Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 6506341, 10 pages
http://dx.doi.org/10.1155/2016/6506341

Research Article

Protecting Mobile Crowd Sensing against Sybil Attacks Using Cloud Based Trust Management System

Shih-Hao Chang and Zhi-Rong Chen

Department of Computer Science and Information Engineering, Tamkang University, New Taipei City 25137, Taiwan

Correspondence should be addressed to Shih-Hao Chang; shhchang@mail.tku.edu.tw

Received 7 August 2015; Accepted 16 February 2016

Academic Editor: Jong-Hyouk Lee

Mobile crowd sensing (MCS) arises as a new sensing paradigm, which leverages citizens for large-scale sensing by various mobile devices to efficiently collect and share local information. Unlike other MCS application challenges that consider user privacy and data trustworthiness, this study focuses on the network trustworthiness problem, namely, Sybil attacks in MCS network. The Sybil attack in computer security is a type of security attack, which illegally forges multiple identities in peer-to-peer networks, namely, Sybil identities. These Sybil identities will falsify multiple identities that negatively influence the effectiveness of sensing data in this MCS network or degrading entire network performance. To cope with this problem, a cloud based trust management scheme (CbTMS) was proposed to detect Sybil attacks in the MCS network. The CbTMS was proffered for performing active and passive checking scheme, in addition to the mobile PCS trustworthiness management, and includes a decision tree algorithm, to verify the covered nodes in the MCS network. Simulation studies show that our CbTMS can efficiently detect the malicious Sybil nodes in the network and cause 6.87 Wh power reduction compared with other malicious Sybil node attack mode.

1. Introduction

In recent years, mobile computing devices on the market, for example, smartphones and tablet computers, have become ubiquitous. Differing from the last century, the mobile phone of today, namely, the smartphone, usually comes with multifunction sensors, such as camera, microphone, GPS, accelerometer, digital compass, and gyroscope. These new technologies have enabled smartphone users to collect sensed data from their local information and upload these sensed data back to an application server using existing wireless communication infrastructure (such as 3G/4G/5G services or even WiMAX access points). Smartphones provide an excellent platform for mobile crowd sensing (MCS) [1]. Hence, a requester of data can create tasks that use the general public to capture geotagged images, videos, audio snippets, or all-out surveys. Participants who have installed the client apps on their smartphones can submit their data and get rewarded.

In recent years, a surfeit of novel and fascinating MCS applications have been developed, ranging from health care to multiple cultural aspects. Two examples of MCS applications are BALANCE [2] and HealthSense [3], used to collect and share data about personal health projects, which monitor the activities and behavior related to diet and encourage healthy living. MCS application provides a very open concept platform, which allows anybody to contribute their local sensing information; however, it may also leak malicious and erroneous attacks to the application. Sharing sensed data tagged with spatiotemporal information could reveal a lot of personal information, such as users' identity, personal activities, political views, and health status, thereby posing threats to the participating users. Malicious participants may unintentionally position the phone in an adverse position or deliberately contribute bad data while collecting sensor readings from mobile phones. Hence, an attacker can use these identities to act maliciously, by providing huge amount of corrupt data to either degrading data correctness or network performance via a Sybil attack in MCS application.

Despite the great number of works in network security literature, a systematic study and classification of the research problems of the Sybil attacks in MCS research domain is lacking to guide further research and development of this

emerging field. The Sybil attack was first introduced by Microsoft researcher Douceur, who described a Sybil attack, relying on the fact that a participatory sensing network possibly includes tremendous and unrealistic data from different resource parities and coordination among entities [4]. Sybil attacks relied on the fact that a MCS network data server cannot confirm that each unknown data-collecting element is a distinct mobile device. Therefore, any malicious MCS network attack can try to inject false information into the network to confuse or even collapse the network applications. Recently, some researchers have revealed Sybil attacks have compromised mobile social network [5] and Internet of Things (IoT) [6] because social network and IoT platform are both vulnerable to Sybil attacks, as Sybil attacks can manipulate pseudoidentities to compromise the effectiveness of social network and IoT system.

Cloud computing provides flexible and on-demand infrastructures which have drawn lots of attention from research and industry in recent years. Cloud computing services commonly denote functionalities such as IaaS (infrastructure as a service), SaaS (software as a service), and PaaS (platform as a service), delineated as a layered system structure for cloud computing. TaaS (trust as a service) decides which types of solutions are appropriate for their unique needs. Recently, several methodologies have been offered for trust management in cloud computing environments [7–9]. For example, CATRAC [9] has proposed security architecture related to combining web services. CATRAC combines both Role-Based Access Control and Trust-Based Access Control in order to arrive at an optimum solution. The authors in [9] described that trust levels are presented as a vector ranging from 0 to 10, indicating "fully distrusted" to "fully trusted," respectively. Five denotes an uncertainty or a neutral level, which is commonly assigned to new clients. Nevertheless, analyzing trust issues from a cloud user will normally respect their data in terms of security and privacy. Therefore, a good reputation system requires reducing alliance of user identification and his/her privacy information.

To solve this problem, a cloud based trust management scheme (CbTMS) is proposed to evaluate the Sybil attacks in MCS applications. The proposed CbTMS framework proffered active and passive checking schemes that leverage mobile PCS and base station to perform the Sybil identity detection over a period of time. Moreover, to address the trustworthiness issue in the proposed system, the CbTMS also provides a Trust Credit Assessment and analytical decision support that perform the trust management service in the cloud to evaluate mobile PCS trustworthiness level. Hence, a high credit score is an indication that a particular smartphone device has been reporting reliable PCS. To verify this idea, the OMNeT++ simulation has been used to present our CbTMS's effectiveness against Sybil attacks. The rest of this paper is organized as follows. Section 2 presents a literature review of current related works and summarizes their conclusions. Section 3 provides the detection factors motivating the need for a reputation system in the context of MCS; it presents an overview of the system architecture. In Section 4, the experimental setup and simulation results are described and Section 5 concludes the paper.

2. Background

In recent years, there have been more and more MCS applications in different fields. For example, in personal health monitoring, BALANCE [2] allows clients to monitor their activities and diet behavior, encouraging healthy living. Food calories are entered via mobile phones and an accelerator detects movement patterns and time to project the calories consumed, thereby achieving health management. HealthSense [3] automatically detects health-related events, such as pain or depression, which cannot be observed directly through current sensor technology. HealthSense analyzes sensor data from the patient by applying machine learning methods. HealthSense also utilizes patient input events to assist in data classification (such as pain or itching). Finally, the user provides feedback on the machine learning process. As mentioned, MCS applications are subject to malicious attacks.

Due to the MCS applications, participants allow anyone with an appropriate mobile device that has the application installed to register as a participant. Such human intervention entails serious security and privacy risks. The free transmission of users' sensor data could result in compromised security and privacy. For instance, users may leak their personal identity information through personal responses. In [5], the mobile Sybil detection is exploited based on mobile user's friend and foe list. Mobile users can detect Sybil attackers with profile matching when they are encountered. Liang et al. [6] explore Internet of Things (IoT) exposed to Sybil attacks where attackers can manipulate fake identities or abuse pseudoidentities to compromise the effectiveness of the IoT and even disseminate spam. Particularly, in [6], the authors also outline three types of Sybil attacks, namely, SA-1, SA-2, and SA-3, according to the Sybil attacker's capabilities. As a result, Sybil detection in research efforts is becoming more popular for the development of both online and mobile Sybil detection and defense schemes in social network and IoT system.

Douceur formalized the Sybil attack in the perspective of peer-to-peer networks [4]. He presented that there is no practical solution for this attack and indicated that Sybil attacks can overthrow the redundancy mechanisms of distributed data storage systems. Problems arise when a reputation system (such as a trusted certification) is tricked into thinking that an attacking computer has a disproportionally large influence. Grover et al. [10] proposed a scheme to protect against the Sybil attack using neighboring nodes' information. In this approach, every node will participate to detect the suspicious node in the network. Every mobile node has a different group of neighbors at different time interval. After these mobile nodes share their network tables, they will match their neighboring tables; if some nodes are simultaneously observed with the same set of neighbors at different interval of time, then these nodes are under Sybil attack. In this case, identities are neighboring nodes associated with specific trust devices. Similar to a central authority creating certificates, there are only few methods to prevent an attacker from reaching multiple devices.

Trust and reputation have been verified as influencing customers or users in selecting high quality service in multiple situations. The concept of trust and reputation is similar

in computational models that can be formally characterized based on history of past interactions. For instance, after the completion of the transaction of rating among parties, the aggregated ratings about a given party can then be used to derive its reputation score. Nevertheless, it seems that threats to users' privacy will be encountered. To solve this problem, Ries [11] intuitively allows the analysis of trust as a subjective prospect, which permits the consideration of context-dependent and individual preferences parameters. However, building up trust and reputation usually requires long duration categorizing that can be a link across numerous transactions.

In cloud computing, trust management is one of the most critical matters and has become a popular research area [7–9]. For example, Brandic et al. [7] presented compliance management using a centralized approach in cloud server-side environments. This method supports customers to select proper trust services in the cloud environment from their own viewpoint. Hwang and Li [8] proposed security-aware cloud architecture, which offers data coloring techniques, and trust negotiation to support the cloud service from a provider perspective. The cloud service consumers' perspective is supported using the trust-overlay networks to deploy reputation-based trust management. Ghali et al. [9] have proposed a security framework called CATRAC to compose web services. CATRAC combines both Role-Based Access Control and Trust-Based Access Control in order to arrive at an optimum solution. However, unlike previous research works [7–9, 12], the proposed CbTMS method utilizes data mining approach to classify Sybil attacks models. Data mining is a new technology and has widely been used by data scientists for research and business purposes. The overall goal of the proposed data mining approach is to extract information from a dataset and convert it into a comprehensible structure for further use. Among many data mining techniques, the decision tree is appropriate for use to extract models and find out how certain variables are associated with important data classes.

Decision trees offer multiple advantages; however, one of the most important advantages is that the knowledge can be extracted and represented in the form of classification (if-then) rules. Decision tree induction is the learning of decision trees from class labeled training tuples. Each rule indicates a unique path from the root to each leaf. In operations research, specifically in decision analysis, a decision tree (or tree diagram) is a decision support tool. A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. There are several machine learning algorithms (MLAs) currently in use for distinguishing the normal and anomalous activities, including the ID3 [13] algorithm as well as Naive Bayes Filter, J48, and C4.5 [14–16] classification model. These classification models can be used in any point of the network, which provide very fast statistical detection of the application, to distinguish the normal and anomalous activities in a cloud. Fastidiousness of future classification by MLAs depends heavily on quality of the training data.

As described above, mechanisms and algorithms for MCS application, Sybil attack, and cloud computing trust models have been proposed. However, their approaches are not applicable to detect Sybil attacks in MCS environments by utilizing trust management system. Therefore, we attempt to identify Sybil attacks in MCS environment by utilizing a cloud based trust management system that differentiates between credible trust nodes' and malicious trust nodes' feedback through a credibility model.

3. Detection of the Sybil Attack in Mobile Crowd Sensing Factors

The mobile crowd entities in the system include smartphone or tablet PC, and the service provider will support interactions between them, that is, inquiries about environment information service. Therefore, such interaction will specify the service content. For example, a user using his/her smartphone, namely, entity A, is interacting with service providers regarding temperature information in his/her current location. Then, entity A here, an interaction initiator, will select a trustable service provider from a set of available service providers; he/she will start evaluating the trustworthiness of the available service providers from the selection list. Then, entity A will examine the direct evidence from previous interactions and recommendations from one or multiple service providers. Hereby, the trust model can be used for combining collected evidence or giving lower weight to recommendations from unreliable sources. The trust model derives trust values for the service providers and then becomes the basis for deciding whether to interact with one of the available services providers and which service provider to select.

Therefore, a cloud based service management framework has been proposed in this paper that consists of trust as a service (TaaS) using the service oriented architecture (SOA). In particular, the proposed cloud based service management framework applies web services to interact with distributed smartphones. This web service is one of the most significant empowering technologies for cloud computing; hence, its similarities to other resources (e.g., software, infrastructures, and platforms) in the cloud are unprotected as services. Therefore, when there is a trusted participant wishing to give his/her trust feedback or inquire about the current trust data in our SOA, he/she can utilize feedback message such as text messaging or multimedia messaging to deliver his/her own data or to get inquired trust data. Figure 1 represents the proposed framework; this framework contains three different layers: the cloud service provider layer, the trust management system layer, and the cloud service consumer layer.

The proposed architecture consists of a cloud service provider layer, which contains multiple service providers who provide cloud services. The minimum symbolic feature of every cloud service is at least providing infrastructure as a service (IaaS); that is, the cloud provider should have a data center that provides the storage, the processing power, and communication capability. Moreover, in the trust management system layer, the proposed layer consists of several distributed trust management system (TMS) nodes that expose interfaces so that cloud service consumers can give their trust

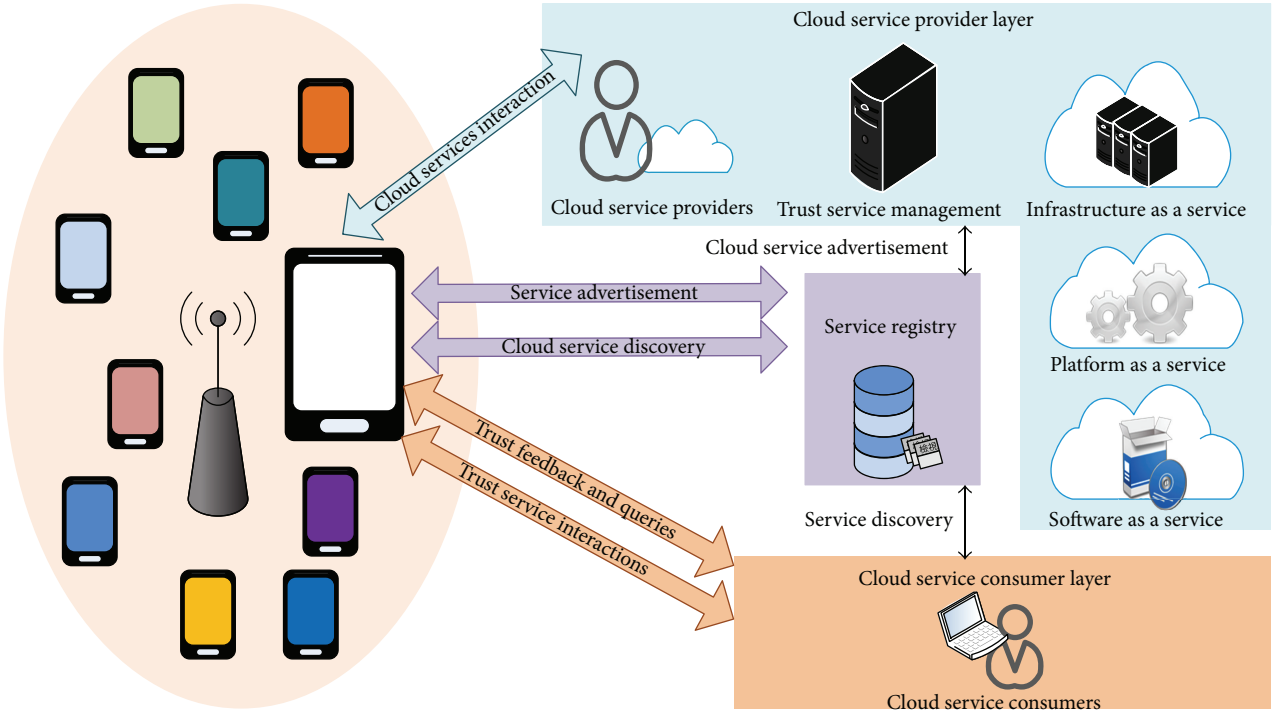

Figure 1: Architecture of the trust as a service framework.

feedback or inquire about the trust results. Furthermore, the cloud service consumer layer consists of different cloud service consumers. For example, a new startup that has limited funding can consume cloud services (e.g., hosting their services in Amazon S3). A cloud service consumer can give trust feedback of a particular cloud service by invoking the TMS.

However, MCS in the wireless environment is exposed to malicious participants deliberately contributing forged nodes and bad data. These malicious participants can also exploit these links to remove the anonymity of the volunteers and compromise their privacy. Like other networks, the security requirements in participatory sensing include services such as confidentiality, integrity, authentication, and access control to defend against malicious participants. Threats such as Sybil attack should be addressed. Therefore, identifying specific Sybil identity features in a MCS network needs to be addressed. For example, while Sybil identities compromise a MCS network, a Sybil identity will impersonate multiple identities. Hence, these Sybil identities will move in a united way because all these impersonating nodes were propagated by a single physical device. As Sybil identities move geographically, all of them will appear or disappear in the network simultaneously as the attacker moves in and out of range. This phenomenon differs from a healthy MCS network where participators are free to move at will.

Therefore, this CbTMS framework exploits Sybil attack characteristics to perform Sybil attack detection based on the following three assumptions. First, it is assumed that the MCS network traffic can be recorded in the cloud. Therefore, the normal network traffic and abnormal network traffic can be observed and analyzed. Second, it is assumed that each user and service provider who wants to participate in the system owns a unique identity, which is acquired at the bootstrapping phase from a party that is trusted by all involved parties (i.e., users, services directory provider, and service providers). Third, it is assumed that each Sybil identity uses a single-channel radio frequency; multiple Sybil identities should transmit consecutively whereas multiple independent nodes can transmit in parallel.

3.1. Passive Checking Scheme. This CbTMS framework includes a passive checking scheme (PCS) and active checking scheme (ACS) that simultaneously keep Sybil identity nodes in check, including traffic volume, signal strength, and network topology. This PCS introduces an adaptive threshold (similar to the watchdog implementation method) to identify the characteristics of Sybil attacks in MCS network. This PCS is implemented in the cloud server side and ACS is implemented in the remote client side. The PCS regularly checks the covered MCS node's conditions to decide whether the node's identity is genuine or has been compromised. The PCS will set multiple adaptive thresholds to monitor covered MCS nodes' characteristics and is implemented as part of the system operations process running on the cloud server. When a requester inquires about the trust credit of an inspector from the CbTMS framework, if the passive PCS does not detect any attack pattern on the node, it returns no attack pattern found to the requester. Otherwise, it will notify the requester to disconnect suspicious malicious node(s).

(1) Traffic Volume. Inside a base station communication range, there may be several thousand mobile devices, with

multiple applications for each device. Hence, the next step is to further classify different groups within the mobile device population with dissimilar characteristics and refine the models. Due to different devices presenting greatly different behaviors and traffic patterns, a naive extension of this model will be to develop a specialized model for every device type. The next step is to further identify groups in device population with similar characteristics and refine the models. As mentioned in our background work, once a Sybil identity has compromised a partial MCS, it will create a number of online identities and use these identities to compromise participant sensing. Therefore, by analyzing this traffic volume, signal strength, and network topology at a regular period, our CbTMS framework can infer whether the system has suspicious Sybil identities.

In our framework, the dynamic traffic of the MCS network is recorded in the cloud. It can be represented as $F = \langle F_1, F_2, \ldots, F_i, \ldots \rangle$, where F_i denotes the traffic at time i. The proposed CbTMS framework may group n entries in F into a single entry. For example, assuming $n = 2$, the new sequence for the traffic volume becomes $\langle F_1F_2, F_3F_4, F_5F_6, F_7F_8, F_9F_{10}, \ldots \rangle$. Thus, the traffic volume can be measured and analyzed with different time resolutions. Our goal is to obtain normal and abnormal traffic models from the collected sensing data. For this purpose, the k-means clustering [4], which is a well-known method for partition clustering, is applied in our framework. The k-means clustering can associate every observation with the nearest mean and hence is useful for cluster analysis, especially for a large number of variables and datasets. More specifically speaking, in this study, the k-means clustering can be used to divide the sensing data space so as to distinguish the normal and the abnormal traffic models. The intracluster heterology V has been used for measuring to select the appropriate value of k. As presented in formula (1), the value of heterology V will be calculated for increasing values of k starting from $k = 2$. Intracluster heterology is defined as

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left(x_j - \mu_i \right)^2, \quad (1)$$

where x_j is a data point residing in the ith cluster, μ_i is the centroid point of the ith cluster, S_i is the collection of all the data points residing in cluster i, and k is the number of clusters. For instance, we can group the normal network traffic volume to S_c and S_r. Now, k-means clustering has been applied to analyze and divide normal and abnormal network traffic into distinct groups. In this study, we can calculate the value of V for increasing values of k. As shown in an example in Figure 2, T8's S_c and S_r ratios are obviously different from the other groups. In this situation, the PCS can analyze the network traffic volume in the cloud DB and assume that suspected Sybil identities existed in the MCS network.

(2) Signal Strength. After the suspected Sybil identities are detected using the traffic volumes as described above, the signal strength of these suspected Sybil identities is further analyzed. The signal strength is determined by considering

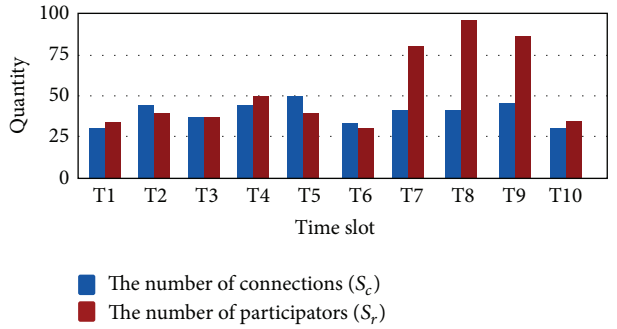

Figure 2: An example diagram of suspicious Sybil attack activity traffic volume.

the number of neighbor nodes inside a base station communication range. For example, when Sybil identities have compromised a MCS network, it will represent multiple fake identities and exchange of data among them. Fortunately, this gives our PCS an opportunity to obtain and check the signal strength of Sybil identities. However, we do not check the entire transmission signal. We only check the transmission signal from Sybil identity successfully received by its neighbor node. For example, we denoted the number S, $0 \leq S \leq 1$, as a signal-received probability that a transmission signal will be picked up by a neighbor node of a Sybil identity. Then, we denoted the number s, $0 \leq s \leq 1$, as the probability of whether this neighbor node will receive the signal. For each transmission, the transmission signal will be checked only if $s < S$.

Assume that R represents the maximum ratio difference, P_r represents received signal strength, and P_e represents expected received signal strength. Given a signal, the ratio difference r is shown in

$$r = 1 - \left(\frac{\min\left(P_r, P_e\right)}{\max\left(P_r, P_e\right)} \right). \quad (2)$$

For any signal that is received by a node, a suspicious signal can be classified if its ratio is different $r > R$. In addition, this signal strength may have precision problem because the received signal measurement result will depend on the transmitter geographical location. An example is shown in Figure 3. Figure 3(a) shows that, in an original network, there are 4 mobile nodes in the base station communication range, and Figure 3(b) shows that there are another 6 suspicious Sybil nodes when Sybil attacks occur.

(3) Network Topology. Because each Sybil group will present a similar topography map, therefore, nodes will very frequently received neighboring nodes' signals even when they are not Sybil identities and will rarely be heard apart as they normally will not move out of radio range. This phenomenon will lead to a false identification rate in topographies due to the higher density in terms of nodes per square meter. Hence, when a Sybil attacker is present in the network, the error rates for a single node spectator will be very obvious. However, in smaller topographies, as all nodes are seen as part of the same identity, there is insufficient mixing to separate

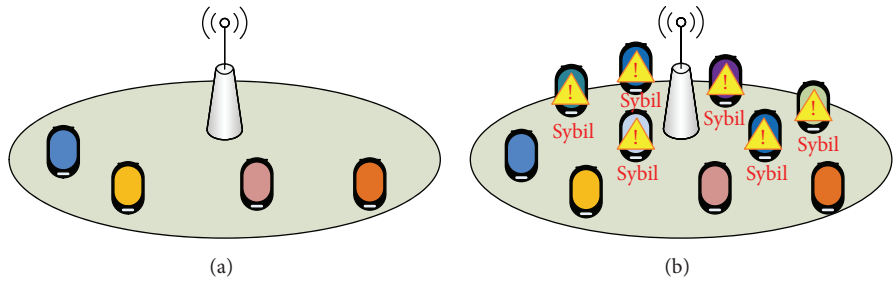

Figure 3: Illustration of suspicious Sybil attack activities in a region.

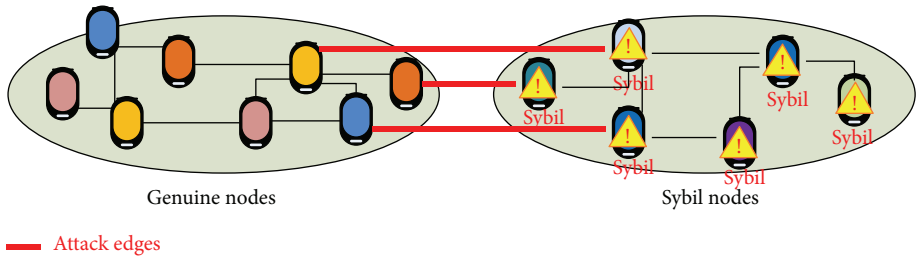

Figure 4: A conceptual network topology of Sybil attacks activities.

Sybil identities from real node, which leads to the high error rate. Once the topography size increases, the number of meaningful observations that can be made increases; and the true positive rate stays high. As the topography size increases further, the number of observations that a single node can make is reduced, as all nodes are spread far apart, and the accuracy of identifying the Sybil identities decreases. As shown in Figure 4, when Sybil attacks occur, the network topology can be conceptually divided into two parts: one consisting of all genuine identities and the other consisting of all Sybil identities. The link connecting a genuine node to a Sybil node is called an attack edge [12].

3.2. Active Checking Scheme. In this section, an active checking scheme (ACS) has been proposed to detect these Sybil nodes and eliminate them in peer-to-peer (P2P) network. The most significant feature in the P2P network is that each peer acts as both server and client. In other words, there is no central server that is used for storing the files and providing download. All nodes download files directly from other peers. Therefore, modern P2P networks suffer from the nuisance of malicious entities, such as DDoS query flooding attacks. We refer to Sybil attacks, which forge multiple identities to negatively impact or even control the entire network. A malicious Sybil identity will cheat its neighbor nodes by creating virtual nodes that are called virtual Sybil nodes. Our challenge is how to detect the Sybil identity with these virtual Sybil nodes and eliminate them to ensure the routing security while routing forwarding.

As the Sybil identity will forge fake identity and location and report its virtual location information to server nodes, it is easy for the malicious Sybil identity to forge reasonable virtual locations if the malicious node knows the location information of its neighbors. For example, we assume that a malicious Sybil identity node is O and it obtains locations of its five neighbor mobile nodes A, B, C, D, and E. Then, it can infer that the five neighbor nodes are in the concentric circles with the center O and satisfy $OA > OB > OC > OD > OE$. Nevertheless, a Sybil identity node has difficulty generating five neighboring mobile nodes' traffic volume, signal strength, and network topology from time to time. Therefore, the proposed ACS will inquire regard with its forged node identity information and hop distances from suspicious Sybil identity node. Then, ACS will pass this information to PCS scheme to do further verification. Hence, the ACS will need to cooperate with a mobile PCS that actively inquires Sybil nodes information and the PCS to verify the response information from the suspicious Sybil identity node. Once the response messages are different from PCS reservation result, then we can find out the Sybil attack area in mobile crowd sensing network and eliminate the Sybil identity in cloud server side.

3.3. Trust Credit Assessment. In the proposed framework, the trust credit of a MCS node is evaluated by Trust Credit Assessment (TCA) scheme. It is characterized by a collection of invocation history records denoted by H. Each requester node r holds a point of view regarding the trustworthiness of a mobile inspector node i in the supplication history record, which is managed by a trust management service. Each supplication history record is represented in a tuple that consists of the MCS node primary identity P, the mobile PCS identity I, a set of trust credits T, and the aggregated trust feedback weighted by the credibility T_c (i.e., $H = (P, I, T, T_c)$). Each credit in T is represented in numerical form with the

range of [0, 1], where 0, +1, and 0.5 signify negative feedback, positive feedback, and neutral feedback, respectively.

Whenever a requester node inquires the trust management service regarding the trustworthiness of a mobile PCS i, the trust result, denoted by $T_r(i)$, is calculated as

$$T_r(i) = \frac{\left(\sum_{k=1}^{l} v(k,i)\, T_c(l,i)\right)}{|v(k,i)|}, \tag{3}$$

where $v(k,i)$ is all of the feedback given to the mobile PCS i and $|v(k,i)|$ represents the length of $v(k,i)$ (i.e., the total amount of feedback given to the mobile PCS i). $T_c(l,i)$ are the trust feedback from the lth cloud consumer weighted by the credibility.

3.4. Analytical Decision Support. In this section, we applied the well-known and widely used C5.0 decision tree algorithm, which was an improved version of C4.5 [16]. In the last few decades, several decision tree learning algorithms have been proposed including Ross Quinlan who invented the Iterative Dichotomiser 3 (ID3) [13] which was used to generate a decision tree from a dataset, as well as Naive Bayes Filter [14], J48 [15], and C4.5 [16] decision tree learning models that can be applied in many applications which provide a very fast statistical method to classify data. The decision tree can use various machine learning algorithms (MLAs) for providing informative diagnosis models according to data features or rules to solve classification problems. The aim of decision tree is to identify numerous ways of splitting a dataset into branch-like segments. This branch-like segment can produce a relationship model on the basis of the data collected from different sources. One of the most noteworthy characteristics of decision trees is represented in the form of classification (if-then) rules. Each rule represents a unique path from the root to each leaf. In operations research, precisely in decision analysis, a decision diagram is a decision support tool.

Different from previous decision tree algorithms, the C5.0 classifier comprises a simple command-line interface, which can generate the decision trees and rules and finally test the classifier. In numerous applications, rule-sets are preferred because they are simpler and easier to understand than decision trees, but compared to C5.0, C4.5's rule-set methods are slow and memory-hungry. As described in [16], C5.0 [17] algorithm has been recognized as an efficient data mining technique compared with C4.5 algorithm. Because C5.0 represents a new algorithm for generating rule-sets, the improvement is substantial. C5.0 model works by splitting the sample based on the field that provides the maximum information gain [18]. The information gain is computed to estimate the gain produced by a split over an attribute.

Let S be the sample:

(i) C_i is Class I, $i = 1, 2, \ldots, m$, $I(s_1, s_2, \ldots, s_m) = -\sum p_i \log_2(p_i)$.

(ii) S_i is the number of samples in Class I. $P_i = S_i/S$; $\log_2$ is the binary logarithm.

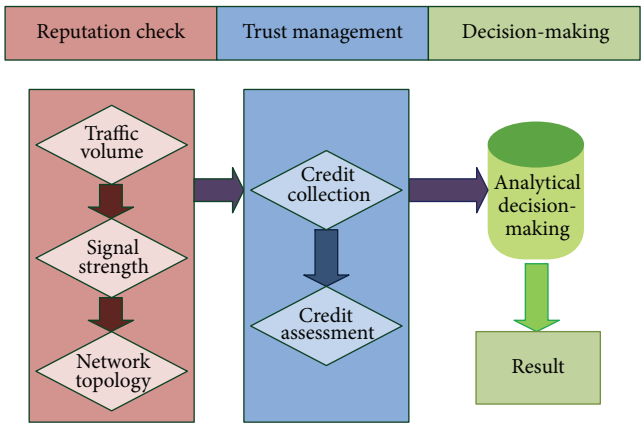

Figure 5: Hybrid reputation monitoring diagram.

(iii) Let Attribute A have v distinct values:

$$\text{Entropy} = E(A) = \sum \left\{ \frac{\left(S_{1j} + S_{2j} + \cdots + S_{mj}\right)}{S} \right\} * I\left(s_{1j}, \ldots, s_{mj}\right), \quad j = 1. \tag{4}$$

(iv) S_{ij} is samples in Class i and subset j of Attribute A:

$$I\left(S_{1j}, S_{2j}, \ldots, S_{mj}\right) = -\sum p_{ij} \log_2\left(p_{ij}\right). \tag{5}$$

(v) $\text{Gain}(A) = I(s_1, s_2, \ldots, s_m) - E(A)$.

(vi) Gain Ratio then chooses, from among the tests with at least average gain, the Gain Ratio = $P(A)$:

$$\sum_{i}^{t} \frac{S_i}{S} \log\left(\frac{S_i}{S}\right). \tag{6}$$

(vii) $\text{Gain Ratio}(A) = \text{Gain}(A)/P(A)$.

3.5. An Example of the Scenario. As this attack has no relation to the identification scheme, we do not further evaluate it. On the other hand, an attacker can utilize Sybil attacks to compromise and control a genuine node. The compromised genuine node will be considered as a Sybil node and not as a genuine node. This Sybil node will focus on creating multiple online user identities called Sybil identities and try to achieve malicious results through these identities. As shown in Figure 5, we will implement our CbTMS algorithm in three phases. In the first phase, the cloud server-side manager will record network traffic to those who participate in the system and define multiple adaptive thresholds, including traffic volume, signal strength, and network topology, to evaluate network trustworthiness.

When a Sybil identity uses a single-channel radio and has been identified as exceeding the adaptive threshold range in our PCS, the PCS module will generate a notification

to the TCA. Then, the TCA will draw these PCS history records from its database and process the credit assessment. Once the Sybil attack pattern has been preliminarily identified, it will enable the analytical decision-making (ADM) to further analyze and determine the Sybil attacks in this network. This framework will check regular network and system statistics and use an adaptive threshold to achieve network trustworthiness. To improve the completeness of the analysis by observing how a Sybil identity behaves in participatory environments, it will require cooperation with telecommunication service cloud providers. In this cloud, we can develop a subset of system calls invoked by the analyzed program in a mobile user environment and receive the result of the computation.

4. Experimental Evaluations

In this section, the proposed algorithm cloud based trust management scheme (CbTMS) has been simulated in NS2 [19]. The main focus of this paper is to simulate Sybil node mobility model, which has been well used in many different application areas. NS2 is a popular and reliable network simulator tool based on C++ and OTcl programming languages, which were developed by UC Berkeley. It provides an open source and collaborative environment to support protocol design and network traffic studies. NS2 has extended version that provides mobility model and adds more supporting features such as sending and receiving packets over wireless channel and provides radio propagation model, MAC protocols, interface queue, link layer configuration, ability to move within a given topology, and ad hoc on-demand distance vector (AODV) routing which is suitable for setting our simulation parameters. This simulation environment considers an urban road scenario with two lanes in each direction. In this scenario, vehicles are placed on the road randomly with the a minimum 5-meter intervehicle space in each lane. Vehicles travel on the road with speeds of predefined formula (miles/hour). The communication range of benign vehicles is 50 meters while the malicious vehicle may adjust its transmission power according to the situation. We inspect the detection time, which is defined as the time interval from when the malicious vehicle starts Sybil attacks to neighboring nodes and when it is identified by other vehicles.

The proposed CbTMS framework has been implemented in NS2 based on the Internet framework and utilizes AODV algorithm and interface queue type Queue/DropTail/PriQueue model for mobility of the nodes because this model can well depict a real world situation and successfully simulate work. This mobility model is based on an entity mobility model where the nodes move independently of each other. The simulation work has taken the parameters for implementation as shown in Table 1 and Figure 6.

4.1. Malicious Sybil Node and Compromised Node Selection. Based on previous sections, the described malicious Sybil identities will be exposed to like mobile malicious participants that can deliberately contribute forge nodes and bad data. In our simulation experience, Sybil identities were designed to modify packet contents and participated in route discovery and route maintenance. They will not forward packets to neighbor mobile nodes, but only to specific compromised nodes. Hence, the packet routing paths will be the same even when new formal nodes join the routing process. Moreover, when the Sybil identity has compromised its neighbor nodes, they will have the same mobility model. Furthermore, in a Sybil attack, the selection of compromised nodes based on detecting node misbehavior was done in a random manner. These compromised nodes will have a random number generator inside them so that every time they will need to see its value before overhearing the channel. If the random number was evaluated as 0, then they will turn on their compromised mode to forward the malicious message to their neighbor nodes or else they have to remain idle. This idle state will also result in a lot of power saving of the compromised nodes without affecting the fault detection.

Table 1: Simulation implementation parameter list.

Parameter	Value
Simulation environment	Wireless channel
Radio propagation model	Two-ray ground
Network interface type	WirelessPhy
MAC type	802.11
Routing protocol	AODV
Interface queue type	Queue/DropTail/PriQueue
Link layer type	LL
Antenna model	OmniAntenna
Max. packet in IFQ	50
Simulation area	1000 m∗800 m
Simulation time	150 s
Number of nodes	18

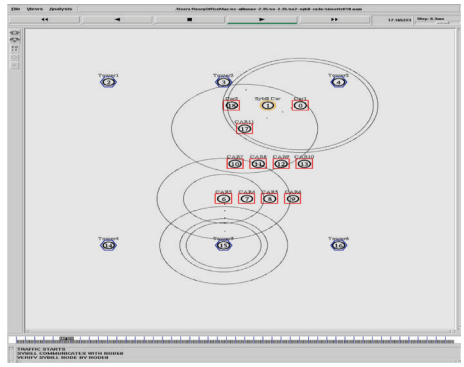

Figure 6: Simulation graphical view of nodes.

4.2. Ad Hoc On-Demand Distance Vector Routing. As described in [20], in AODV routing protocol, the routing agent will cache the packet first and broadcast a request to try to find a route. Once the packet has reached link layer, the link layer looks up ARP table to map IP address to MAC address, and then it delivers packet to interface queue. Wireless MAC will be used to avoid collision and if a

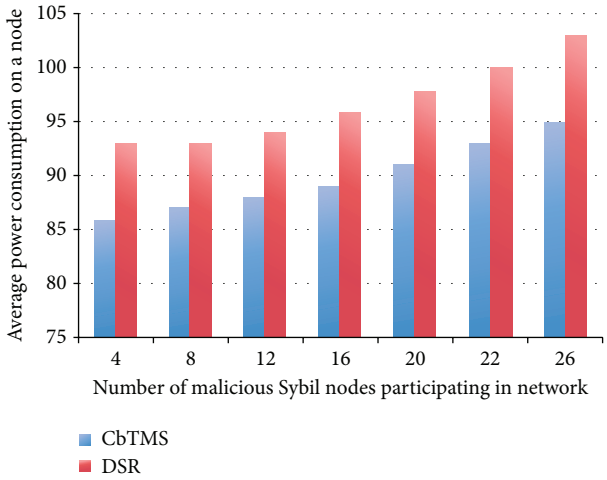

Figure 7: Average power consumption comparative diagram.

packet can be transmitted safely, it will be handed to network interface. The network interface simulates the smartphone wireless behavior in the real world. Finally, the modulated data will be transmitted over the wireless channel.

4.3. Experiment Results. Ideally, the average power consumption for a MCS node mode is 89.85 Wh as defined normal mode as shown in Figure 7. The Wh is a unit of energy equivalent to one watt of power expended for one hour of time. On the other hand, in a malicious Sybil node attack mode, the average power consumption is much higher than in a normal mode while utilizing AODV routing protocol. The simulation result shows that each node will consume 96.72 Wh on average. In the case of AODV routing protocol, it is based on the nodes having to cooperate to find a path between nodes. It allows nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. Therefore, it will consume enormous power in the network in our malicious Sybil node mode. But compared with the proposed CbTMS algorithm, we will detect malicious Sybil node and compromised nodes to prevent communication overhead. In addition, in our simulation setup, there is only 1 node that has been set up as Sybil identity node. The proposed CbTMS provides lower power consumption which causes 6.87 Wh reduction compared with malicious Sybil node attack mode while utilizing AODV routing protocol.

5. Conclusion

In this paper, a cloud based trust management scheme (CbTMS) was proposed for detecting Sybil attacks in mobile crowd sensing (MCS) networks. Sybil attacks create multiple online user identities called Sybil identities and try to compromise systems with their malicious information through these identities. The proposed CbTMS framework can perform trust management and reputation checker to verify the nodes in the MCS network. It combines two schemes, namely, Characteristics Checking Scheme (PCS) and Trust Credit Assessment (TCA), to detect suspicious Sybil nodes. PCS was proposed for passively monitoring the characteristics of the suspicious Sybil nodes, including time, density, and topology in the MCS, whereas TCA was proposed for evaluating the trustworthiness of the suspicious Sybil nodes. Our simulation studies show that our CbTMS can efficiently detect the malicious Sybil nodes in the network and cause 6.87 Wh reduction compared with malicious Sybil node attack mode.

Competing Interests

The authors declare that they have no competing interests.

References

[1] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 29–35, 2014.

[2] T. Denning, A. Andrew, R. Chaudhri et al., "BALANCE: towards a usable pervasive wellness application with accurate activity inference," in *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile '09)*, vol. 5, pp. 15–16, Santa Cruz, Calif, USA, February 2009.

[3] E. P. Stuntebeck, J. S. Davis II, G. D. Abowd, and M. Blount, "HealthSense: classification of health-related sensor data through user-assisted machine learning," in *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications (HotMobile '08)*, pp. 1–5, February 2008.

[4] R. Douceur, "The Sybil attack," in *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, Cambridge, Mass, USA, March 2002.

[5] X. Liang, X. Lin, and X. S. Shen, "Enabling trustworthy service evaluation in service-oriented mobile social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 310–320, 2014.

[6] K. Zhang, X. Liang, R. Lu, and X. Shen, "Sybil attacks and their defenses in the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 5, pp. 372–383, 2014.

[7] I. Brandic, S. Dustdar, T. Anstett, D. Schumm, F. Leymann, and R. Konrad, "Compliant Cloud Computing (C3): architecture and language support for user-driven compliance management in Clouds," in *Proceedings of the 3rd IEEE International Conference on Cloud Computing (CLOUD '10)*, pp. 244–251, July 2010.

[8] K. Hwang and D. Li, "Trusted cloud computing with secure resources and data coloring," *IEEE Internet Computing*, vol. 14, no. 5, pp. 14–22, 2010.

[9] C. Ghali, A. Chehab, and A. Kayssi, "CATRAC: contextaware trust- and role-based access control for composite web services," in *Proceedings of the 10th IEEE International Conference on Computer and Information Technology*, pp. 1085–1089, Bradford, UK, 2010.

[10] J. Grover, M. S. Gaur, V. Laxmi, and N. K. Prajapati, "A sybil attack detection approach using neighboring vehicles in VANET," in *Proceedings of the 4th International Conference on Security of Information and Networks (SIN '11)*, pp. 151–158, Sydney, Australia, November 2011.

[11] S. Ries, "Extending Bayesian trust models regarding context-dependence and user friendly representation," in *Proceedings of the 24th Annual ACM Symposium on Applied Computing (SAC '09)*, pp. 1294–1301, ACM Press, Honolulu, Hawaii, USA, March 2009.

[12] S.-H. Chang and T.-S. Huang, "A fuzzy knowledge based fault tolerance algorithm in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '12)*, pp. 891–896, IEEE, fukuoka, Japan, March 2012.

[13] D. Jiang, *Information Theory and Coding*, Science and Technology of China University Press, 2001.

[14] S. F. Chen and Z. Q. Chen, *Artificial Intelligence in Knowledge Engineering*, Nanjing University Press, Nanjing, China, 1997.

[15] A. P. Muniyandi, R. Rajeshwari, and R. Rajaram, *Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree Algorithm*, vol. 30, Elsevier, New York, NY, USA, 2012.

[16] *Is See5/C5.0 Better Than C4.5?*, 2009, http://www.rulequest.com/see5-comparison.html.

[17] Information on See5/C5.0—RuleQuest Research Data Mining Tools, 2011, http://www.rulequest.com/see5-info.html.

[18] X. Zhu, J. Wang, S. Wu, and H. Yan, "Research and application of the improved algorithm C4.5 on decision tree," in *Proceedings of the International Conference on Test and Measurement (ICTM '09)*, pp. 184–187, IEEE, Hong Kong, December 2009.

[19] S. McCanne and S. Floyd, "Network Simulator Version 2," http://www.isi.edu/nsnam/ns.

[20] Z. Wang, Y. P. Chen, and C. Li, *Implementation of the AODV Routing Protocol in ns2 for Multi-Hop Wireless Networks*, Memorial University, 2010.