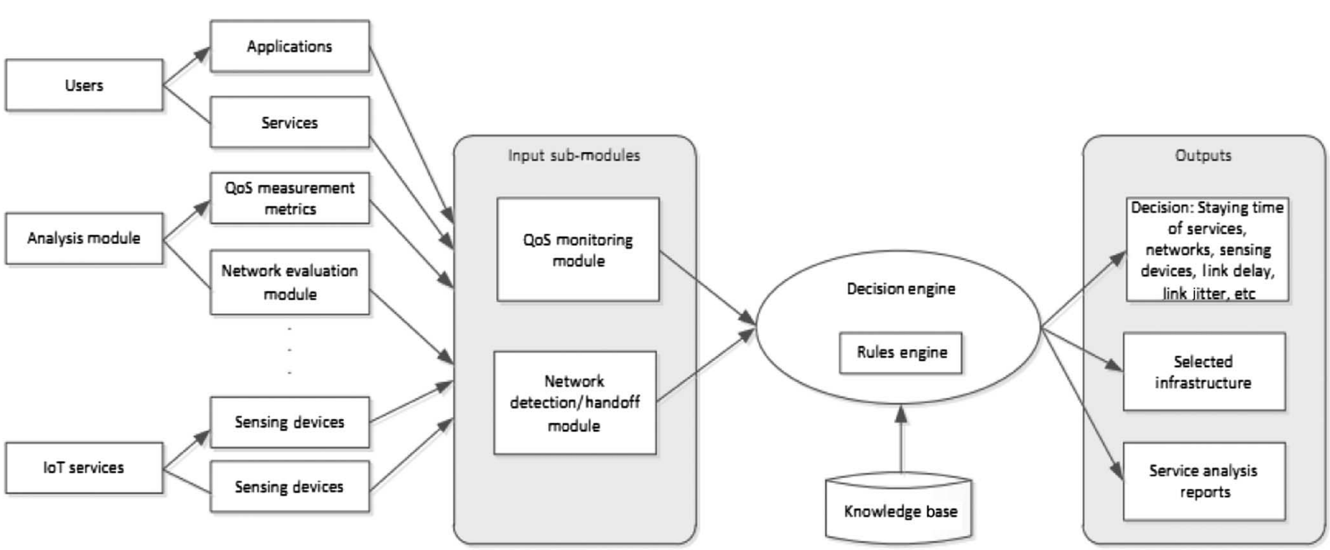



Fig. 2. Decision-making for QoS in IoT.

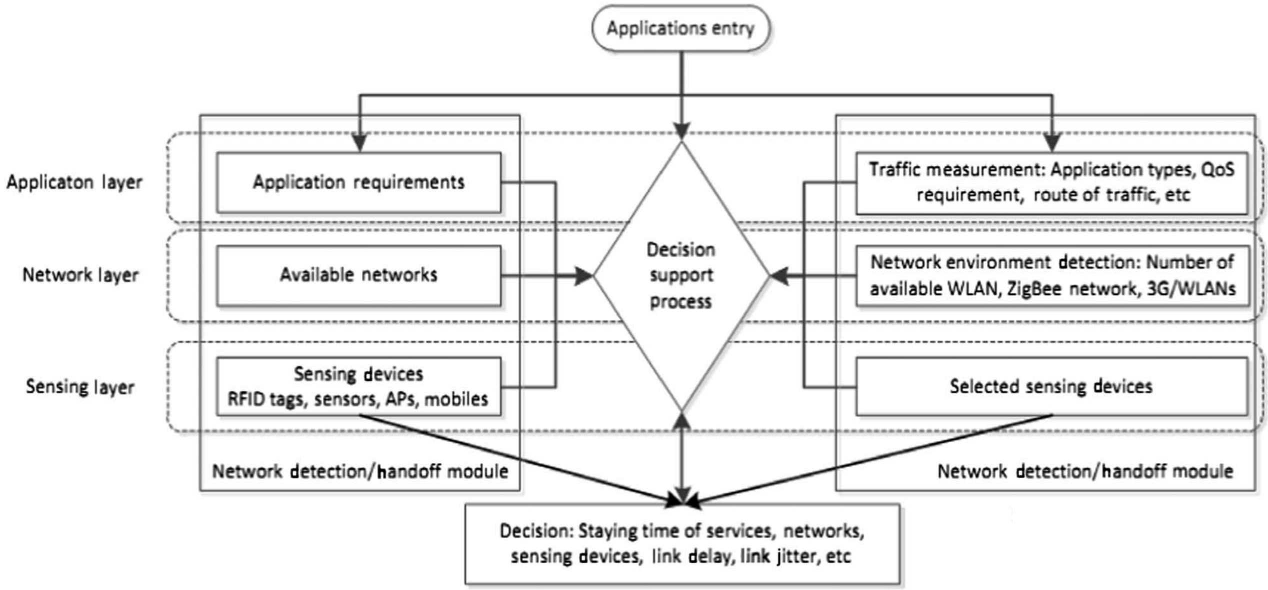



Fig. 3. QoS decision in network layer of IoT.

*1) QoS in Application Layer:* For services in IoT, specific QoS requirements may be requested, as well as the commonly used QoS metrics such as *throughput, bandwidth, delay, and performance* of the protocols in use. Many existing QoS models for traditional Internet are available for this layer, such as differentiated services (DiffServ) [22], flexible QoS model for ad-hoc networks (FQMM) [23], and integrated mobile ad-hoc network (MANET) QoS (iMAQ) model [24]. In IoT, we investigate the composite QoS in the service-oriented environment. The existing QoS models did not consider the inherent stochastic and dynamic nature of IoT and may cause service failure. The QoS in application layer focuses on reliable service composition algorithms based on Markov decision, which can significantly improve success rate of IoT service.

At the application layer, an application is selected to establish a connection, and the decisions are made by the QoS schedule engine. The main functionalities of this layer are as follows: 1) efficiently schedule the services according to the QoS and 2) transfer the resource requests to the QoS module in network layer of IoT.

*2) QoS in Network Layer:* Fig. 3 shows that the QoS module at network layer of IoT contains the allocation of network resources. The recent trends indicate that services in IoT may require the coexistence of heterogeneous networks (especially for wireless networks) to provide universal coverage and network access due to their different characteristics. However, the different sensing devices and information that acquired by them becomes increasingly difficult to access and make use of over such huge networks. In IoT, services can be provided over multiple access network nodes within the same network layer technology. The service providers can provide the reserved network connections for its users to support the applications that users required. Seamless and efficient QoS support from different infrastructures is a challenging problem in the applications development toward the IoT.

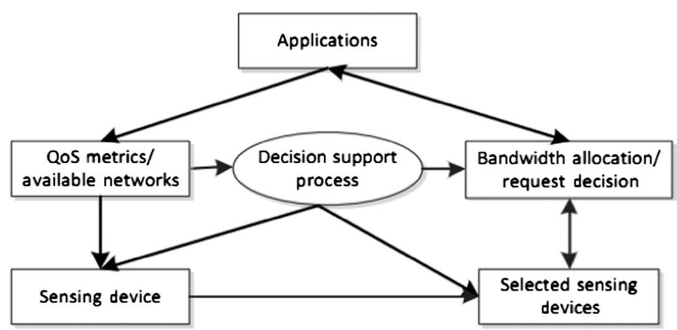



Fig. 4. QoS decision in sensing layer of IoT.

*3) QoS in Sensing Layer:* It is important for IoT to measure the QoS attributes provided by IoT, which is much different with that of traditional networks. In our works, we defined the services in IoT by three levels: 1) application level; 2) network level; and 3) sensing level, as shown in Fig. 4. The services in application level and network level are same with the traditional services, which mean that the QoS support in application level and network level can be measured with traditional QoS attributes. However, the sensing level is the physical infrastructure of IoT, which includes nodes with different resources, such as powerful data process centre, sensor networks, RFID, mobile devices, and other heterogeneous devices. The availability of services in this layer determines the success or failure of a service request.

### C. QoS Metrics in IoT

In QoS framework design, it is essential to measure the QoS attributes that are provided by IoT, where services are much different with that in traditional networks. Similar to the QoS optimization in tradition networks, application layer and network layer can evaluate service quality with existing QoS attributes.

However, at the sensing layer, the QoS metrics involve the resources allocation and information process at diverse edge-nodes of IoT, such as information accuracy and availability of services is determined by the sensing layer.

*1) The Application Layer:* The QoS metrics used in IoT are similar to that of existing QoS in web services, where the following metrics can be used.

1) *Services perform cost* denotes the cost to invoke a service and is provided by the service provider, and it is a constant in a service round.
2) *Perform time* denotes the perform time from accepting of service to the finishing of it.
3) *Load* is used to evaluate the service of real-time usage.
4) *Reliability* denotes that the reliability of service can be obtained to the user.

*2) The Network Layer:* In network layer, the network traffic can be classified into two categories: 1) delay-sensitive service, such as real-time applications; and 2) BE class service, such as data-based applications. In this level, the following metrics might be involved.

1) *Bandwidth allocation* denotes the bandwidth that can be allocated to connections provided for a service.
2) The *capacity of network layer*, which aims to achieve maximum throughput under QoS constraints.
3) *Allocation rate* by taking the delay, throughput, and jitter into consideration.
4) *Power efficient strategies* and *throughput optimization* over the network layers.

*3) The Sensing Layer:* In IoT, the availability of services in application layer depend the support of both network layer and sensing layer. In sensing layer, when multiple devices are available that can serve for a service request, a decision process is needed to determine the selection of sensing devices. The services in service-oriented IoT may subject to compromise of information by low sensing quality and noise. As a result, the accuracy of acquired information is a metric to measure the discrepancy between the real world and the provided result. In most sensing-based applications, a probability with which a given accuracy can be achieved is related with services, which needs to make a correct/optimal decision. On the other hand, power consumption at sensing nodes and cost are two important factors of QoS in sensing layer. Energy/power is related to the lifetime of sensing networks, and the cost is influenced by the deployment of devices.

In this study, we use the *information accuracy*, *sensing precision*, *energy consumption, life-time* of sensing networks, and *cost* to measure the QoS of sensing networks, as well as a four-tuple *(accuracy, precision, energy/power, lifetime, and cost)*.

### D. QoS Monitoring Modules

It is necessary to automatically monitor the attributes of QoS in IoT. When a service arrives, it might be divided into multiple subtasks at application layer that can be executed over network layer and then further sent to the sensing layer. In network layer, the scheduler decides the network type and sensing layer devices, i.e., for data acquisition based subtasks, the scheduler will arrange the proper RFID or sensor network to execute it, the policy can be predefined before tasks distribution. At the sensing level, the QoS manager needs to optimize the cost for fixed budget, i.e., selection of nodes, power consumptions, connectivity, size of largest component in a graph, etc. A decision-making process is needed to trade-off the cost-quality density. The main functionalities of QoS monitoring modules include:

1) decomposing the QoS requirements of arriving applications into requirements for individual services used in the services;
2) synthesizing descriptions of the QoS attributes among the three layers;
3) compiling the synthesized descriptions into executable subtasks.

## IV. Decision-Making Process for QoS

The decisions-making in application, network, and sensing layers of IoT must be made to support the QoS requirements for different services. At the application layer, an application may contain one or more services, and the decisions are made by the application layer that depends the QoS requirements of service(s). The QoS scheduler needs to allocate resources to the services according to the QoS requirements. In some cases, the arriving services may require some substantial resources that affect the number of subtasks that can serve by the system. The decision-making at network layer is similar to that of the traditional QoS guarantee in existing networks, including the human factors (such as stability of service, users, delays, and so on) and technical factors (such as reliability and scalability). The decision-making at the sensing layer involves information accuracy, energy consumptions for selected networks, such as WSN, RFID, Wireless Mobile Networks or WLAN, based on the availability of resource and requirements of QoS. The decision-making at sensing layer also deals with how much users or area have been covered by the IoT, and how much cost is used for a service.

### A. QoS Decision-Making Process in Application Layer

In service-oriented IoT, an application from user is always involved to multiple services or composition of multiple services to fulfil the goal. In QoS design for IoT, the service request arrives randomly, which causes the uncertainty of services, and the inherent stochastic and dynamic nature of services should be considered. Actually, it is difficult to dynamically update the QoS requirements for real-time-based applications. In the QoS decision process at application layer, a Markov decision process (MDP) is used to model the services QoS, where some new metrics are introduced.

*1) Measurement Metrics at Application Layer of QoS:* For the services in application layer, many metrics are available for measurement of QoS. The service cost, service time, service load, the reliability, and the reputation are the most important factors, which are defined as follows (let $\xi$ denote a service in IoT).

1) *Service Cost* $C(\xi)$ is the cost for the services provided by the end service provider and is a constant in IoT.
2) *Service Time* $S(\xi)$ denotes the service time from the acceptance of the request to finishing of the task. $S(\xi)$

depends the process capability and the network resource allocation. Here, we assume that $k$ services IoT nodes batch process as a $M/M/k$ queue model to anticipate the $S(\xi)$ as $\mathbb{E}(S(\xi)) = (\rho\pi_k)/((1-\rho)^2\lambda + 1/\mu)$.

3) *Load* $\mathbb{L}(\xi)$ is used to measure the usage of service in IoT, which depends on the request arrival rate $\lambda$ and service rate $\mu$ as $\mathbb{L}(\xi) = \lambda\mu$.
4) *Reliability* $A(\xi)$ denotes the probability of failure of services on client nodes; here, we modeled it as a random variable with $E(A) = \sum_{(i=1)}^{n} a_i/n$ and $D$.
5) *Reputation* $R(\xi)$ denotes the trust level of service at IoT nodes, the higher the $R(\xi)$ for service $\xi$ is, the better reputation the IoT node hold.

*2) QoS Architecture:* In IoT, the QoS manager records the communications between tasks and services; by doing this, QoS manager is able to monitor the QoS metrics that can be provided by IoT. Assume that $X$ denotes the QoS metrics results, QoS manager may adaptively update the metrics with a confidence level of $\alpha$, which depends the $\mathbb{E}(X)$ and $\mathbb{D}(X)$. The confidence interval is defined as $(\mathbb{E}(X) - \sqrt{(\mathbb{D}(X)/(1-\alpha))}$, $\mathbb{E}(X) + \sqrt{(\mathbb{D}(X)/(1-\alpha))})$.

*3) MDP Model for Tasks:* We define *MDP* as a five-tuple $(S, A, P, \mathrm{Re})$, in which $S$ is the set of states, $A$ is the actions set. $P$ denotes the probability function, which describes the changes of state when one or more actions are executed $(P :\in [0,1])$. $\mathrm{Re}(s,a)$ denotes the reward function for status changes from $s$ by action $a$, and we have $\mathrm{Re} : S \times A \to \mathrm{Re}$.

Here, we use the expected sum of rewards to represent the quality of a policy. Actually, it can be discounted to ensure that the $\mathbb{E}(\mathrm{Re})$ can converge to a finite value. The sequence of states under policy $\pi$ can be represented as $V^{\pi}(s)$, which can be obtained by (1)

$$V(S) = \underset{a\in A}{\operatorname{argmax}}\left(\mathrm{Re}(s,a) + \gamma\sum_{s'\in S} p(s,a,s')V^{\pi}(s)\right). \quad (1)$$

Since we have

$$\mathrm{Re}(s,a) = \sum_{s'\in S}\mathrm{Re}(s,a) \times p(s,a,s'). \quad (2)$$

It is noted that an optimal policy can be derived from (1), which can be further optimized as

$$V(S) = \max_{a\in A}\left(\mathrm{Re}(s,a) + \gamma\sum_{s'\in S} p(s,a,s')V^{\pi}(s)\right) \quad (3)$$

where $V_k$ converges to the optimal policy value when $k \to \infty$.

Let three-tuple $(t, h, p)$ denote the state of an IoT service, in which $t \in \{t_1, t_2, \ldots, t_T\}$ is the operating time of network, $h$ is a value that can be used to measure the service state in previous time-step, and $p$ is the amount of energy consumed.

The transition probability from state $s_i = (t_i, h_i, p_i)$ to $s_j = (t_j, h_j, p_j)$, when action $a \in A$ operated is $p(s_i, a, s_j)$, can be exactly defined by probability of *state model, bandwidth, and energy consumption* as $p_T(t_i,t_j) \cdot p_B(b_i,b_j) \cdot p_P(p_i,a,p_j)$, where

$$p_T(t_i,t_j) = \begin{cases} 1, & \text{if } i=j=T \\ -1, & \text{if } j=i+1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

is the state model and

$$p_B(b_i,b_j) = \begin{cases} p_B^{\text{same}}, & \text{if } i=j; \\ 2p_B^{\text{chage}}, & \text{if } i=B, j=2B; \\ p_B^{\text{chage}}, & \text{if } |i-j|=1; \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

models the bandwidth change in IoT. The energy consumption can also be modeled with probabilities $P(a)$ when an action $a \in A$ is executed.

$$p_P(p_i,a,p_j) = \begin{cases} 1, & \text{if } i=P; \\ p_P(a), & \text{if } i=j; \\ 1-p_P(a), & \text{if } i+1=j; \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The reward function is determined only on the change rate and available bandwidth that is reported by node. Here, we use a penalty to model this process

$$Re((t,b,p),a) = \begin{cases} -Re^{\text{powerout}}, & \text{if } p=P, t<T; \\ k_R \cdot a \cdot b, & \text{otherwise} \end{cases} \quad (7)$$

in which $k_R$ is a constant of proportionality.

Assume that each task at each node can only be run once for a service, then the (6) can be easily solved with an iteration algorithm. This problem can be initialized with arbitrary values as $V_0(s)$, then it can be updated at iteration $k > 0$ as follows:

$$V_k(S) = \max_{a\in A}\left(\mathrm{Re}(s,a) + \gamma\sum_{s'\in S} p(s,a,s')V_{k-1}(s)\right). \quad (8)$$

As mentioned above, the reward function $\mathrm{Re}(s,a)$ is defined as

$$\mathrm{Re}(s,a) = \sum_{s'\in S}\mathrm{Re}(s,a,s')p(s,a,s'). \quad (9)$$

Then, the optimal function to select the services can be defined as

$$V_k(s) = \sum_{a\in A} p(s,a,s')[\mathrm{Re}(s,a) + \gamma V_{k-1}(s')]. \quad (10)$$

For services $\xi$, the optimal policy is $f^*(s) = \{v^*(s_1), v^*(s_2), \ldots, v^*(s_N)\}$.

---

**Algorithm 1:** QoS-aware services iteration algorithm

---

**Input:** Initialized $V_0(s)$, $t = N$, $\gamma$, $v_0$, $e$.

**Output:** $V^*(s)$

**repeat**

**repeat**

For $s_t \in S$, calculate $\sum_{s_{t-1} \in S} R(s_t, a, s_{t-1})$;

Calculate $p(s_t, a, s_{t-1}) = p_T(s_t, s_{t-1}) \cdot p_B(s_t, s_{t-1}) \cdot p_T(p_t, a, p_{t-1})$;

Calculate $R(s_t, a) = \sum_{s_{t-1} \in S} R(s_t, a, s_{t-1}) p(s_t, a, s_{t-1})$;

$v_k(s) \leftarrow \max_{a \in A} \sum_{s_{t-1} \in S} p(s_t, a, s_{t-1}) \; [R(s_t, a) + \gamma v_{k-1}(s_{t-1})]$;

$k \leftarrow k + 1$;

**until** $k > K$ *or* $\|v_k - v_{k-1}\| \leq e$;

$v^*(s_t) \leftarrow \arg\max_{a \in A}$;

$t \leftarrow t + 1$;

**until** $t > T$;

$f^*(s) = \{v^*(s_1), v^*(s_2), \ldots, v^*(s_N)\}$.

### *B. Decision Model in Network Layer*

In network layer, the network traffic can be classified into two classes depending the types of service: 1) *delay or jitter* sensitive service that has an expectation to deliver on time, such as real-time-based services and 2) *BE* class service that includes all nondetrimental services, such as peer-to-peer applications. The sensitive service requires the network layer to provide connections with a minimum transmission rate as $R_{\min}^{(i)}$ for $i = 1, \ldots, K$.

The BE-based services should be able to achieve optimal proportional fairness. Each edge-node in sensing layer obtains the resource from the available access networks to the networked layer. For nodes in sensing layer $n_j (j = 1, \ldots, N)$ and nodes in network layer $m_i (i = 1, \ldots, M)$, the bandwidth allocation function and the total bandwidth and power consumption are denoted as $B_{ij}$ and $P_{ij}$, respectively. Then the throughput at node $n_i$ can be modeled as

$$r_i = \sum_{j=1}^{N} (1 - \eta_{ij}) \beta_j x_{ij} \log_2 \left( 1 + \frac{g_{ij} p_{ij}}{x_{ij}} \right) \tag{11}$$

where $\eta_{ij}$ denotes the average bit error rate (BER) over link $(i, j)$; $\beta_j$ denotes that the system efficiency, which is used to guarantee the node $n_j$, can access all $s_i$; $x_{ij}$ is the bandwidth allocated to link $l_{ij}$; $p_{ij}$ denote the transmission power of node $m_i$; and $g_{ij}$ is the channel gain.

It is clear that for $i = 1, 2, \ldots, K$,

$$r_i \geq R_{\min}^{(i)}. \tag{12}$$

In an IoT that includes heterogeneous nodes, the system capacity maximization with QoS support problem can be modeled as

$$\max \boldsymbol{R}(\mathbf{x}, \mathbf{p}) = \max \sum_{i}^{M} \sum_{j}^{N} (1 - \eta_{ij}) \beta_j x_{ij} \log_2 \left( 1 + \frac{g_{ij} p_{ij}}{x_{ij}} \right) \tag{13a}$$

$$\text{s.t. } x_{ij} \leq X_j \quad \forall j \tag{13b}$$

$$p_{ij} \leq P_i \quad \forall j \tag{13c}$$

$$\sum_{j}^{N} (1 - \eta_{ij}) \beta_j x_{ij} \log_2 \left( 1 + \frac{g_{ij} p_{ij}}{x_{ij}} \right) \geq R_{\min}^{(i)} \tag{13d}$$

$$p_i = p_{i+1} \quad \forall i = K + 1, \ldots, M - 1 \tag{13e}$$

in which $X_j$ denotes the total bandwidth at node $n_j$, and $P_i$ is the maximum available power of node $m_i$. It is clear that (13) can be solved using convex optimization method, in which the total bandwidth at each nodes in both network layer and sensing layer can be obtained.

### *C. Decision-Making Process for QoS in Sensing Layer*

The characteristics of sensing layer include autonomous (in which some edge-nodes are densely and randomly deployed in an autonomous manner), limited resources (some edge-nodes of IoT have limited resources and power supply, and are typically small in size), susceptible (the physical environment in which the sensing layer is deployed in out-doors and can be easily damaged), and specific data-centric (the services supported by IoT applications).

As mentioned above, in QoS design at the sensing layer following factors should be considered: 1) the information accuracy, which is very different from the traditional QoS requirement due to the information in IoT suffers from information compromise; 2) the links with high bandwidth and low cost have priority for a service; 3) the number of connections for services should be minimized and the communication load over current communications should be optimized; and 4) coverage of IoT, it depends the deployment of edge-nodes in IoT.

In this work, we highlight the following issues, information accuracy, energy efficiency, adaptability, and coverage. In sensing layer, the QoS performance metrics significantly differ from the existing QoS mechanisms and the key metrics that should be considered in sensing layer are as follows.

1) *Information accuracy*: Each node might be deployed separately to efficiently sense interested data. It includes the data accuracy, sensing time accuracy, and spatial accuracy. In order to reduce the data redundancy and transmission burden, data collected from different IoT edge-nodes might be packed into a packet which then will be forwarded to up layers.
2) *Energy consumption*: In IoT, the energy limitation of end sensing nodes is of crucial importance to the lifetime of the sensing nodes.
3) *Coverage*: The diversity of applications in IoT might led to interpretation of IoT coverage which targets to sure each sensing region in within the sensing range of at least one sensor node.

The *information accuracy* can be quantified with probability: a larger probability means a higher accuracy. Let random variable $\mathbf{x} \in R^N$ denotes the information that including $N$ status, the

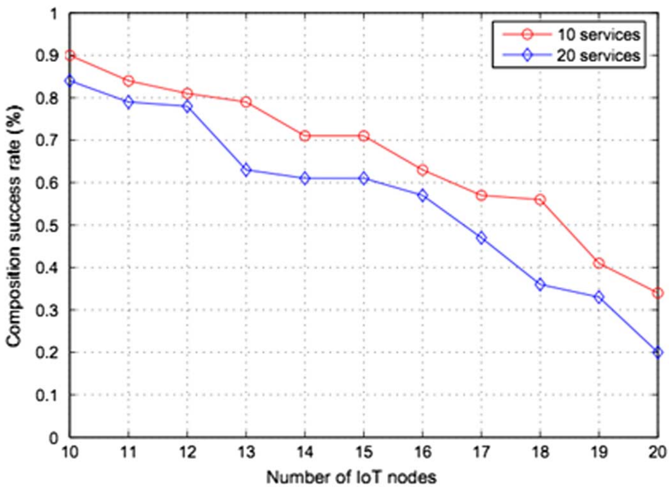



Fig. 5. Service composition success rate.

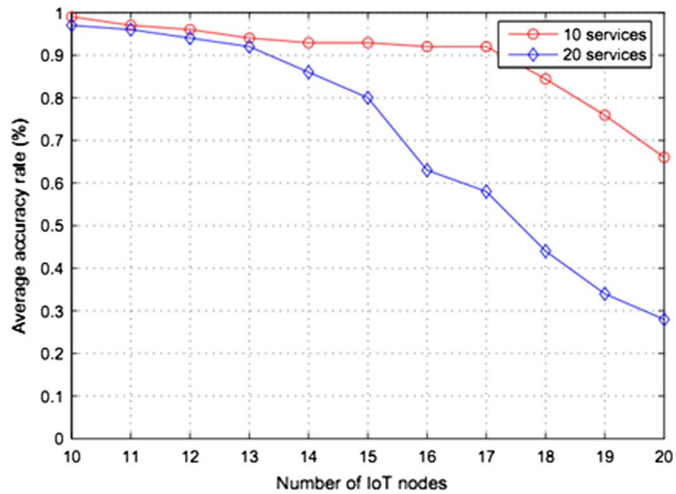



Fig. 6. Average accuracy rate.

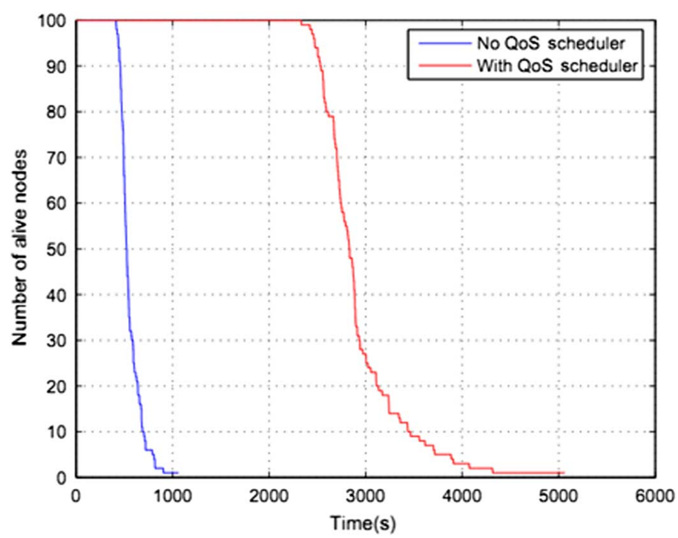



Fig. 7. Lifetime of IoT network with 100 nodes.

information accuracy of acquired information $\mathbf{x}'$ can be represented with the qualified closeness to the actual status.

$$q = P(\mathbf{x}' = s_i | \mathbf{x} = s_i) \tag{14}$$

in which $i = 1, 2, \ldots, N$. It is clear that $\theta = \frac{1-q}{N-1}$ denotes the wrong probability that $s_i$ to be $s_j$. Assume that the acquired information $\boldsymbol{y}$ has

$$P(\boldsymbol{y}_i) = \sum_{j=1} P(\mathbf{x}_i)(\boldsymbol{y}_i|\mathbf{x}_i) = P_i \cdot q + \theta(1 - P_i). \tag{15}$$

The energy consumption at an edge-node $i$ can be estimated as

$$E_i = E_i^R + E_i^T = (\lambda_i^r + \lambda_i^t) \cdot b_i + \gamma \cdot b_i \cdot d_{(i,j)}^m \tag{16}$$

where $\lambda_i^r$ and $\lambda_i^t$ denote the receive energy coefficient and transmit energy coefficient at node $i$, respectively. $b$ denotes the traffic bit-rate, $d_{(i,j)}$ denotes the distance of link to node $j$, and $m$ is the *path-loss-exponent.* The coverage of node $i$ is difficult to define in IoT, since in IoT the coverage may involve sensing area or business events, in this model we use $C_i^{\text{cov}}$ to represent the coverage level.

Therefore, an optimal QoS requirement needs to maximize all the three parameters, and it can be described as

$$\arg\max \left( \frac{1}{\sum_{i=1}^{N} E_i}, P(\mathbf{y}|\mathbf{x}), C^{\text{cov}} \right). \tag{17}$$

## V. Simulation

In order to verify the proposed QoS mechanism, in this section we simulate the QoS scheduling for multiple services in an IoT with 100 sensor nodes. The service requests are randomly arrived according to a normal distribution, and the simulation environment is as follows: *Intel i*3 CPU 2.3 GHz, 4*G* RAM, *Windows* 7, *and MATLAB* 2009*R*. To evaluate the QoS service scheduling ability in application layer, we simulate 10 and 20 services in an IoT environment. Fig. 5 depicts the mean values for the services that successfully served rate when the number of IoT nodes varies from 10 to 20, and a total of 100 trails are simulated for each setting. It can be seen that the proposed scheme can effectively improve the service accurate. In Fig. 6, we investigate the *information accuracy* when a data acquisition application is served in IoT. It is shown that the error between the acquired data and the real data remains near to 5% when the number of IoT nodes is small than 13 for both 10 and 20 services are scheduled, but increases quickly when the services number increases beyond 17. In Fig. 7, we investigate the energy consumption of IoT network with 100 nodes. When 10 data acquisition services are invoked, the proposed decision scheme can significantly reduce the energy consumption at IoT nodes. When no QoS scheduler is performed, the first dead node occurred at time $t = 415$, meanwhile the first dead node occurrence time is $t = 2334$ when QoS scheduler is performed. The lifetime of IoT is significantly prolonged.

Meanwhile, the energy consumption and network costs varies slightly as the services number, which suggests that the QoS in network layer and sensing layer tend to provide optimal solutions.

## VI. Conclusion

It is foreseeable that the IoT will be a part of future internet where "things" can be wirelessly organized as a global network that can provide dynamic services for applications and users. IoT is able to bridge the gap between the virtual network and the real things world. As reported in [26], [27], the edge-nodes of service-oriented IoT may be able to interact and communication with the environment and work cooperatively in business, information, and social process. This works focus on the QoS support for the diverse applications in IoT. We firstly investigated the existing QoS mechanisms in different networks that are involved in IoT, and then examined the constraints in IoT. We also proposed a three-layer architecture which aims to enhance the QoS in IoT. This framework includes the application layer, network layer, and the sensing layer. The QoS optimization algorithms also have been given, depending on the specific service requirements.

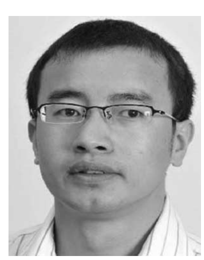
## References

[1] L. D. Xu, "Enterprise systems: State-of-the-art and future trends," *IEEE Trans. Ind. Informat.*, vol. 7, no. 4, pp. 630–640, Sep. 2011.
[2] O. Vermesan, M. Harrison, H. Vogt, K. Kalaboukas, and M. Tomasella *et al.* (Eds.) "The internet of things—Strategic research roadmap," in *Cluster of European Research Projects on the Internet of Things, CERP-IoT,* 2009.
[3] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, and A. Bassi *et al.*, *Internet of Things: Global Technological and Societal Trends.* Aalborg. Denmark: River Publishers, pp. 9–52, May 2011.
[4] F. Xia, L. T. Yang, L. Wang, and A. Vinel, "Internet of things," *Int. J. Commun. Syst.*, vol. 25, no. 9, pp. 1101–1102, Sep. 2012.
[5] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, "Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things," *J. Netw. Comput. Appl.*, vol. 36, no. 6, pp. 1531–1539, Nov. 2013.
[6] J. Leu, C. Chen, and K. Hsu, "Improving heterogeneous SOA-based IoT message stability by shortest processing time scheduling," *IEEE Trans. Serv. Comput.*, to be published, doi: 10.1109/TSC.2013.30.
[7] Y. Ding, Y. Jin, L. Ren, and K. Hao, "An intelligent self-organization scheme for the internet of things," *IEEE Comput. Intell. Mag.*, vol. 8, no. 3, pp. 41–53, Aug. 2013.
[8] P. Vlacheas, R. Giaffreda, V. Stavroulaki, D. Kelaidonis, V. Foteinos, G. Poulios *et al.*, "Enabling smart cities through a cognitive management framework for the internet of things," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 102–111, Jun. 2013.
[9] M. T. Lazarescu, "Design of a WSN platform for long-term environmental monitoring for IoT applications," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 3, no. 1, pp. 45–54, Mar. 2013.
[10] D. F. L. Filho and J. R. Amazonas, "TCNet: Trellis coded network—Implementation of QoS-aware routing protocols in WSNs," *IEEE Latin Amer. Trans.*, vol. 11, no. 3, pp. 969–974, May 2013.
[11] M. Liu, Z. Li, X. Guo, and E. Dutkiewicz, "Performance analysis and optimization of handoff algorithms in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 7, pp. 846–858, Jul. 2008.
[12] S. Hodges, S. Taylor, N. Villar, J. Scott, D. Bial, and P. T. Fischer, "Prototyping connected devices for the internet of things," *Computer*, vol. 46, no. 2, pp. 26–34, Feb. 2013.
[13] A. E. Al-Fagih, F. M. Al-Turjman, W. M. Alsalih, and H. S. Hassanein, "A priced public sensing framework for heterogeneous IoT architectures," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 133–147, Jun. 2013.
[14] S. Li, L. Xu, X. Wang, and J. Wang, "Integration of hybrid wireless networks in cloud services oriented enterprise information systems," *Enterp. Inf. Syst.*, vol. 6, no. 2, pp. 165–187, Jan. 2012.
[15] W. He and L. Xu, "Integration of distributed enterprise applications: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 35–42, Feb. 2014.
[16] J. Zhao, "Enabling multi-hop communications through cross-layer design for hybrid WSNs with transmit-only nodes," *IEEE Global Telecommun. Conf.,* 2011, pp. 1–5.
[17] J. Zhao, C. Qiao, R. S. Sudhaakar, and S. Yoon, "Improve efficiency and reliability in single-hop WSNs with transmit-only nodes," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 3, pp. 520–534, Mar. 2013.
[18] W. Viriyasitavat, L. D. Xu, and A. Martin, "SWSpec: The requirements specification language in service workflow environments," *IEEE Trans. Ind. Informat.*, vol. 8, no. 3, pp. 631–638, Aug. 2012.
[19] J. Guo, L. D. Xu, G. Xiao, and Z. Gong, "Improving multilingual semantic interoperation in cross-organizational enterprise systems through concept disambiguation," *IEEE Trans. Ind. Informat.*, vol. 8, no. 3, pp. 647–658, Aug. 2012.
[20] V. M. Rohokale, N. R. Prasad, and R. Prasad, "A Cooperative Internet of Things (IoT) for rural healthcare monitoring and control," *Proc. 2nd Int. Conf. Wireless Commun., Veh. Technol., Inform. Theory Aerosp. Electron. Syst. Technol. (Wireless VITAE)*, pp. 1–6, 2011.
[21] S. Hachani, L. Gzara, and H. Verjus, "A service-oriented approach for flexible process support within enterprises: Application on PLM systems," *Enterp. Inf. Syst.*, vol. 7, no. 1, pp. 79–99, 2013.
[22] T. Melodia and I. F. Akyildiz, "Cross-layer QoS-aware communication for ultra wide band wireless multimedia sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 5, pp. 653–663, Jun. 2010.
[23] S. Ehsan and B. Hamdaoui, "A survey on energy-efficient routing techniques with QoS assurances for wireless multimedia sensor networks," *Commun. Surveys Tut.*, vol. 14, no. 2, pp. 265–278, 2012.
[24] Y. Wang, M. C. Vuran, and S. Goddard, "Cross-layer analysis of the end-to-end delay distribution in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 305–318, Feb. 2012.
[25] S. Li, L. D. Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2177–2186, Nov. 2013.
[26] S. Li, L. D. Xu, and X. Wang, "A continuous biomedical signal acquisition system based on compressed sensing in body sensor networks," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1764–1771, Aug. 2013.
[27] X. Sheng, J. Tang, X. Xiao, and G. Xue, "Sensing as a service: Challenges, solutions and future directions," *IEEE Sens. J.*, vol. 13, no. 10, pp. 3733–3741, Oct. 2013.
[28] E. I. Gaura, J. Brusey, M. Allen, R. Wilkins, D. Goldsmith, and R. Rednic, "Edge mining the internet of things," *IEEE Sens. J.*, vol. 13, no. 10, pp. 3816–3825, Oct. 2013.
[29] S. Li, G. Oikonomou, T. Tryfonas *et al.*, 2014). A Distributed Consensus Algorithm for Decision 2 Making in Service-Oriented Internet of Things, *IEEE Trans Ind. Informat.*, [Online]. Available: doi: 10.1109/TII.2014. 2306331 in press.

**Ling Li** received the Master's and Doctorate degrees in production/operations and logistics from the Ohio State University, OH, USA, in 1994 and 1996, respectively.

She is a Professor of Production/Operations with Old Dominion University, VA, USA, and a Fellow in Production and Inventory Management, Association for Operations Management, Chicago, IL, USA.

Dr. Li has served as Associate Editor for IEEE Transactions on Information Technology in Biomedicine, IEEE Transactions on Industrial Informatics, and other IEEE journals.



**Shancang Li** received the B.S. and M.S. degrees in mechanical engineering, and the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively.

Currently, he is a Research Assistant with the Faculty of Engineering, and a Member of the Cryptography Research Group, University of Bristol, Bristol, U.K. His research interests include mobile security, wireless sensor networks, Internet of Things, and applications of wireless technologies.

**Shanshan Zhao** received the B.S. and M.S. degrees in mechanical engineering from Xi'an Technological, Xi'an, China, in 2001 and 2004, respectively. She received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, in 2008.

Currently, she is an Academic Visitor with the University of the West of Scotland, Renfrewshire, U.K. Her current research interests include intelligent informations, IoT, and augmented reality.