IEICE TRANS. INF. & SYST., VOL.E97–D, NO.10 OCTOBER 2014

INVITED PAPER *Special Section on Frontiers of Internet of Things*

A Hybrid Trust Management Framework for Wireless Sensor and Actuator Networks in Cyber-Physical Systems

Ruidong LI[†a)], ***Member***, **Jie LI**[††], ***Nonmember***, ***and*** **Hitoshi ASAEDA**[†], ***Senior Member***

SUMMARY To secure a wireless sensor and actuator network (WSAN) in cyber-physical systems, trust management framework copes with misbehavior problem of nodes and stimulate nodes to cooperate with each other. The existing trust management frameworks can be classified into reputation-based framework and trust establishment framework. There, however, are still many problems with these existing trust management frameworks, which remain unsolved, such as frangibility under possible attacks. To design a robust trust management framework, we identify the attacks to the existing frameworks, present the countermeasures to them, and propose a hybrid trust management framework (HTMF) to construct trust environment for WSANs in the paper. HTMF includes second-hand information and confidence value into trustworthiness evaluation and integrates the countermeasures into the trust formation. We preform extensive performance evaluations, which show that the proposed HTMF is more robust and reliable than the existing frameworks.

key words: *cyber-physical systems, wireless sensor actuator network, trust management framework, security*

1. Introduction

Cyber-Physical Systems (CPS) [11] are the integrations of computation, networking, and physical processes. Recently the researchers identify that trust plays important role to secure CPS [6], [31], [32]. In this paper, we investigate trust management for one of the main components of CPS, wireless sensor and actuator networks (WSANs) [8], [26], which consist of a large number of typically small devices, each incorporating sensing, processing, and wireless communications capabilities.

WSANs [8], [26] are multi-hop wireless networks characterized by absence of any infrastructure, dynamic topology, wireless links, and constrained resources, which have great needs to be enabled to be trustworthy [3], [31], [32]. WSANs have interesting applications for information sharing, and opportunistic communications in various domains, such as agriculture, industry, and environment.

We focus on one of the most important parts to construct trust environment for WSANs, *trust management framework* [7], [25], [31], [32]. It is intended to cope with *misbehavior* problem of nodes and stimulate nodes to cooperate. Trust management framework has a wide range of applications including public key authentication [12], [24], [28], peer-to-peer networks [29], [30], and mobile ad hoc networks [22]. Because WSAN can be treated as a special case for ad hoc network with constrained resources, we also take the trust management framework for ad hoc network as the design references. Currently trust management becomes the foundation for many cryptography-based security mechanisms in WSANs [32]. Also detailed analysis on trust management framework using game theory has also been performed [1], [16], [25].

Herein, the *trust* is defined as the *belief level* that one node can put on another node for a specific action based on direct or indirect observations on behaviors of that node, similarly to [13]. Trust management framework is the framework to manage this kind of trust relations. Currently there are two categories of trust management frameworks for WSANs. One is the *reputation-based framework* (RBF) [7], [9], [20]. The other is the *trust establishment framework* (TEF) [4], [21], [22], [27]. It is noticeable that different names may be utilized for the final evaluated *trust* in different trust management frameworks, for example, reputation in RBF. In this paper, we use trustworthiness value as the final evaluated *trust*. By the RBF, trusts of other nodes are evaluated objectively based on direct observations and second-hand information. In contrast with RBF, for a TEF, trusts between nodes with direct interactions are evaluated based on direct observations and trusts between nodes without direct interaction are established through combination of the trusts of intermediate nodes.

Recently research attentions have been put on the intrinsic problems with trust management framework itself [2], [21]. The attacker not only can perform misbehaviors on forwarding packets, but can perform misbehaviors to make trust management framework malfunction. In [2], the false rating attack has been identified for RBF. But there are still some other unsolved problems with the method proposed in [2], for example, absence of considerations on another important parameter *confidence value*, vulnerability under on-off attack and conflicting behavior attack. In [21], a TEF was presented, by which some attacks can be handled. However, we discover two novel attacks that the framework in [21] cannot cope with. These two novel attacks are denoted by *selective misbehavior attack* and *location-dependent attack*.

To design a robust trust management framework, we firstly investigate the intrinsic problems with the existing trust management frameworks including the above two novel attacks. These problems cannot be solved by any single existing framework. After the corresponding counter-

Manuscript received January 15, 2014.

[†]The authors are with NICT, Koganei-shi, 184–8795 Japan.

[††]The author is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba-shi, 305–8573 Japan.

a) E-mail: lijie@cs.tsukuba.ac.jp

DOI: 10.1587/transinf.2013THP0010

measures are identified, we propose a hybrid trust management framework (HTMF) for WSANs, which combines the merits of RBF and TEF while removing the problems associated with each of the two categories of frameworks. In the HTMF, trust is formed based not only on direct observations but second-hand information compared to TEF, and confidence value has been included into the trust evaluation in contrast with RBF. By HTMF, the observations are expired by influence exponential decrease method and the trust for the provider of the second-hand information is employed when evaluating trust. We perform performance evaluations for the HTMF. From the results, we can see that the proposed HTMF can obtain the more reliable trust compared with the existing RBF and it can inhibit the selective misbehavior attack and location-dependent attack more effectively compared with the existing TEF.

The remainder of the paper is organized as follows. In Sect. 2, the intrinsic problems of existing frameworks will be provided. Then, we provide the countermeasures to the intrinsic problems and integrate them into HTMF in Sect. 3. In Sect. 4, we introduce the proposed HTMF, which is designed based on a novel modified Bayesian approach. In Sect. 5, we provide performance evaluations to compare the proposed HTMF with the existing frameworks. Finally, we conclude our work in Sect. 6.

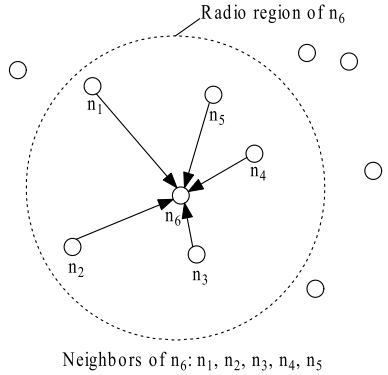

Fig. 1 A topology for descriptions of possible attacks.

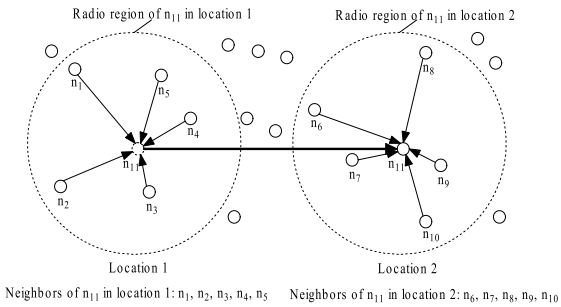

Fig. 2 Location-dependent attack.

2. Attacks to the Existing Frameworks

A WSAN is composed of many sensor and actuator nodes that have responsibility to forward packets for other nodes besides their own communications. The existing trust management frameworks themselves are vulnerable under various attacks, which will be identified in this section. We identify that these attacks still cannot be solved by any single existing framework till now.

2.1 Selective Misbehavior Attack

Consider that an attacker performs misbehaviors to victim nodes who it wants to attack and normal behaviors to the nodes that play crucial role to provide network service. We call this attack selective misbehavior attack. It is an attack similar to packet drop attack in [18]. This attack is harmful to TEF.

As for the TEF, trust from one node to another node is evaluated subjectively only based on direct observations obtained by watchdog mechanism. Watchdog mechanism [14] is implemented by comparing the sent packets with the overheard packets to see if there is a match. Take the topology in Fig. 1 as an example. Here, n_6 is assumed to be an attacker. The attacker, n_6, forwards the packets from n_2 with drop ratio 90%, but the packets from other neighbors with drop ratio 10%. By the TEF, the behaviors from n_6 to n_2 only can be reflected in the evaluated trust from n_2 to n_6. However, they cannot influence the trusts from nodes n_1, n_3, n_4, n_5 to n_6. Thus, n_6 performs misbehavior to n_2, but is contradictorily thought as a good guy by other nodes.

2.2 Location-Dependent Attack

Consider that an attacker misbehaves at one location and behaves normally at another location. We call this attack location-dependent attack. This attack is harmful to TEF, which also roots in the subjective characteristic, because the behaviors at one location cannot influence the trust evaluation of nodes at another location.

An example for location-dependent attack is shown in Fig. 2. In Fig. 2, n_{11} is assumed to be an attacker. At location 1, n_{11} forwards the packets from all the neighbors with high drop ratio 90%, which makes its trust low. But when it is desired to send packets, it moves to location 2. Here it forwards the packets for these new neighbors with drop ratio 10%, which makes its trust high. At location 2, the trusts from nodes $n_6, n_7, n_8, n_9, n_{10}$ to n_{11} have not been influenced by the misbehavior of n_{11} at location 1. Thus, n_{11} can obtain normal service at location 2 in spite of its misbehaviors at location 1.

2.3 Other Attacks

We also consider other attacks identified in [2], [21]. By on-off attack [21], attack nodes perform normal behaviors at one time period and then perform misbehaviors at another time period. Since the *trust* of attack node in existing frameworks does not decrease sharply, the attack node can perform many misbehaviors before it is stopped. This attack is harmful for both RBF and TEF.

Malicious nodes can attack the framework by providing false recommendations including false accusation and false praise, which is referred to as bad mouthing attack [21] and false ratings [2]. This attack is harmful for the RBF in form of false reports on the observed behaviors. Also this attack can put effect on the TEF to provide deliberate or bias recommendations.

By the conflicting behavior attack, attack nodes will behave differently to the nodes in different groups to make the opinions from different good groups to the attacker conflicting, and then make them be unable to trust each other. This attack is harmful for both the RBF and TEF, when they are equipped with recommendation generation system.

By sybil attack [15], [19], [21], attacker uses the faked IDs to cooperate with each other to make the trust system run out. New comer attack [10], [21] means that attacker can remove their bad history by registering as a new user. Both these two attacks can be inhibited by the usage of authentication scheme, which is beyond the topic of this paper.

2.4 Other Problems with RBF

To describe the trust for a node accurately, there are two important parameters, *trust value* and *confidence value*. Trust value corresponds to the estimation of a node's trust on a specific action. Confidence value is another indispensable parameter which characterizes the statistical reliability of the computed trust value. For the RBF, however, confidence value has not been involved, which makes the evaluated trust value sceptical. Moreover, the detailed second-hand information distribution method has not been provided in RBF.

3. Countermeasures of the Attacks

We firstly address selective misbehavior attack and location-dependent attack. Both two attacks root in the subjective characteristic of TEF. Therefore, it is necessary to design an objective trust management framework, by which the trust is formed based on not only direct observations but also second-hand information.

As for on-off attack, we propose the exponential decrease method with the adaptive discount factor. That is, only if a node performs normal behavior continuously for long time period, it can be put on high trust. However, if few misbehaviors of a node are observed, its trust will decrease sharply. Also, if few observations of normal behaviors are collected after many observations of misbehavior, its trust can be raised sharply to encourage such behavior.

The solution to deal with bad mouthing attack and conflicting behavior attack consists of two parts, detection part and process part. In the detection part, deviation test and the check on trust level of information provider are used to discover these two attacks. Deviation test is based on the statistical characteristic of observations [2]. After deviation test, *recommendation generation system* is proposed as a framework to manage the trust levels of these recommendations issued by different nodes, which is used to differentiate these two attacks. In the process part, if bad mouthing attack is detected, recommendation generation system is used to punish the attacker by lowering the trust level. Otherwise, the attacker is punished by including this second-hand information into trust evaluation for the attacker.

For other newly discovered problems with existing RBF, the confidence value will be included into the proposed trust management framework and a second-hand information distribution method in more detailed form will be presented.

To design a robust trust management framework, we combine the merits of the existing frameworks while removing the problems of them. All these proposed countermeasures are integrated into a hybrid framework, HTMF. Within HTMF, trusts for nodes in the network are evaluated based on both direct observations and second-hand information. Deviation test and the recommendation generation system are included into second-hand information processing procedure. Also, influence exponential decrease method with adaptive discount factor is integrated into trust evaluation. Moreover, trust value and confidence value are evaluated and combined into a whole metric, *trustworthiness*. Here, trustworthiness is the whole metric to show the trust levels of nodes in this paper.

4. Proposed Hybrid Trust Management Framework

The proposed hybrid trust management framework (HTMF) is designed based on a novel modified Bayesian approach. Here, we firstly introduce standard Bayesian approach [2], [7], [22]. Assume that subject node believes object node behaves normally with probability θ, which can also be described as $p(B)$. Here B will be *belief*. Also we simplify *Observation* to be O. Similarly to [7], the formula for standard Bayesian approach is provided as follows.

$$p(B|O) = \frac{p(O|B) * p(B)}{Normalizing\ Constant} \tag{1}$$

where $p(B)$ is the prior probability, $p(O|B)$ is the likelihood function, and $p(B|O)$ is the posterior distribution.

Beta distribution is the most promising distribution to represent $p(B)$, since it is flexible and simple and its conjugate is also a Beta distribution [2], [7], [9], [22]. Therefore, θ in HTMF is assumed to follow Beta distribution [5] as follows.

$$Beta(\theta, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \quad \forall\ 0 \le \theta \le 1, \alpha \ge 0, \beta \ge 0 \tag{2}$$

From Eq. (2), we can see that there are two parameters to characterize a Beta distribution, α and β, which is very suitable for trust management. Within HTMF, α and β are used to denote magnitude of normal behaviors and misbehaviors, respectively.

In this paper, the notation, $\{subject : object, action\}$, is used to denote the trust relation from a subject node to an object node on a specific action. We use ITF to denote

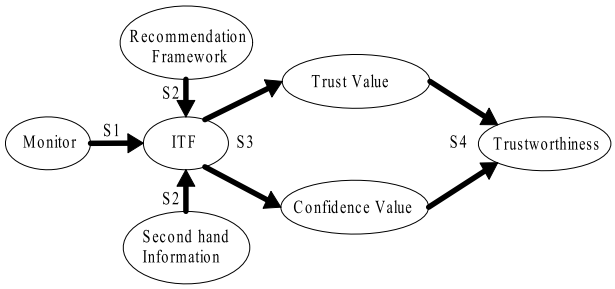

Fig. 3 HTMF: Hybrid trust management framework.

initial trust form that is formed by the collected data. $ITF\{i : j, action\}$, the initial trust form from node i to node j on a specific action, *action*, is defined as $(\alpha_{ij}, \beta_{ij})$. Here α_{ij} and β_{ij} are the number of normal behaviors and the number of misbehaviors of node j observed by node i, respectively. At the same time, second-hand information S_{kj} is similarly defined as the pair $(\alpha_{kj}, \beta_{kj})$.

The utility of standard Bayesian approach is provided as follows. Initially, θ is uniformly distributed between 0 and 1, which is described as $Beta(\theta, 1, 1)$. Then if there are s observations with normal behaviors and f observations with misbehaviors, the posterior distribution is updated by $\alpha = \alpha + s$ and $\beta = \beta + f$. After training by a large number of observations, θ will be close to $\frac{\alpha}{\alpha+\beta}$, with high probability. It can be concluded that if one node performs more normal behaviors, θ will converge to 1 and this node is more trustable [2], [13].

By the standard Bayesian approach, the same weight is given to each observation, regardless of the time of occurrence or who the provider is. Here to integrate all the countermeasures proposed in Sect. 3 into our framework, we develop a novel modified Bayesian approach. Firstly, to expire old observations and defense against on-off attack, influence exponential decrease method is used. When s observations with normal behaviors and f observations with misbehaviors are collected during time period t_d, α and β are updated by $\alpha = w_1^{t_d} * (\alpha - 1) + 1 + s$ and $\beta = w_1^{t_d} * (\beta - 1) + 1 + f$, where w_1 is the discount factor. To defense against on-off attack, w_1 should be an adaptive value. Secondly, to punish nodes performing bad mouthing attack, we use the trustworthiness of the information provider in recommendation generation system as the weight for the second-hand information it published.

4.1 HTMF Overview

We provide the skeleton for the HTMF as in Fig. 3, which consists of four steps, $S1$, $S2$, $S3$, $S4$ as below.

Step S1: Update ITF through Direct Information. Each node in the network monitors the behavior of its neighbors using watchdog mechanism [14]. In this step, the influence exponential decrease method with adaptive discount factor is used to expire old observations and defend against on-off attack.

Step S2: Distribute and process second-hand information. The direct observations obtained by one node k about a neighboring node, node j, can be used by another node i as second-hand information about the behaviors of node j. The second-hand information is flooded in the network. The nodes receiving these information check them by deviation test and other checks and then use the *trustworthiness* of information provider in the recommendation generation system (see Sect. 4.6) as the weight. This is used to inhibit bad mouthing attack and conflicting behavior attack. Due to the watchdog mechanism, the behaviors (of node j) observed by any two neighbors of node j will never overlap each other.

Step S3: Evaluate *trust* and *confidence value* evaluation. One node forms the elementary opinion for another node, *trust value* and *confidence value*, based on ITF obtained through steps S1 and S2. A high *trust value* means that the subject node trusts that the object node can perform an action well. The *confidence value* means the accuracy of the calculated *trust value*. A high *confidence value* represents that the object node has passed a large number of tests which have been given by the subject and other nodes. Obviously opinions with a high confidence are more useful in making decisions. Here the vulnerability for absence of *confidence value* has been solved by including this parameter into opinion formation.

Step S4: Evaluate trustworthiness. Since two parameters are difficult for trust comparison [22], two parameters formed in S3, namely trust value and confidence value, are combined into a whole trust metric, trustworthiness, to another node.

We will elaborate each step in more detailed form in the following subsections.

4.2 ITF Update through Direct Information

At this step, the ITF is firstly initialized as $(1, 1)$. Then each node in the network observes the behaviors of its neighboring nodes, and updates the ITF in succession. When an observation for node j is obtained by node i, the ITF should be updated. Let $s \in \{0, 1\}$ be the set of symbols for observations. That is, if a normal behavior is observed, $s = 1$; otherwise $s = 0$. The ITF is updated as follows:

$$\begin{aligned} &\alpha_{ij} = w_1^{CT-t_{last}} * (\alpha_{ij} - 1) + 1 + s \\ &\beta_{ij} = w_1^{CT-t_{last}} * (\beta_{ij} - 1) + 1 + 1 - s \\ &(0 \le w_1 \le 1) \end{aligned} \tag{3}$$

where w_1 is a discount factor, CT is current time, and t_{last} is the time point that last update was performed. w_1 is an adaptive value between 0 and 1. $w_1^{CT-t_{last}}$ is the factor to expire old observations exponentially, which is called influence exponential decrease method in this paper. In Eq. (3), we use $\alpha_{ij} - 1$ and $\beta_{ij} - 1$, because they are the actual number of observations on the behaviors and the minimum value for both

α_{ij} and β_{ij} is 1. Here we utilize memoryless characteristic for exponential distribution.

To defense against on-off attack, w_1 is set as an adaptive value which changes under different cases. To differentiate these cases, we use two characteristics for the collected information. One is normal behavior ratio for last fixed number of observations, which is denoted by *NBR* (Normal Behavior Ratio). For example, if we set this fixed number is 100 and there are 94 normal behaviors in the last 100 observations, *NBR* will be 94/100. The other characteristic is the detail information for recent fixed number of observations, which is denoted by *RO* (Recent Observations). For example, if this fixed number is 4, the last 2 observations on the behavior of one node are misbehaviors and the observations from last 3 to last 4 are normal behaviors, *RO* can be set as 1100. Here we use *abcd* to denote each bit for *RO*. Using these two characteristics, we provide four cases as follows.

Case 1: $NBR \geq Threshold$, $d = 1$ and $a+b+c+d >= 3$. This case corresponds to the situation that there are many normal behaviors which have been observed in the past time, current observations are normal behaviors and most of recent observations are also normal behavior. Under this situation, this trend should be encouraged. w_1 will increase until it approaches 1.

Case 2: $NBR \geq Threshold$, $d = 0$ or $d = 1$ and $a+b+c+d \leq 2$. This case corresponds to the situation that there are many normal behaviors in the past time, but current observation is misbehavior or current observation is normal behavior but most of recent observations are misbehaviors. The nodes under this situation should be punished strictly to prevent the trend for performing misbehavior. Thus, w_1 will drop greatly to a low value. It will decrease until the trust for this node reaches a threshold. Then for next misbehaviors, w_1 will increase gradually. This means whatever you have done many normal behaviors, if you perform misbehavior, the normal behaviors you did before will be much useless.

Case 3: $NBR < Threshold$, $d = 1$ or $d = 0$ and $a+b+c+d \geq 3$. Here, $d = 0$ and $a+b+c+d \geq 3$. This case corresponds to the situation that there are many misbehaviors in the past time, but current observation is normal behavior or current observation is misbehavior but the most recent observations are mostly normal behaviors. This trend should be encouraged, since the node is trying to perform normal behavior. For this situation, w_1 will drop greatly to a low value. It will decrease until the trust for this node is above a threshold. Then for next normal behaviors, w_1 will increase gradually. This means if you did a good behavior, your past of misbehaviors will be forgot quickly.

Case 4: $NBR < Threshold$, $d = 0$ and $a+b+c+d \leq 2$. This case corresponds to the situation that there are many misbehaviors which have been observed in the past time, current observation is still misbehavior and most of recent observations are misbehaviors. In this case, w_1 will increase gradually to punish the node until it approaches 1.

The descriptions above correspond to the situation that some observations have been collected during time interval, t_d. But if there is no observation obtained during t_d, the ITF will also be updated as follows:

$$\begin{aligned}&\alpha_{ij} = w_1^{CT-t_{last}} * (\alpha_{ij} - 1) + 1 = w_1^{t_d} * (\alpha_{ij} - 1) + 1\\&\beta_{ij} = w_1^{CT-t_{last}} * (\beta_{ij} - 1) + 1 = w_1^{t_d} * (\beta_{ij} - 1) + 1\\&(0 \leq w_1 \leq 1)\end{aligned} \tag{4}$$

At the same time, second-hand information is obtained every period T. At the beginning of every period, second-hand information, S_{kj}, is initialized as $(0, 0)$. If node k obtains an observation for j, the S_{kj} should be updated. Here also let $s \in \{0, 1\}$ be the set of symbols for observations. That is, if the observation is normal behavior, $s = 1$; otherwise $s = 0$. The S_{kj} should be updated as follows:

$$\begin{aligned}&\alpha_{kj} = w_1^{CT-t_{last}} * \alpha_{kj} + s\\&\beta_{kj} = w_1^{CT-t_{last}} * \beta_{kj} + 1 - s \quad (0 \leq w_1 \leq 1)\end{aligned} \tag{5}$$

Similarly, if node k has not obtained any observation during a time interval t_d, the S_{kj} will be updated as follows:

$$\begin{aligned}&\alpha_{kj} = w_1^{t_d} * \alpha_{kj}\\&\beta_{kj} = w_1^{t_d} * \beta_{kj} \quad (0 \leq w_1 \leq 1)\end{aligned} \tag{6}$$

The second-hand information is reset every period, T. When one period T reaches, it is kept as one piece of second-hand information. Meanwhile, S_{kj} is reset to (0, 0).

4.3 Second-Hand Information Distribution and Processing

To disseminate second-hand information throughout the network, we provide the detailed method for second-hand information distribution and processing here in contrast with [2], where it has not been provided in detail. Note that the recommendation generation system which we will present shortly in Sect. 4.6 is used in this step (i.e., Step S2).

After the formation of the second-hand information, it should be flooded throughout the network. We consider the situation that a node receives a published second-hand information. The algorithm it will perform is provided as below.

Algorithm :
if(*it has not been received before*)
{*receive this information and perform deviation test and one check*;
if(*bad mouthing attack is detected*)
{
drop this information;
update the trustworthiness of information provider in recommendation generation system.
}*else*{
obtain the trustworthiness of the provider from recommendation generation system;
update ITF;

distribute such message to its neighbors.
}
}*else*{
drop the message.
}

In the above algorithm, the node firstly should check whether it has received this information before. If it has, only drop this information. Otherwise, it will verify the reliability of such information to recognize bad mouthing attack and conflicting behavior attack. As mentioned in previous section, bad mouthing attack can be performed by issuing false information to disturb system, and conflicting behavior attack can be employed to disturb recommendation generation system. Thus, it is important to perform second-hand information verification to differentiate both of them. Thus after the node receiving a second-hand information, it will perform a deviation test. The deviation test is provided as follows.

$$|E(Beta(\theta,\alpha_{kj},\beta_{kj})) - E(Beta(\theta,\alpha_{ij},\beta_{ij}))| \le m \tag{7}$$

where m is the deviation threshold. If this test is passed, the received S_{kj} is reliable and start processing it. Otherwise, there are two cases which should be considered. Case 1: node k performs bad mouthing attack. Case 2: node j performs conflicting misbehavior attack. According to countermeasures in Sect. 3, here we use the check on trust level of information provider in recommendation generation system to differentiate them. If the trust for node k in recommendation generation system is lower than a threshold, node i will think node k performs bad mouthing attack. Thus this second-hand information will be dropped and one misbehavior of node k on recommendation is collected. Otherwise, node j is thought to perform conflicting behavior attack and this second-hand information will be included into the trust evaluation for node j, because this information is the real information on the behaviors of node j.

$$\begin{aligned}
\alpha_{ij} &= w_1^{CT-t_{last}} * (\alpha_{ij} - 1) \\
&\quad + w_1^{CT-t_{publishing\ time}} * w_2 * \alpha_{kj} + 1 \\
\beta_{ij} &= w_1^{CT-t_{last}} * (\beta_{ij} - 1) \\
&\quad + w_1^{CT-t_{publishing\ time}} * w_2 * \beta_{kj} + 1 \quad (0 \le w_1 \le 1) \\
w_2 &= T(i : k, recommendation)
\end{aligned} \tag{8}$$

where $w_1^{CT-t_{last}}$ and $w_1^{CT-t_{publishing\ time}}$ are the exponential decrease factor for expiring current ITF and the received second-hand information, respectively. To make trustworthiness in recommendation generation system influence trust evaluation, here it is used as the weight put on the received second-hand information. In (8), w_2 is set as $T(i : k, recommendation)$. $T(i : k, recommendation)$ represents the trustworthiness from node i to node k on the action, *recommendation*, in the recommendation generation system.

4.4 Trust and Confidence Value Evaluation

In HTMF, elementary trust from the subject node, node i, to the object node, node j, is composed of trust value and confidence value. Here confidence value is included into trust evaluation in contrast with RBF. The definition for it is similar to [22]. It is noticeable that a TEF has been proposed in [28], which is intrinsically different from the proposed HTMF, where second-hand information is included in trust evaluation. Trust value is to specify the trust estimation of node i to node j. Confidence value is to describe the accuracy of the evaluated trust value. Some notations are defined as follows.

- $t\{i : j, action\}$: Trust value that node i puts on node j for a specific action *action*. It has the property $0 \le t\{i : j, action\} \le 1$.
- $\sigma\{i : j, action\}$: Standard deviation of trust value from node i to node j on a specific action *action*.
- $c\{i : j, action\}$: Confidence value of trust value from node i to node j on a specific action *action*. It also has the property $0 \le c\{i : j, action\} \le 1$

Here we investigate calculation method for these parameters. Since the relation between the characteristic of Beta function and the trust is clarified in the first part of this Section, the *trust value* can be calculated as the expectation value of $beta(\theta,\alpha,\beta)$.

$$t\{i : j, action\} = E(Beta(\theta,\alpha,\beta)) = \frac{\alpha}{\alpha+\beta} \tag{9}$$

Here if $t\{i : j, action\}$ approaches to 1, it means that node i trusts node j to perform the action *action*. On the contrary, if $t\{i : j, action\}$ approaches to 0, it means that node i distrusts node j to perform the action *action*.

The other important parameter, $c\{i : j, action\}$, is used for characterizing the statistical reliability of the computed $t\{i : j, action\}$. It is a value between 0 and 1. Similarly to [22], $\sigma\{i : j, action\}$ and $c\{i : j, action\}$ are calculated as formula (10) and (11), respectively.

$$\begin{aligned}
\sigma\{i : j, action\} &= \sigma(Beta(\theta,\alpha,\beta)) \\
&= \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}
\end{aligned} \tag{10}$$

$$\begin{aligned}
c\{i : j, action\} &= 1 - \sqrt{12}\sigma(Beta(\theta,\alpha,\beta)) \\
&= 1 - \sqrt{\frac{12\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}
\end{aligned} \tag{11}$$

Here if $c\{i : j, action\}$ approaches to 1, it means that the evaluated *trust value* from node i to node j on the action *action* is believable because enough observations on behaviors have been collected. On the contrary, if $c\{i : j, action\}$ approaches to 0, it means that the evaluated *trust value* is untrustworthy because of the lack of observation collection.

4.5 Trustworthiness Evaluation

Here we combine (t, c) into one parameter, trustworthiness, which is the final evaluated *trust* for nodes. It can be utilized to judge whether a node is a good guy or not more easily.

We use $T\{i : j, action\}$ to represent the trustworthiness from node i to node j on a specific action $action$. Similarly to [22], the obtained $T\{i : j, action\}$ has the following properties.

- $0 \leq T\{i : j, action\} \leq 1$.
- $T\{i : j, action\}$ is induced from $t\{i : j, action\}$ and $c\{i : j, action\}$, but there are some rules for the calculation. Given a pair of trust value and confidence value, if the confidence value is high, trust value plays more important role for the trustworthiness formation. Thus under this situation, $t\{i : j, action\}$, should be put larger weight than confidence value $c\{i : j, action\}$. On the contrary, if the confidence value is low, obviously the confidence value is more important than trust value when evaluating trust. Therefore, $t\{i : j, action\}$, should be put less weight than confidence value $c\{i : j, action\}$.

Similarly to [22], the value of trustworthiness can be defined as

$$T\{i : j, action\} = 1 - \frac{\sqrt{\frac{(t\{i:j,action\}-1)^2}{x^2} + \frac{(c\{i:j,action\}-1)^2}{y^2}}}{\sqrt{\frac{1}{x^2} + \frac{1}{y^2}}} \tag{12}$$

where x and y are constants. The research in [22] shows that the most appropriate values for the trustworthiness parameters are $x = \sqrt{2}$ and $y = \sqrt{9}$. Therefore, in this paper, we also set x be $\sqrt{2}$ and y be $\sqrt{9}$.

Also a threshold value of trustworthiness is defined as

$$T_{threshold} := T(0.5, 0.5) = 0.5 \tag{13}$$

which represents the *trustworthiness* value assigned to a node with the *trust value* to be 0.5 and *confidence value* to be 0.5, respectively. This threshold value can be used to classify the nodes into good guys or bad guys. That is, if the *trustworthiness* from one node, i, to another node, j, is larger than $T_{threshold}$, it means that node i trusts that node j is a good guy and is preferable to perform a specific action. Otherwise, it denotes that node i does not believe that node j is preferable to perform an action.

4.6 Recommendation Generation System

The recommendation generation system is used to prevent the nodes in the network from providing false recommendation, which exhibits as the deliberate false second-hand information in the proposed HTMF. Since the HTMF is a general framework, it is also fit to construct a recommendation generation system. Therefore, recommendation generation system is designed as the trust management framework with action to be "recommendation".

In the recommendation generation system, $ITF(i : j, recommendation)$ is initiated as $(1, 1)$. If an observation of node j is obtained by node i, node i should firstly judge whether it is a normal behavior or misbehavior. The method to differentiate them is the deviation test and another check which are provided in Sect. 3. If one node is thought to perform bad mouthing attack, one misbhavior for it is observed. Then this ITF will be updated similarly as in Sect. 4.2. Also the collected observations should be published and processed as second-hand information similarly as in Sect. 4.3. After that, the trustworthiness in the recommendation generation system can be calculated as in Sects. 4.4 and 4.5. The obtained trustworthiness can be finally used as the weight on the second-hand information as in 4.3.

5. Performance Evaluation

Among these intrinsic problems, selective misbehavior attack and location-dependent attack are two novel attacks discovered in this paper. The countermeasures to other attacks are similar to those proposed in [2], [21]. Their effectiveness have already been verified. Thus, here we only clarify the robustness of HTMF under the two newly discovered attacks and the effectiveness for including confidence value into trust evaluation.

5.1 Selective Misbehavior Attack

To demonstrate that the proposed HTMF can inhibit the selective misbehavior attack, which occurs in the TEF, we investigate the following two metrics.

1. The trustworthiness values to the attacker, which are the trust levels from other nodes to the attacker.
2. The throughput of the attacker, which is defined as the total successfully delivered message divided by the simulation time.

5.1.1 Trustworthiness Value

We consider the scenario depicted in Fig. 1. Here, we will not consider influence exponential decrease of observations. In this scenario, n_6 is the attacker, who performs selective misbehavior attack. Here, it is assumed that n_6 forwards the packets from n_2 with drop ratio 90%, and with drop ratio 10% for other neighbors. In the mean time, there are 2000 packets for n_6 to forward for each neighbors, n_1, n_2, n_3, n_4, n_5.

Under this situation, we can obtain the result as in Fig. 4. In Fig. 4, by the TEF, the trustworthiness from n_2 to n_6 is much lower. However, the trustworthiness from other neighbors to n_6 is much higher. Obviously, the misbehaviors from n_6 to n_2 have not influenced the trustworthiness from other neighbors to n_6. In contrast, by the proposed HTMF and the RBF, the trustworthiness and the corresponding reputation are the same for each neighbor. This is because by each of them, the trust for one node is evaluated objectively. Thus the misbehavior from n_6 to n_2 also put effect on the trust level from other neighbors to n_6. That is, the attacker cannot perform misbehaviors and keep its trustworthiness at

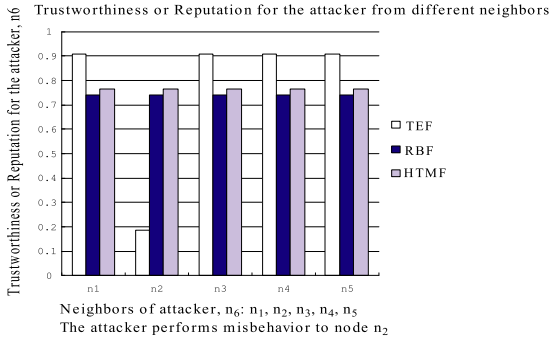

Fig. 4 Trustworthiness to the attacker from different nodes in the neighborhood.

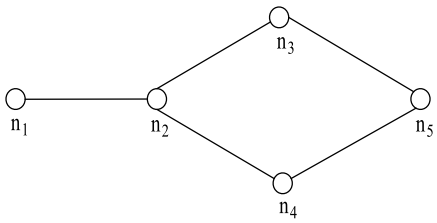

Fig. 5 The topology of the experiment network.

a high level at the same time. Therefore, we can see that the HTMF and the RBF can inhibit the selective misbehavior.

5.1.2 Throughput of the Attacker

To explain that the proposed HTMF can restrain the selective misbehavior attack, we carry out simulations to investigate how the throughput of the attacker changes with the drop ratio from the attacker to the victim node on the forwarding packets.

We simulate the proposed objective framework using an object-oriented modular discrete event simulator called OMNET++ [33]. In our simulation, each node is a compound module and the communications between the modules are made via message exchange. The routing model we use is the routing protocol, AODV [17], since it is representative protocol for multihop dynamic wireless network.

The topology shown in Fig. 5 is used for simulations. In this topology, the connections between any two nodes are the wireless links. That is, n_1, n_3, n_4 exist in the radio region of n_2, but n_5 is out of the radio region of n_2.

In this scenario, n_1 and n_2 send packets to n_5 with constant rate, 100 packets/second. At the same time, n_3 and n_4 send packets to n_1 with constant rate, 1 packet/second. Here n_2 is the attacker. In our implementation, we let n_2 forward the packets from n_1 with drop ratio 90%, 80%, 70%, 60%, 50%, but forward the packets from n_3, n_4, n_5 normally. In the simulation, the attacker concentrates the attack on the delivery of three kinds of packets, data packets, RREQ (Route Request) and RREP (Route Reply).

In the implementation of the proposed HTMF, the node in the network evaluates the trustworthiness based on all the

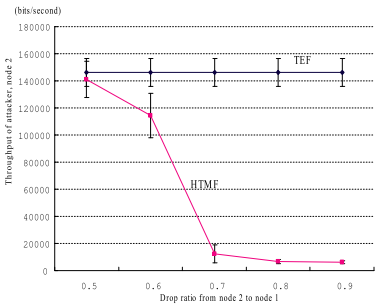

Fig. 6 The comparison of throughput for node 1.

observed behaviors from other nodes. In the implementation of TEF, the node evaluates the trustworthiness only based on the direct observations. Additionally, the normal nodes forward the packets from other node according to the trustworthiness of this node. In our simulation, if the trustworthiness from one node, n_i, to another node, n_j, is lower than $T_{threshold}$, n_i will think n_j is an attacker and will not forward the packets from n_j.

The simulation time for each run is 500 seconds. We run the simulation 40 times with different seeds. The confidence level in our simulation is 95%, and confidence interval is 10%. We can obtain the results for the throughput of the attacker, n_2, as in Fig. 6.

From Fig. 6, we can see that by the proposed HTMF, the throughput of the attacker drops greatly with its drop ratio to the packets from the victim node increases. This is because the misbehaving node is punished by high drop ratio from other nodes to the packets it sends when its trustworthiness drops. However, by TEF, the throughput has not been influenced by the misbehavior from the attacker to the victim node. At the same time, we can see that the throughput by the proposed framework drops greatly at an interval of the drop ratio to the victim, which is from 0.6 to 0.7. The reason for that is the critical point for the trustworthiness value equaling to $T_{threshold}$ exists in this interval. Totally speaking, the simulation result demonstrates the effectiveness of HTMF in inhibiting the selective misbehavior attack compared with the existing framework.

5.2 Location-Dependent Attack

For location-dependent attack, we also investigate trustworthiness value to the attacker, and will not consider influence exponential decrease method. Here we consider the scenario depicted in Fig. 2. In this scenario, n_{11} is the attacker performing location-dependent attack.

Without loss of generality, it is assumed that each node in any location has 2000 packets, which need n_{11} to forward. Meanwhile, we assume that n_{11} forwards the packets for the neighbors at location 1 with drop ratio 90%, and for the neighbors at location 2 with drop ratio 10%.

For this situation, we obtain the results as in Fig. 7. In this Figure, by TEF, the trustworthiness from the neighbors of n_{11} at location 1 is much lower, while the trustworthiness

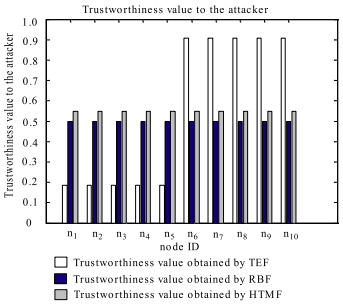

Fig. 7 Trustworthiness to the attacker for different nodes at different places when the attacker is at place 2.

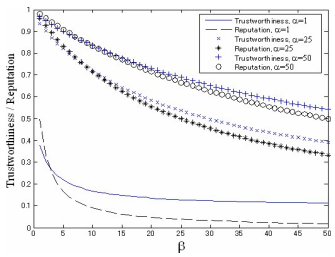

Fig. 8 Trustworthiness obtained by HTMF Vs Reputation obtained by RBF under the following cases: case 1: α=1; case 2: α=25; case 3: α=50; In all cases, β varies from 1 to 50.

from the neighbors at location 2 is very high. It is obvious that n_{11} still can obtain good service at location 2 even it has performed many misbehaviors at location 1. The reason that this unfair status occurs is also because the trust for a node is evaluated subjectively only based on direct observations. In contrast, by the proposed HTMF or RBF, the trust for a node is evaluated objectively. Therefore, the behaviors at location 1 put effect on the trust evaluation at location 2 and location-dependent attack can be prevented by HTMF.

5.3 Absence of Consideration on Confidence Value

To show the necessity of introducing confidence value to the framework, we compare the evaluated trusts by the proposed HTMF and the RBF. We will not consider the influence exponential decrease for observations here. The trust metrics in HTMF and RBF are expressed as trustworthiness and reputation, respectively.

We consider three cases: case 1: α=1; case 2: α=25; case 3: α=50. That is, the number of observations on normal behaviors are set as 1, 25 and 50. In all these cases, β varies from 1 to 50, which means the number of observations on misbehaviors varies from 1 to 50. We can obtain the results as Fig. 8. From this figure, we can see that for all cases when the number of observations is low, the evaluated trustworthiness by HTMF is lower than the reputation obtained by RBF. This is because that the low confidence value influences on the evaluated trust. On the other hand, when the number of observations becomes larger, the confidence value will become higher which reflects in the higher trust for the HTMF than that for RBF.

Also, with increasing of number of observations, the difference between HTMF and RBF increases, because HTMF introduces confidence value in trust formation. When more observations are collected, the evaluated trust is more trustable in HTMF. In contrast, RBF has not included confidence value and it is not influenced by the number of the collected observations. It is obvious that the trustworthiness of HTMF coincides with the human intuition. The trustworthiness with more observed behaviors should be higher than the nodes with less observed behaviors.

Therefore, we can see that by HTMF, the more reliable trust can be obtained than existing RBF, because HTMF can correspond well to decrease the evaluated trust when the number of collected observations is small and to increase the evaluated trust when the number of observations increases compared to RBF.

6. Conclusions

In the paper, we clarify the intrinsic problems with existing frameworks, and provide the countermeasures for them. Then we propose a novel hybrid trust management framework called HTMF. The proposed HTMF holds objective feature by which trust for a node is evaluated based on not only direct observations but second-hand information. It makes HTMF robust under selective misbehavior attack and location-dependent attack in contrast to the TEF. The proposed HTMF can also inhibit other possible attacks in the existing frameworks, such as on-off attack, bad mouthing attack, and conflicting behavior attack. Also confidence value has been included into the trust evaluation and detailed second-hand information distribution and processing method have been provided. We perform performance evaluations for the comparison between the proposed HTMF and existing frameworks. From the evaluation results, we can see that HTMF is more robust and reliable than existing frameworks.

Acknowledgements

This work is partially supported by Grant-in-Aid for Scientific Research of Japan Society for Promotion of Science (JSPS) and Collaboration Research Grant of National Institute of Informatics (NII), Japan.

References

[1] J.S. Baras, T. Jiang, and P. Purkayastha, "Constrained coalitional games and networks of autonomous agents," Proc. Third International Symposium on Communications, Control and Signal Processing, pp.972–979, St. Julians, Malta, March 2008.
[2] S. Buchegger and J.-Y. Le Boudec, "A robust reputation system for P2P and mobile ad-hoc networks," Proc. P2PEcon 2004, Harvard University, Cambridge MA, USA, June 2004.
[3] E. Chang and T.S. Dillon, "Trust, reputation, and risk in cyber physical systems," AIAI 2013, 2013.
[4] D. Chen, G. Chang, D. Sun, J. Li, J. Jia, and X, Wang, "TRM-IoT: A trust management model based on fuzzy reputation for Internet of things," Comput. Sci. Inf. Syst., vol.8, no.4, pp.1207–1228, 2011.

[5] A. Davison, "Statistical models," Cambridge University Press, Cambridge Series in Statistical and Probabilistic Mathematics, June 2003.
[6] J. Daly, "Securing cyber-physical systems in the age of connectivity," http://www.fedtechmagazine.com/article/2013/12/securing-cyber-physical-systems-age-connectivity, FedTech Magazine, Dec. 2013.
[7] S. Ganeriwal and M. Srivastava, "Reputation-based framework for high integrity sensor networks," Proc. ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2004), Washington, D.C., USA, Oct. 2004.
[8] M. Haenggi, "Mobile sensor-actuator networks: Opportunities and challenges," 7th IEEE International Workshop on Cellular Neural Networks and Their Applications, 2002.
[9] A. Jϕsang and R. Ismail, "The beta reputation system," Proc. 15th Bled Conference on Electronic Commerce, Bled, Slovenia, June 2002.
[10] M. Kinateder, E. Baschny, and K. Rothermel, "Towards a generic trust model - Comparison of various trust update algorithms," iTrust, pp.177–192, 2005.
[11] E.A. Lee, "Cyber physical systems: Design challenges," Proc. 2008 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing (ISORC'08), pp.363–369, 2008.
[12] R. Levien and A. Aiken, "Attack-resistant trust metrics for public key certification," Proc. 7th USENIX Security Symposium, pp.229–242, Jan. 1998.
[13] J. Li, R. Li, and J. Kato, "Future trust management framework for mobile ad hoc networks," IEEE Commun. Mag., vol.46, no.4, pp.108–114, April 2008.
[14] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," Proc. MobiCom 2000, pp.255–265, Aug. 2000.
[15] J. Newsome, E. Shi, D. Song, and A. Perrig, "The sybil attack in sensor networks: Analysis and defenses," Proc. Third International Symposium on Information Processing in Sensor Networks (IPSN), 2004.
[16] T.G. Papaioannou and G.D. Stamoulis, "Achieving honest rating with reputation-based fines in electronic markets," Proc. IEEE Infocom 2008, April 2008.
[17] C. Perkins and E. Royer, "Ad hoc on-demand distance vector routing," 2nd IEEE Workshop on Selected Areas in Communication, pp.90–100, New Orleans, LA, Feb. 1999.
[18] Y. Cho, G. Qu, and Y. Wu, "Insider threats against trust mechanism with watchdog and defending approaches in wireless sensor networks," IEEE Symposium on Security and Privacy Workshops, pp.134–141, May 2012.
[19] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara, "Reputation systems," Commun. ACM, vol.43, no.12, pp.45–48, 2000.
[20] J. Sen, "A distributed trust management framework for detecting malicious packet dropping nodes in a mobile ad hoc network," Int. J. Network Security & Its Applications (IJNSA), vol.2, no.4, pp.92–104, Oct. 2010.
[21] Y. Sun, Z. Han, W. Yu, and K.J.R. Liu, "A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks," Proc. IEEE Infocom 2006, Barcelona, Spain, April 2006.
[22] C. Zouridaki, B.L. Mark, M. Hejmo, and R.K. Thomas, "A quantitative trust establishment framework for reliable data packet delivery in MANETs," Proc. 3rd ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2005), Alexandria, VA, USA, Nov. 2005.
[23] B. Tajes, M. Rajkumar, K. Sushan, and K. Chandrasekaran, "Trust management in ad hoc networks: A social network based approach," Network and Complex Systems, vol.1, no.1, pp.24–32, 2011.
[24] Y. Takehana, I. Nishimura, N. Yosaka, T. Nagase, and Y. Yoshioka, "Building trust among certificates management nodes in mobile ad-hoc networks," Proc. 26th International Conference on Advanced Information Networking and Applications Workshop, pp.564–568, 2012.
[25] G. Theodorakopoulos and S. Baras, "Malicious users in unstructed networks," Proc. IEEE Infocom 2007, Alaska, USA, April 2007.
[26] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, Wireless Sensor and Actuator Networks: Technologies, Analysis and Design, Academic Press, 2008.
[27] H. Xia, Z. Jia, L. Ju, X. Li, and Y. Zhu, "A subjective trust management model with multiple decision factors for MANET based on AHP and fuzzy logic rules," 2011 IEEE/ACM International Conference on Green Computing and Communications (GreenCom2011), pp.124–130, Aug. 2011.
[28] H. Yong, C. Yu, C. Zhou, and W. Song, "A distributed key management framework with cooperative message authentication in VANETs," IEEE J. Sel. Areas Commun., vol.29, no.3, pp.616–629, 2011.
[29] B. Yu, M.P. Singh, and K. Sycara, "Developing trust in large-scale peer-to-peer systems," Proc. First IEEE Symposium on Multi-Agent Security and Survivability, 2004.
[30] R. Zhou and K. Hwang, "PowerTrust: A robust and scalable reputation system for trusted peer-to-peer computing," IEEE Trans. Parallel and Distributed Systems, vol.18, no.4, pp.460–473, April 2007.
[31] "Designed-in cyber security for cyber-physical systems," http://www.cybersecurityresearch.org/documents/CSRA_Workshop_Report.pdf, 4-5 April 2013, Gaithersburg, Maryland.
[32] "Internet of Things in 2020: Roadmap for the future," http://www.smart-systems-integration.org/public/documents/publications/Internet-of-Things_in_2020_EC-EPoSS_Workshop_Report_2008_v3.pdf, 05 Sept., 2008.
[33] "OMNET++ Community Site," http://www.omnetpp.org/index.php

Ruidong Li is a researcher of the network architecture laboratory at NICT. He received a bachelor in engineering from Zhejiang University, China, in 2001. He received a master and doctorate of engineering from the University of Tsukuba in 2005 and 2008, respectively. Since 2008, He is a member of the AKARI architecture design project and network architecture laboratory in NICT. His current research interests include information-centric network, internet of things, security/secure architectures of future networks, and regional platform network.

Jie Li is a professor in Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. His research interests are in mobile distributed multimedia computing and networking, OS, network security, modeling and performance evaluation of information systems. He is a senior member of IEEE and ACM, and a member of IPSJ (Information Processing Society of Japan). He has served as a secretary for Study Group on System Evaluation of IPSJ and on several editorial boards for IPSJ Journal, IEEE Transactions on Vehicular Technology, Wiley Wireless Communications and Mobile Computing, and so on, and on Steering Committees of the SIG of System EVAluation (EVA) of IPSJ, the SIG of DataBase System (DBS) of IPSJ, and the SIG of MoBiLe computing and ubiquitous communications of IPSJ. He has also served on the program committees for several international conferences such as IEEE ICDCS, IEEE INFOCOM, IEEE GLOBECOM, and IEEE MASS.

Hitoshi Asaeda is a Planning Manager of Network Research Headquarters, National Institute of Information Communications Technology (NICT). From 1991 to 2001, he was with IBM Japan, Ltd. From 2001 to 2004, he was a Research Engineer Specialist at INRIA Sophia Antipolis, France. He was Project Associate Professor of Graduate School of Media and Governance, Keio University, where he was during 2005-2012. He holds a Ph.D. from Keio University. His research interests include routing architectures and future Internet technologies.