Channel State Information-Based Detection of Sybil attacks in Wireless Networks

Chundong Wang[1], Likun Zhu[1]*, Liangyi Gong[1], Zhentang Zhao[1], Lei Yang[1], Zheli Liu[2], and Xiaochun Cheng[3]
[1]Tianjin University of Technology, Tianjin 300384 China
{michael3769, kurtcobian4ever}@163.com, gongliangyi@tjut.edu.cn,
deviltangv@163.com, 778750188@qq.com
[2]Nankai University, Tianjin 300350 China
liuzheli@nankai.edu.cn
[3]Middlesex University, London NW4 4BT UK
X.Cheng@mdx.ac.uk

Abstract

Single authentication mechanisms and broadcast characteristics of wireless networks make the Access Point (AP) vulnerable to spoofing attacks and Sybil attacks. However, Sybil attacks seriously affect network performance. Sybil nodes act with different identity, and prevent the normal clients from transmission. In this paper, a self-adaptive MUSIC algorithm is proposed, which improves the accuracy of the angle of the indoor wireless device by eliminating the phase offset in channel state information (CSI), and designs different types' detection algorithm of Sybil attacks and spoofing attacks based on different Sybil attack models. And we experiment on mobile and commercial WiFi devices. The average detection error of angle is below 6.3°. After combining analysis of received signal strength indicator (RSSI), our detection algorithm can effectively detect whether the nodes launched by Sybil attacks, and the identity of other clients disguised by spoofing attacks. According to the experimental results, the scheme can distinguish the Sybil clients and the normal clients accurately, and the average success rate of the Sybil attack detection system is 98.5%.

Keywords: Channel State Information, Sybil attack, Spoofing Attack, Indoor localization

1 Introduction

The key factor in secure communication of wireless networks is whether or not they can be effectively secured without eavesdropping. There are broadcasting natural characteristics of wireless channel. The one-way authentication APs can't effectively resist the attack on the forged identity of the wireless networks, although equipped with 802.11i security protocol[12]. One serious consequence is that the network is susceptible to various identity-based attacks such as Sybil attacks[6]. As a result, wireless networks physical layer information[31, 14] has attracted a lot of attention in the absence of upper data encryption, while the MIMO technologies[18] in the latest 802.11n and 802.11a/c WiFi standards provide richer and more accurate information of the wireless networks channel characteristics.

Sybil attacks were first discussed by J. R. Douceur[5], where the attacker (Sybil node) tries to forge multiple identification in a certain region. A malicious node renders multiple nodes, stats other nodes, and falsifies the identity of the network that does not exist in the network. Analogously in spoofing attacks[3, 24], a malicious device claims to be a specific client or AP other than itself. Sybil nodes

Journal of Internet Services and Information Security (JISIS), volume: 8, number: 1 (February 2018), pp. 2-17
*Corresponding author: Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, 300384 Tianjin, China, Tel: +86-130-7721-7616

may block communication in the network, the forged node will send a large number of high-speed data packets, and prevent the normal node to communicate with the AP of sever. As a special denial of service attack[27], the attack seriously threatens the performance of the wireless network system. Sybil attacks have been studied extensively in Wireless Sensor Networks[19]. There are two kinds of solutions, encryption authentication and non-encryption mechanism.

The former uses a shared key for authentication[2, 30, 1], the node needs to store the common key, traffic and storage capacity is too large, excessive energy consumption, shorten the life of the networks. The latter is mainly through the combination of RSSI, to determine the location of the node. Murat Demirbas et al.[4] proposed a Sybil attack detection based on RSSI. The Sybil node is detected by comparing the RSSI ratio of the two identities at the four detection nodes. It can detect the Sybil attack but the detection nodes are four and only after the network is already built. The literature [20] stores the identity information, RSSI and location information of the neighbor node in the table of the head node. This method detects the Sybil attack by comparing the current RSSI value of the neighbor nodes and the threshold in the table. But the accuracy of this detection method is less than 20m and use too many detection nodes. Liang X et al.[23] have proposed a channel-based authentication technique to detect Sybil attacks in wireless networks, utilizing the uniqueness of channel responses in rich-scattering environments. However, channel-based detection relies on the limited spatial information associated with channel path loss.

Our method combines CSI and RSSI to distinguish clients and Sybil nodes. After that, we present a detection algorithm based on different Sybil attack models. The multiple signal classification (MUSIC)[15] algorithm is widely employed to estimate the arrive of angle (AoA) of incoming signals. Due to the presence of sampling frequency offset (SFO) and sampling time offset (STO)[16], the CSI generates phase offset and line-of-sight path time offset. In order to eliminate the offsets, we improve the traditional MUSIC algorithm[28, 7] and propose a self-adaptive MUSIC algorithm, which can enhance the accuracy of clients' angle effectively. Based on this, we don't use multiple AP for clients positioning, we only use one AP on the existence of the clients positioning, which is different from most CSI-based localization algorithms. Meanwhile we combine AoA technology with signal strength RSSI to achieve effective positioning of real clients. Then turn off the attack on the Sybil clients. We use the mobile as a client to detect the angle at the beginning. Then, we turn to commercial WiFi devices whose angle is detected at different angles and compare the accuracy of angles in different AoA algorithms. Finally, we emulate that the Sybil nodes are attacked at different angles and different positions. Once the physical location of the node is determined, no matter how the intrusion client tampers or falsifies the virtual Sybil clients, it can be easily detected.

The structure of this paper is as follows: In Section 2. we introduce the CSI and Sybil attack Models in detail. In Section 3, we use the traditional MUSIC algorithm and CSI to calculate the phase, and propose self-adaptive MUSIC algorithm to eliminate the angle deviation. We propose our Sybil attack detection algorithm in Section 3.3. We then present simulation results and experimental results in Sections 4. Finally, we draw a short conclusion, in Section 5.

2 CSI and Attack Models

2.1 Channel State Information

Modern WiFi NICs measure the wireless channel for every received WiFi frame in order to decode the payloads of the frame [15]. In the time domain, the received signal $r(t)$ is the convolution of the transmitted signal $s(t)$ and the channel impulse response (CIR) $h(t)$. According to the convolution theorem,

the signal frequency domain responds as in formula 1.

$$R(f) = S(f) \times H(f) \tag{1}$$

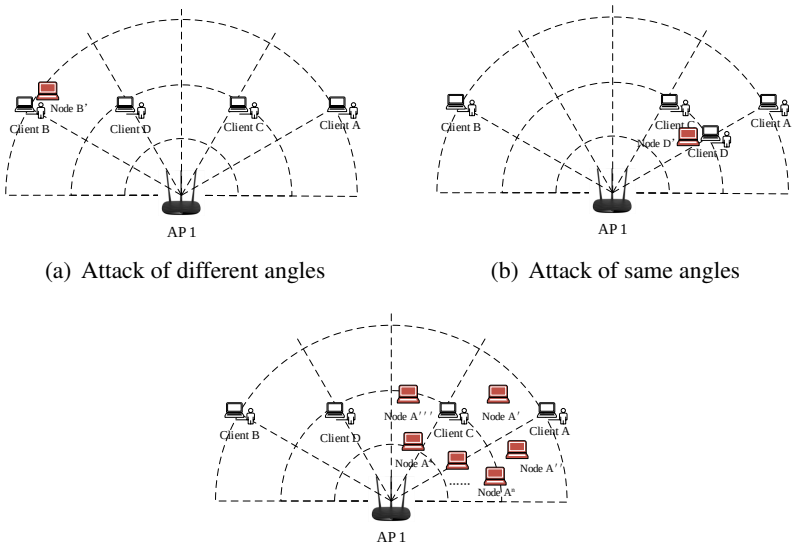
$R(f)$ is the received signal spectrum, $S(f)$ is the transmit signal spectrum and $H(f)$ is called the channel frequency response (CFR). CFR values are reported by WiFi NICs in the form of CSI measurements, where each CSI measurement contains S matrices with dimensions of N_{TX}, N_{RX}. S is the number of OFDM subcarriers and N_{TX}, N_{RX} are the number of transmitting/receiving antennas, respectively. Each entry in the CSI matrix is a CFR value between a pair of transmitting/receiving antenna at a given OFDM subcarrier for one received WiFi frame. Usually, CSI are measured at $S = 30$ subcarriers and we call the time-series of CFR values on a given subcarrier for a given antenna pair as a CSI stream in this paper.

The traditional way of obtaining CFR is measured by using a unique device Vector Network Analyzer. Use the software radio platform (SDR)[13] to implement and modify the 802.11 protocol . Recently, Halperin [8] obtain the the channel state information (CSI) by modifying the Intel 5300 network card driver from the ordinary WiFi device. In 802.11a/g/n, the CSI is the sampled version of the channel frequency response (CFR). Meanwhile, Atheros CSI tool proposed by Yaxiong[11] is supposed to all types of Atheros 802.11n WiFi chipset now. Each CSI indicates the magnitude and phase information of the subcarriers in the channel: Where $H(f_k)$ is the CSI of the subcarrier with the center frequency f_k. Compared with the traditional RSSI, CSI not only provides the subcarrier amplitude information, but also provides the subcarrier phase information, and RSSI is simply the strength of the information. CSI provides the amplitude and phase information of subcarriers with multiple frequencies, and RSSI is only one dimension. Since the subcarriers of different frequencies are not exactly the same in space, it is possible to better characterize the space by CSI and improve the accuracy of WiFi Localization.

Z Yang[26] give the channel response characteristics in detail from the RSSI to CSI in the multi-path environment, and analysis of the CSI in the time domain in different aspects. It laid a deep theoretical foundation for further research on indoor localization. Hence in recent years, most author use fine-grained CSI to estimate location [22, 29, 21, 26]. Array track [25] achieves a precise positioning of the human by 6 AP with 16 antennas. The localization algorithm and spatial smoothing technique [17] have a strong guiding value for indoor AoA localization. But they use too many transmitters and receivers. Considering the cost of the actual indoor wireless network layout, we think multiple nodes to detect Sybil attacks is clearly too redundant. Spotfi [9] estimates ToF and AoA of WiFi devices, achieving a median accuracy of 40 cm. Since this paper is mainly to detect the existence of Sybil attacks, we only need to confirm whether there are Sybil nodes or not. Therefore, there is no use for geographic location parameters of all clients, but we need to confirm the Sybil nodes' angle for recognition by our self-adaptive MUSIC algorithm. Experiments show that it achieves high-precision angles and distinguishes line of sight (LoS)[32] angles accurately. Whereas Dynamic-MUSIC[10] can identify the human target's angle of mobile path for localization without any devices indoor. In this paper, we only need to use one witnessing AP. It can detect the angle of each client by its physical layer information. We apply it in 5G monitor mode to get phase information.

2.2 Sybil attack Models

In wireless network, most clients have a fixed position. Sybil attacker may forge the identity of normal clients, such as IP addresses, Mac addresses, or public keys. We present generalized Sybil attack Models in Fig 1. We assume that there are four clients A, B, C, D in the indoor wireless network, which are located at $[30^\circ, 150^\circ, 60^\circ, 120^\circ]$ of the witnessing AP 1. And client A, client B located in 3 m radius of the circle. Similarly, client C, client D located in 2 meters radius of the circle. AP 1 is the center of these circles. It receives service requests from those clients in a period of time.



(a) Attack of different angles

(b) Attack of same angles

(c) A large number of Sybil nodes attack

Figure 1: Sybil attack models

In Fig 1(a), we assume client A is a Sybil client, which located in 30° of AP 1 and steals the identity of the client B. When client A forged the client B's ID as node B', client A send massive data to AP 1. AP 1 think the data are from client B, and exchange massage with it. That is spoofing attack, and node B' is Sybil node. Spoofing attacks may also occur at different radians of the same angle above the radius. Unlike figure 1, client A,client D are located in the same degree 30°. Fig 1(b) shows that when client A steals the identity of client D, AP 1 claims node D' for the real node so that node D' can accept all the messages of client D.

As shown in Fig 1(c), it is also assume that client A is a Sybil client. Client A claims that it has multiple identities such as node A', node A'', node A''' and so on, which can consume the resources of witnessing AP 1 by sending massive high-rate service requests. Client B, C, D will fail to access the network especially when the Sybil node is large enough. This is the typical generalized Sybil attack. Simultaneously, those nodes forged by client A are Sybil nodes.

We present all of Sybil attack models in the indoor environments. If the attack happens in the outdoor, it will be easy to distinguish by the sharp declines in RSSI. Whereas, constrain of Sybil attacker is that all the identities are part of the same physical device, they must appear or disappear simultaneously. Hence we can propose our Sybil attack detection algorithm in Section 3.3 in detail. In wireless networks, there may be more than one AP, each AP can receive the request from the clients, we can also detect the witnessing AP attacked by Sybil in our algorithm.

3 System

Our work is based on the following three steps:

1. Collect the channel state information by center AP and estimate the angle of arrival of all the clients at different locations and different devices through traditional MUSIC algorithm.

2. After eliminating the phase offset and sampling time offset, we apply the spatial smoothing MUSIC algorithm to obtain an accurate line-of-sight path angle while comparing the sum of the three

channel RSSIs to distinguish between different angles of the clients and the APs.

3. Use above information to compare the real location of different nodes in the wireless network. According to different cases of Sybil attack and spoofing attack, we determine whether it is Sybil node by our algorithm.

3.1 Traditional MUSIC Algorithm

In the indoor environment, due to the presence of objects such as walls, furniture, and people, the transmission signal will produce reflection during the propagation process, but the Los path can be used as the shortest propagation path for the shortest time. In order to better understand our self-adaptive music algorithm, the traditional music algorithm of Direction Of Arrival(DoA) implementation principle will be of great help. After this, we will briefly introduce the traditional MUSIC algorithm and make our improvements.

The basic idea of the multiple signal classification (MUSIC) algorithm[15] is to characterize the covariance matrix of any array of output data to obtain the signal subspace and the noise subspace. Finally, the orthogonality of these two subspaces is used to construct the spatial spectral function, and the angle of the signal is detected by spectral peak search.

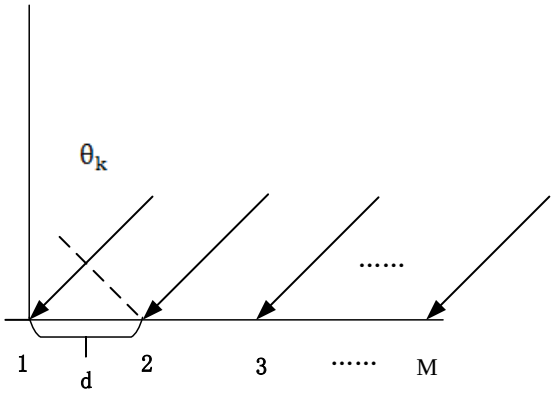


Figure 2: MUSIC algorithm DOA model

The basic AoA estimation model is shown in Fig 2. Since the singal has to travel further to arrive at each different antenna. Hence there is phase shift. We can get the phase of the two antenna signals when arriving at different antennas through the CSI tools citehalperin2011tool so that the calculated phase difference can be calculated by the formula 2:

$$\Delta\phi = -2\pi\frac{dcos(\theta)}{\lambda} \tag{2}$$

Where θ is the breadth of the angle we need. d is the distance between two antennas. Through the signal frequency f, we can get the signal wavelength λ, so the wave angle θ can be easily calculated:

$$\theta = \arccos\left(\frac{\Delta\phi\cdot\lambda}{-2\pi d}\right) \tag{3}$$

As the multipath effect of the indoor environment, the multipath signal makes the calculation of the radar angle completely inaccurate. By increasing the number of antennas and using the 30 subcarriers

of the CSI signal, we can use this characteristic to further eliminate multipath effects. Assuming that the antenna array is M and the number of signals is D, the received signal $x(t)$ can be expressed as

$$\mathrm{x}(\mathrm{t}) = [\mathrm{a}(\theta_1)\,\mathrm{a}(\theta_2)\cdots\mathrm{a}(\theta_D)] \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_D(t) \end{bmatrix} + n(k) \tag{4}$$

$$[\mathrm{a}(\theta_1)\,\mathrm{a}(\theta_2)\cdots\mathrm{a}(\theta_D)] = \begin{bmatrix} 1 & 1 & & 1 \\ e^{-j\Delta\varphi_1} & e^{-j\Delta\varphi_2} & \cdots & e^{-j\Delta\varphi_D} \\ \vdots & \vdots & & \vdots \\ e^{-j(M-1)\Delta\varphi_1} & e^{-j(M-1)\Delta\varphi_2} & & e^{-j(M-1)\Delta\varphi_D} \end{bmatrix} \tag{5}$$

Where $\mathrm{a}(\theta_i)$ is the steering vector, corresponding to the different antenna in the receiver at different wave angle phase shift $\Delta\phi_k$. $n(k)$ is an additive Gaussian white noise with mean of 0, the variance of σ^2, so the autocorrelation function of the received signal can be expressed as:

$$\begin{aligned} R_X &= E\left(xx^H\right) \\ &= E\left[(AS+N)\left(AS+N\right)^H\right] \\ &= AE\left[SS^H\right]A^H + E\left[NN^H\right] \\ &= AR_{ss}A^H + \sigma^2 I \end{aligned} \tag{6}$$

Where R_{ss} corresponds to our signal subspace, and $\sigma^2 I$ corresponds to the noise subspace. We obtain the M eigenvalues and the corresponding eigenvectors of the R_X of the autocorrelation function. Usually the noise is much smaller than the signal, and we sort the known eigenvalues. From equation 6 we can see that there are M eigenvalues, where M-D eigenvalues corresponds to the signal and the smaller D eigenvalues corresponding to noise. We know that the eigenvectors corresponding to the noise eigenvalues are orthogonal to the column vectors of the matrix A. The columns of A correspond to the direction of the signal. So the noise feature vector E:

$$\mathrm{E} = [v_{D+1}, v_{D+2}, \cdots v_M] \tag{7}$$

Define the spatial spectral function:

$$\mathrm{P}_{MU}(\theta) = \frac{1}{a^H(\theta)EE^H a(\theta)} \tag{8}$$

In the formula 8, the denominator is the inner product of the signal vector and the noise matrix. When the $a(\theta)$ is orthogonal to the E columns, the denominator is zero, and the minimum value is actually taken due to the presence of noise. $\mathrm{P}_{MU}(\theta)$ has a spike. By traversing the different angles θ, we perform the spectral search, we can find DoA. In this experiment, we use M = 3 transmit antennas, D = 1 AoA signal, we find the covariance of the signal. The experiment shows that we use the covariance mean R_{xx} of 30 subcarriers to be higher than the covariance R_X of the individual signals stability, resulting in a more stable signal AoA. If the received signal has a component of the LoS path, the LoS path has the strongest signal strength value, then the peak of the signal's pseudo-spectrum is the DoA of the signal on the LoS path. However, we find that traditional MUSIC algorithm may have big error because of the phase offset on CSI. It was caused by the radio frequency link connected to the RF oscillator. To reduce the phase offset, we present a self-adoptive MUSIC in next section.

3.2 Self-Adaptive MUSIC Algorithm

In a standard WiFi network, the transmitter and receiver are not time-synchronized, so they are not synchronized to the sampling clock at the DAC and ADC. The sampling time offset (STO) will deviate from the arrival time and phase of the LoS path and NLoS path, which cause that the direct path of AoA maybe not the real wave of the angle. The non-line-of-sight path may be determined as the line of sight and produce a offset. Meanwhile, when transmitting device RF link connection to RF oscillator, it will generate sampling frequency offset (SFO). The SFO changes the sampling time offset from the packet of the same packet, which in turn leads to another additional time offset of the ToF estimate across the packet and unknown phase β.

The total time delays Δt caused by STO and SFO are very different across two packets' transmissions between a pair of transmitter and receiver. Hence, the phase of the i-th subcarrier can be expressed as $\widehat{\phi}_i = \phi_i + 2\pi f_i \Delta t + \beta + N_{\mathrm{i}}$ is the genuine channel response phase we are searching for and N_{i} is some measurement noise. we use the classification algorithm to ϕ_i, and the value of Δt and β we are always the same. Suppose $\phi_{\mathrm{i}}(m,k)$ is the unwarpped phase of the CSI at the k^{th} subcarrier of the i^{th} packet received at m^{th} antenna,we can obtain the optimal linear fit of the phase for the i^{th} as:

$$\widehat{\tau}_{\mathrm{i}} = \arg\min_{\Delta t} \sum_{m,k=1}^{M,K} (\phi_{\mathrm{i}}(m,k) + 2\pi f_i (k-1)\Delta t + \beta) \tag{9}$$

The $\widehat{\tau}_{\mathrm{i}}$ is time delay of the i^{th} packet, and we can get modified CSI phase $\phi_i(m,k) = \widehat{\phi}_i(m,k) - 2\pi f_{\mathrm{i}}(k-1)\widehat{\tau}_{\mathrm{i}}$. Our method is similar to Sanitizing ToF Estimates in Spotfi and this method eliminates the residual synchronization error in the CSI phase. However, this method destroys the independence of the subcarrier CSI phases so that the processed result can only be used to the digital signature, and the phase deviation of the SFO is not effectively eliminated. In order to eliminate the phase offset of the SFO and obtain a more accurate angle value, we apply the self-adaptive MUSIC algorithm to calculate the phase offset between the antennas.

We assume that the phase deviation between the antennas is $\langle \delta_0, \delta_1 \rangle$. Because $\langle \delta_0, \delta_1 \rangle$ is a hidden random variable, we can not directly get the size of these two variables, so this algorithm uses the search method. It selects all combinations of $\langle \delta_0, \delta_1 \rangle$ that can be used to achieve the best combination of the reception effect as our estimated antenna deviation. We verify that we estimate the estimated antenna deviation by a number of experiments.

Because we selected a better calibration results from a combination of $\langle \delta_0, \delta_1 \rangle$. The true value of the angle is not exactly equal to the value we measured. In addition, it is possible that all combinations do not allow the peak of the pseudo spectrum to be exactly equal to the measured value. To improve the stability of the evaluation system, we need to choose a robust evaluation function to evaluate the pseudo spectrum of each combination.

- When the peak of the dummy spectrum is equal to the measured DOA value, the combination has the greatest probability of being the best calibration.
- When the peak of the dummy spectrum is close to the measured DOA value, the combination also has a larger probability of becoming the best calibration combination.

Consider the above two points, we designed our evaluation function $\eta(\rho)$, $\rho$is the pseudo spectrum, as follows:

- Find a normalizing constant k such that $\rho' = k\rho$ is one, and set $\int k\rho(\theta)\, d(\theta) = 1$.
- Construct a Gaussian mask $\mathrm{g}_\alpha(\theta)$ with an expected value α and a variance according to the desired level of error tolerance. Set $\bar{\mathrm{g}}_\alpha(\theta) = 1 - \mathrm{g}_\alpha(\theta)$.

- Calculate $\mu(\rho) = \frac{g_\alpha(\theta)\rho'(\theta)d\theta}{\bar{g}_\alpha(\theta)\rho'(\theta)d\theta}$

Our algorithm estimates the best deviation for each packet and then calculates the offset of all packets. We will generate two phase deviation values $\langle \delta_0, \delta_1 \rangle$ in the multiple clustering algorithm and select the most frequent phase deviation combination as our final estimate. Calibrate the antenna 2 and the antenna 3 while applying the calibrated CFR to the traditional MUSIC algorithm. This inherent deviation occurs when the device is started and does not change when the device is running, but the phase deviation is reset when the device is restarted. We apply this algorithm to the spatial smoothing MUSIC algorithm to automatically correct each phase offset at the start of the device.

At this point, we eliminate the time offset in the different CSI packets generated by the STO, make the direct path more accurate, and evaluate the optimal phase offset by the self-calibration method, which is generated by the SFO. In the fourth quarter, a large number of experiments show that for different angles of the AP, self-calibration algorithm can effectively detect its position.

What we only want is the direct path. So we therefore adopt spatial smoothing to get the direct path accurately. Though our witnessing AP only has 3 antennas array, it can detect two AoA in the indoor environment. So we group our antennas into 2 group $\{\text{antenna}1,2;\text{antenna}2,3\}$ as input singal to detect the direct path. In detail, spatial smoothing generating signals x_1,x_2,x_3 would output two signals $\dot{x}_1$, $\dot{x}_2$ where $\dot{x}_i = \frac{1}{2}(\dot{x}_1 + \dot{x}_2), i = 1,2$. At last, we implemented two algorithms only using a three-antenna service AP, our algorithm compared to Phaser[7] and ArrayTrack[25]with fewer antenna arrays, while achieving the same average error of 6.3°. Accurately enhance 5.6° than traditional algorithms. Similarly, compared with the SpotFi[9] with a certain complexity, we do not need the clustering algorithm to calculate the AP position and ToF estimation.

3.3 Sybil attack Detection

Next, we attempt to tell whether the nodes on the same degree are located in the same position. Further, CSI we catch on the severing AP also contains the RSSI of different channel. RSSI was first found by William Jakes, who said Rayleigh fading process could be described by the sum of a series of complex sinusoid signals. Suppose node i receives radio signal from node 0, then the RSSI is $R_i = \frac{P_0 K}{\alpha \cdot d_i}$ where P_0 represents transmitter power, R_i is RSSI, K is constant of impulse response of Rayleigh channel model. d_i is Euclidean distance, and α is distance-power gradient. Hence, at the same transmit power, the RSSI is inversely proportional to the distance d. As we can prove in Fig 3, we compare different angle's total RSSI of the 30 subcarrier in the same radius of 2m with that in radius of 3m. We can see clearly the gap between the RSSI of different degrees. The average difference is 1.33dB. In addition to 15° gap is relatively small, the other degree of RSSI differences are more than 2.6dB. Therefore, we can distinguish between nodes in different angles at different locations effectively. For further explanation, we will detail the specific Sybil attack model in figure 1, to see whether our algorithm can effectively detect the attacker.

For the case of Fig 1(a), the client A falsifies the identity of the client B as node B' and sends a request message to steal the returned data sent from AP 1. Since AP 1 receives the CSI phase information sent by node B', AP 1 can locate the client for which is the actual sending data by combining our improved MUSIC algorithm. That is the client A who is located in the direction of 30°. We can be sure that the client B' should be identified as a Sybil node. Fig 1(b) and figure Fig 1(a) are similar, the only difference is client A, client D at the same angle, but different radius. Thus, while our MUSIC algorithm can determine that D' and D have the same angle, they have different RSSI values. node D' can also be determined as a Sybil node.

In the attack model of Fig 1(c), when A falsifies a large number of Sybil nodes, AP 1 does not care about the contents of the packet sent by the virtual node, and only determines the angle of the client

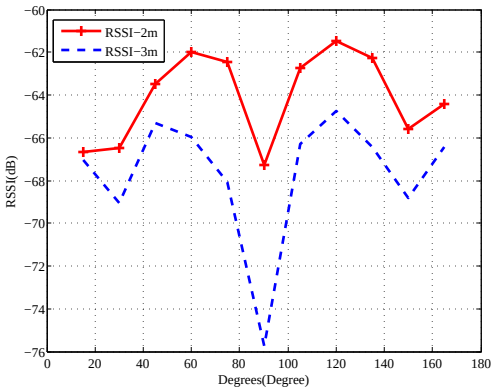


Figure 3: RSSI of 2m and 3m

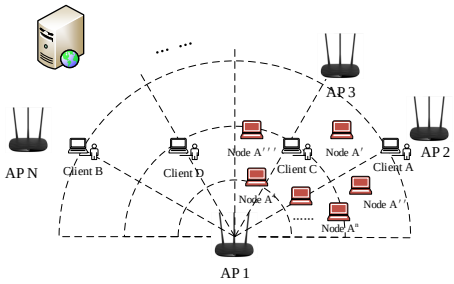


Figure 4: Sybil attack Detect System Model

based on the CSI by our algorithm. If the number of nodes does not affect the client B,client C,client D access to the network, we can determine that there are only four angles in the network sending the data and are at [30°, 150°, 60°, 120°]. When client B, client C, client D suffers from DoS attacks due to a large number of nodes, we can only detect that there is only one node in the network where the data is sent and is at 30°, so we can determine that the remaining nodes are Sybil nodes which forged by client A.

According to our Sybil attack models, we design a system to combine AoA and RSSI from one center witnessing AP to tell whether the node is a Sybil node. As shown in Fig 4, when all clients access to the wireless network, center witnessing AP equipped with off-the-shelf Intel 5300 WiFi NICs or other WiFi NICs is responsible for the data collection from all clients and send sample of packets to the server. After this, the center server calculates all CSI and RSSI measurement. That is to say, it run self-adaptive MUSIC algorithm to create a table about AoA and RSSI for each client. Once the Sybil client accesses to the wireless network, the system compares the AoA and RSSI values of the Sybil client and the actual client in the table. When they are below the threshold, they are sent from the same node, which claim to come from different clients. If so, send alarm massage to AP to stop receiving data from the Sybil client. However, in actual wireless networks, there may deploy multiple AP in order to improve communication quality and increase the available user ceiling. When multiple APs accept CSI at the same time, they can improve the accuracy of wireless network Sybil attack detection by applying the self-adopted MUSIC algorithm. Multiple APs can also detect one of center APs are attacked by Sybil node.

Based on our system, we propose two kinds of detection algorithms for AP and client respectively on the server. The algorithm steps are as follows:

- Sybil attack detection algorithm of Clients:

step 1 The client v_i send data to the $V_i \rightarrow AP_i : \{ID(V_i), CSI(RSSI, PHASE)(V_i)\}$.

step 2 AP_i sends data to the server to calculate each nodes' AOA, and updates the tables. $AP_i \rightarrow Server : \{ID(V_i), CSI(RSSI, PHASE)(Vi)\}_{AP_i}$. Update Client tables, $\{ID(V_i), RSSI(V_i), AOA(V_i)\}$.

step 3 Determine whether v_i is a Sybil client.

$$Server : \{AOA(V_i) - aoa(V_i) > \text{Threshold}\}_{AP_i}$$
$$Server : \{RSSI(V_i) - rssi(V_i) > \text{Threshold}\}_{AP_i}$$
$$Server \rightarrow O_i : \{ID(V_i), Alarm(V_i)\}$$

step 4 AP_i sends Alarm messages to neighbor nodes and sends Alarm messages to other common clients.

$$AP_i \rightarrow AP_{i+1} : \{ID(V_i), Alarm(V_i), RSSI(V_i), AOA(V_i)\}$$
$$AP_i \rightarrow V_{i+1} : \{ID(V_i), Alarm(V_i), RSSI(V_i), AOA(V_i)\}$$

- Sybil attack detection algorithm of APs:It is same as former 4 steps,only change v_i to AP_i send data to AP_{i+1}.AP_{i+1} confirm whether AP_i is Sybil node. If so, AP_{i+1} sends Alarm messages to neighbor AP nodes.

step 5 $O_{i+1} \rightarrow O_{i+2} : \{ID(O_i), Warnning(O_i)AOA(O_i), RSSI(O_i)\}$.

step 6 O_{i+2} confirm to the server. If O_{i+2} treat O_i is a Sybil node, then inform the clients and other APs.

$$O_{i+1} \rightarrow O_{i+2} : \{ID(O_i), \text{Alarm}(O_i)AOA(O_i), RSSI(O_i)\}$$
$$O_{i+1} \rightarrow V_i : \{ID(O_i), \text{Alarm}(O_i)AOA(O_i), RSSI(O_i)\}$$

4 Experimental Evaluation

The experiment is divided into three parts. Firstly, our experimental environment is 6.35 m×8.5 m meeting room. Receivers and transmitters is miniPC equipped with Intel 5300 WiFi NICs. They all contain 3 antennas. We use 1 antenna at transmitter and 3 antennas at receiver. We only employed one receiver. We installed CSI tools[8] on these miniPCs, which can receive CSI of 30 subcarriers. In order to eliminate the interference in the environment, while further reducing the impact of phase offsets, we use the monitor mode and 5.32GHz band. Our system can not only work in the 5Ghz band, also work in the 2.4Ghz band. We only need the center AP equipment with WiFi NICs which can receive CSI signals. Regardless of sending any type of WiFi device and any type of packet, we can receive the CSI when it accesses to the wireless network and sends it to the transmitter. Secondly, we use smart phone HTC M8 which support 802.11a/c protocol as transmitter. We use one mini pc with Intel 5300 WiFi NICs as a receiver. We do experiments in the office, filled with desks and computers. The size is about 64 m^2. Then we put the phone and the antennas at the same level of height and adjust the distance between the receiving antennas to 6cm to prevent interference between the receiving antennas. Finally, we conducted multiple sets of experiments to detect different Sybil attack models and calculated their detection efficiency in the first part of the experimental environment. At the same time, in order to verify our detection algorithm for efficiency of Sybil node, we use Matlab simulation to increase the number of nodes to view efficiency of Sybil attack detection, while comparing RSSI detection method.

4.1 Arrive of Angle Estimation Accuracy of Access Point

In this part of the experiment, we let the transmitter at 11 degrees of transmitter by the step of 15° range from 15° to 165°. The distance between the transmitter and the receiver is 3m. After collecting the CSI data, we changed the distance to 2m and carried out the same experiment. Through our improved

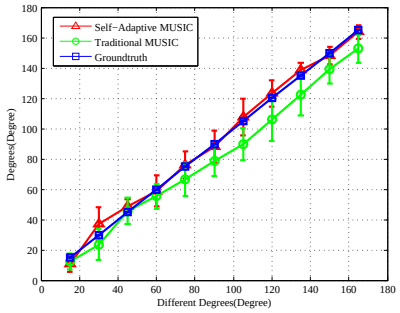


Figure 5: Error bar of traditional MUSIC and self-adaptive MUSIC algorithm

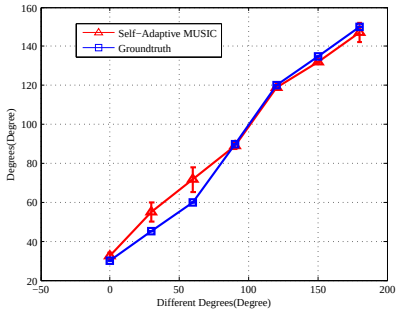


Figure 6: Mobile angle error bar of self-adaptive MUSIC algorithm

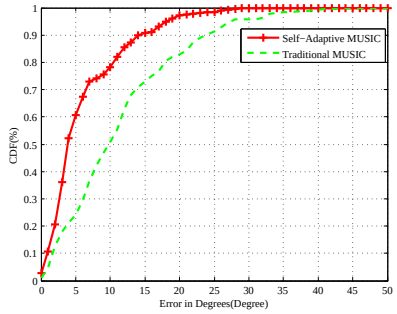


Figure 7: AoA Estimation error

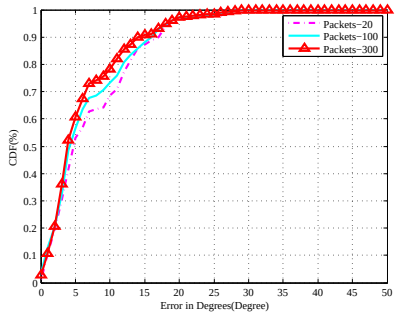


Figure 8: AoA Estimation error with different number of packets

MUSIC algorithm, we can determine the inherent phase offsets $\langle 8°, 20° \rangle$ of antenna 2,3. Fig 5 is the error bar of the traditional MUSIC algorithm and our algorithm in different angles measured data, which randomly selected 300 packets. We can see that the average of our algorithm is almost the same as the true angle with the average error is 6.3°, but the average error of the traditional operator is 11.9°.

Fig 7 shows the AOA's CDFs estimate error with all APs, and we can see that 80% of the AP's detection error in a large number of packets does not exceed 10°, whereas the traditional algorithm only 50%. We simulate the number of different packages as shown in Fig 8, the experiment shows that our algorithm can effectively improve the positioning accuracy with the CSI packets increase. At the same time Fig 3 ,which is the average RSSI of 8000 packets ,shows that difference between 2m and 3m is 1.3dB.

4.2 Arrive of Angle Estimation Accuracy of Mobile

Due to the many obstacles in the environment, we are making sure that the external environment has not changed. On the semicircle with a distance of 1 m, we changed the position of the mobile phone [30°,45°,60°,90°,120°,135°,150°] respectively as the measurement angle, respectively. Fig 6 shows that we can also effectively detect the angle of the phone in the wireless network with the accuracy of 7.2°. The average RSSI difference in different locations is 1.6dB.

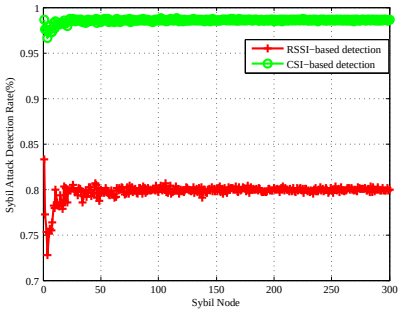


Figure 9: Sybil attacks' Sybil node Detection Rate

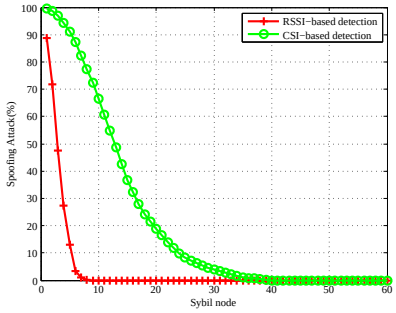


Figure 10: Sproofing Attacks' Sybil node Detection Rate

4.3 Detection of Sybil Nodes

This part is based on the experimental environment of the first part, where 20 sets of different nodes are randomly arranged at positions 2, 3, 4, 5m and the angle of [30°,60°,90°,120°] for Sybil attack detection. According to our Sybil attack model for multiple experiments, we set 6.3° degrees as AOA threshold, 1.3db as RSSI threshold. Because we only use one AP as a detection node, RSSI based Sybil attack detection efficiency is 76.5%.Whereas, our detection system based on CSI has the detection rate of 100%. We assume that we can detect the nodes in the error range AOA and RSSI based on the above ideal environment.

In order to further explore, we use Matlab randomly generate a large number of nodes with AOA and RSSI. We set 500 WiFi clients randomly, and simulate 200 times Sybil attacks. We can detect whether the Sybil attacks happen or not accurately as the Sybil nodes increasing. Figure 9 shows the detection rate of the virtual Sybil nodes generated by Sybil clients. We can see that our algorithm can achieve an average detection efficiency of 98.5% for the Sybil nodes, and the RSSI detection method can only achieve 79.8%. With the number of spoofing clients' Sybil nodes increasing, Fig 10 shows the probability that each node is detected by a spoofing attack. When the spoofing nodes are four, our detection efficiency is 94.2%. With the increase in spoofing at the same time, our accuracy is gradually declining. When a client virtual out of 40 Sybil nodes, we can't accurately distinguish every nodes generated by clients of spoofing attacks. But we can still know that some nodes are being attacked. Because as long as there is a node has not been accurately detected, we think that the entire test is a failure.

5 Conclusion

Different from the traditional RSSI detection algorithm, we first improve the accuracy of AOA by self-adaptive MUSIC algorithm based on modified CSI. Then we provide Sybil attack detection algorithm of APs and clients, which combine with AOA and RSSI. The algorithm can maximize the effective detection of Sybil nodes with detection rate of 98.5%. However, our algorithm is susceptible to environmental change, thus affecting the accuracy of AOA. Although we can detect a large number of Sybil nodes that move simultaneously, it can be mistaken for one single Sybil node when the node moves without informing the server. This will be done in future works.

Acknowledgments

Our work was supported by the Foundation of the Educational Commission of Tianjin, China(Grant No.2013080), the General Project of Tianjin Municipal Science and Technology Commission under Grant(No.15JCYBJC15600), the Major Project of Tianjin Municipal Science and Technology Commission under Grant(No.15ZXDSGX00030), and NSFC:The United Foundation of General Technology and Fundamental Research (No.U1536122). The authors would like to give thanks to all the pioneers in this field, and also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer networks*, 50(13):2127–2159, September 2006.

[2] D. Balfanz, D. K. Smetters, P. Stewart, and H. C. Wong. Talking to strangers: Authentication in ad-hoc wireless networks. In *Proc. of the Network and Distributed System Security Symposium (NDSS'02), San Diego, CA, USA*, February 2002.

[3] Y. Chen, W. Trappe, and R. P. Martin. Detecting and localizing wireless spoofing attacks. In *Proc. of the 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'07), San Diego, California, USA*, pages 193–202. IEEE, June 2007.

[4] M. Demirbas and Y. Song. An rssi-based scheme for sybil attack detection in wireless sensor networks. In *Proc. of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks (WoWMoM'06), Buffalo-Niagara Falls, New York, USA*, pages 564–570. IEEE, June 2006.

[5] J. R. Douceur. The sybil attack. In *Proc. of the 2002 International Workshop on Peer-to-Peer Systems (IPTPS'02), Cambridge, Massachusetts, USA*, volume 2429 of *Lecture Notes in Computer Science*, pages 251–260. Springer-Verlag, March 2002.

[6] D. B. Faria and D. R. Cheriton. Detecting identity-based attacks in wireless networks using signalprints. In *Proc. of the 5th ACM workshop on Wireless security (WISe'06), Los Angeles, California, USA*, pages 43–52. ACM, September 2006.

[7] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson. Phaser: Enabling phased array signal processing on commodity wifi access points. In *Proc. of the 20th annual international conference on Mobile computing and networking (MobiCom'14), Maui, Hawaii, USA*, pages 153–164. ACM, September 2014.

[8] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, 41(1):53–53, January 2011.

[9] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. Spotfi: Decimeter level localization using wifi. In *Proc. of the 2015 ACM Conference on Special Internet Group on Data Communication (SIGCOMM'15), London, UK*, pages 269–282. ACM, August 2015.

[10] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei. Dynamic-music: accurate device-free indoor localization. In *Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16), Heidelberg, Germany*, pages 196–207. ACM, September 2016.

[11] Z. Li, Y. Xie, M. Li, and K. Jamieson. Recitation: Rehearsing wireless packet reception in software. In *Proc. of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom'15), Paris, France*, pages 291–303. ACM, September 2015.

[12] C. Mitchell. Security analysis and improvements for ieee 802.11 i. In *Proc. of the 12th Annual Network and Distributed System Security Symposium (NDSS'05), San Diego, CA, USA*, pages 90–110, February 2005.

[13] J. Mitola and G. Q. Maguire. Cognitive radio: making software radios more personal. *IEEE personal communications*, 6(4):13–18, August 1999.

[14] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst. Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Communications Surveys & Tutorials*, 16(3):1550–1573, February 2014.

[15] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280, March 1986.

[16] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka. You are facing the mona lisa: spot localization using phy layer information. In *Proc. of the 10th international conference on Mobile systems, applications, and services (MobiSys'12), Low Wood Bay, Lake District, UK*, pages 183–196. ACM, June 2012.

[17] T.-J. Shan, M. Wax, and T. Kailath. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):806–811, 1985.

[18] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt. Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels. *IEEE Transactions on Signal Processing*, 52(2):461–471, January 2004.

[19] S. Tree. Wireless sensor networks. *Self*, 1(R2):C0, 2014.

[20] J. Wang, G. Yang, Y. Sun, and S. Chen. Sybil attack detection based on rssi for wireless sensor network. In *Proc. of the 2007 International Conference on Wireless Communications, Networking and Mobile Computing (WiCom'07), Shanghai, China*, pages 2684–2687. IEEE, September 2007.

[21] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni. Fila: Fine-grained indoor localization. In *Proc. of the 31st Annual International Conference on Computer Communications (INFOCOM'12). Orlando, Florida, USA*, pages 2210–2218. IEEE, March 2012.

[22] J. Xiao, K. Wu, Y. Yi, and L. M. Ni. Fifs: Fine-grained indoor fingerprinting system. In *Proc. of the 21st International Conference on Computer Communications and Networks (ICCCN'12), Munich, Germany*, pages 1–7. IEEE, July 2012.

[23] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe. Channel-based detection of sybil attacks in wireless networks. *IEEE Transactions on information forensics and security*, 4(3):492–503, September 2009.

[24] L. Xiao, A. Reznik, W. Trappe, C. Ye, Y. Shah, L. Greenstein, and N. Mandayam. Phy-authentication protocol for spoofing detection in wireless networks. In *Proc. of the 2010 IEEE Global Telecommunications Conference (GLOBECOM'10), Miami, Florida, USA*, pages 1–6. IEEE, December 2010.

[25] J. Xiong and K. Jamieson. Arraytrack: A fine-grained indoor location system. In *Proc. of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI'13), Lombard, Illinois, USA*, pages 71–84. Usenix, 2013.

[26] Z. Yang, Z. Zhou, and Y. Liu. From rssi to csi: Indoor localization via channel response. *ACM Computing Surveys (CSUR)*, 46(2):25, November 2013.

[27] H. Zhang, P. Cheng, L. Shi, and J. Chen. Optimal dos attack scheduling in wireless networked control system. *IEEE Transactions on Control Systems Technology*, 24(3):843–852, August 2016.

[28] W. Zhang, A. Hoorfar, and L. Li. Through-the-wall target localization with time reversal music method. *Progress In Electromagnetics Research*, 106:75–89, 2010.

[29] Y. Zhang, X. Chen, H. Wang, and L. Chen. Accurate indoor localization with channel state information. In *Proc. of the 3rd International Workshop on Energy Harvesting & Energy Neutral Sensing Systems (ENSsys'15), Seoul, South Korea*, pages 35–36. ACM, November 2015.

[30] Y. Zhang, W. Liu, W. Lou, and Y. Fang. Location-based compromise-tolerant security mechanisms for wireless sensor networks. *IEEE Journal on selected areas in communications*, 24(2):247–260, 2006.

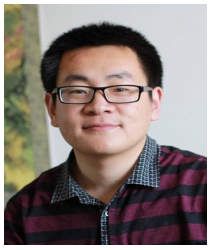
[31] N. Zhao, F. R. Yu, M. Li, Q. Yan, and V. C. Leung. Physical layer security issues in interference-alignment-based wireless networks. *IEEE Communications Magazine*, 54(8):162–168, August 2016.

[32] Z. Zhou, Z. Yang, C. Wu, W. Sun, and Y. Liu. Lifi: Line-of-sight identification with wifi. In *Proc. of the 33rd Annual International Conference on Computer Communications (INFOCOM'14), Toronto, Canada*, pages 2688–2696. IEEE, April 2014.

Author Biography

Chundong Wang received the BSc degrees in computer science from Tianjin Normal University, China, in 1991,He received the MSc and PhD degrees in computer science from Nankai University, China, in 2002 and 2007,respectively. Currently, he works at Tianjin University of Technology as an Professor. His current research interests include network information security, pervasive computing, mobile computing and intelligent information processing.

Likun Zhu received his B.S. degree in communication engineering from North China University of Science and Technology in 2016, and he is currently working towards the MSc degree in Information and Communication Engineering at Tianjin University of Technology, China. His main research is directed to the architecture of wireless security, indoor location and wireless perception.



Liangyi Gong received his B.E. degree in 2010 and Ph.D. degree in 2016 from the Department of Computer Science and Technology of Harbin Engineering University, Harbin, China. His main research interests include wireless networks and mobile computing. And development of the WiFi radar system, while participating in wireless Mesh video network research.

Zhentang Zhao received her B.S. degree from Liaocheng University in 2015. He is currently studying for a master's degree at Tianjin University of Technology, Tianjin, China. His research interests includes Wireless security and Security of Vehicle networking.

Lei Yang received her B.S. degree from Nanjing University of Postsand Telecommunications in June 2016. She is currently studying for a master's degree at Tianjin University of Technology, Tianjin, China. Her research interests includes privacy protection and Database watermark.

Zheli Liu received the BSc and MSc degrees in computer science from Jilin University, China, in 2002 and 2005, respectively. He received the PhD degree in computer application from Jilin University in 2009. After a postdoctoral fellowship in Nankai University, he joined the College of Computer and Control Engineering of Nankai University in 2011. Currently, he works at Nankai University as an Associate Professor. His current research interests include applied cryptography and data privacy protection.

Xiaochun Cheng Dr Xiaochun Cheng had his BEng on Computer Software in 1992 and his PhD on Artificial Intelligence in 1996. He has been a senior member of IEEE since 2004. He is the secretary for IEEE SMC UK&RI. He is a member of IEEE SMC: Technical Committee on Systems Safety and Security.He is also a committee member of European Systems Safety Society.