A Joint Detection Method for Identifying Pseudo Base Station Based on Abnormal Access Parameters

Jing Shao, Dali Zhu* & Hao Jin
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

Rong Qiao
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

ABSTRACT: The attack of pseudo base station is one of the security problems on mobile terminals, which affects normal communications and disguises as legit users for illegal purposes. To prevent pseudo base stations from interfering with users, this paper proposes a joint detection method that can judge pseudo base station. The abnormal access parameters in the process of terminal access to pseudo base station are divided into two categories, one is base station parameters, and the other is terminal parameters. Based on these feature parameters including LAC, CI, RSSI and mobile phone mode, the method concludes algorithm by Naive Bayesian classification. The experiment result shows that the recall rate increases to 70% and the false positive rate decreases to 30%, which means this method is better than traditional ways.

Keywords: pseudo base station; security; detection; abnormal access; feature parameters

1 INTRODUCTION

In recent years, security problems on mobile terminals have been highly concerned. The attack of Pseudo base station is one of those serious issues, which affects normal communications and can disguise as legit users for illegal purposes. Therefore, finding a way to detect pseudo base station and prevent their spread is necessary.

This paper proposes a joint detection method that identifies pseudo base station based on multi-mode intelligent terminals. In the process of terminal initial access to pseudo base station, we design a detection algorithm to monitor the abnormality of RSSI, mode switch and terminal’s characters for most multi-mode intelligent terminals.

The rest of the paper is organized as follows. In section 2 we describe some of the related works. In section 3 we discuss the principle of the pseudo base station. In section 4 we propose the joint detection method. In section 5 we present the experiments and results. We conclude our work and plan the future work in section 6.

2 RELATED WORK

Currently, the pseudo base station detection methods are mainly concentrated in the following three ways: (1) Base station power detection: The signal intensity of normal station reception is from -110dbm to -60dbm, and the power of pseudo base station received by the terminal can reach -50dbm. Paper [1] proposes detection method, which locates the pseudo base station by power energy maps. In paper [2], the method monitors the change of RSSI directly. (2) Base station parameters detection: According to the base station cell identification code, location area update identification code and other base station characteristic parameters, paper [3] monitors the parameters when terminals access to the pseudo base station. (3) Short message detection: In paper [4], digital signature was introduced in short message service center, so message was encrypted by SMS. After receiving the message, terminals verify the legitimacy of the digital signature.
In the first detection method, the average RSSI of pseudo base station received by terminal is about -50dBm[5], while sometimes in the strong signal area, the average RSSI of normal base station can also reach -60dBm, so the threshold to identify pseudo base station is not easy to decide. And pseudo base station is always moving, so the time efficiency of the power energy maps is very short. The sample graph of RSSI in ten seconds is shown as Figure 1.

![Figure 1. The change of RSSI in 10 sec.](image)

As we can see from Figure 1, the variation range of RSSI can reach 30dBm, so we can’t identify pseudo base station accurately only by power change.

In the second detection method, even though the main purpose of the current pseudo base is to send spam messages [6], which does not deliberately camouflage recognition code of public network, actually characteristic parameters of base station such as cell identification code can be forged by pseudo base station. The pseudo base station can send system broadcast information and carry information that are fully consistent with the signal format and content of public base station, namely that most base station parameters can be forged [7]. Pseudo base stations which steal secret information are likely to use station parameters that are consistent with the public ones.

The third method needs to modify SMS protocol and terminal protocol, which means high implementation cost and low practicability [8]. Therefore, it is not easy to judge whether the base station is a pseudo one according to the change of one parameter. We should combine with various parameters and establish a model to detect the pseudo base station.

3 PRINCIPLE OF THE PSEUDO BASE STATION

3.1 Background of pseudo base station

The current mainstream of pseudo base station platform is based on foreign open source GSM protocol stack. As a result of the 3G, 4G communication protocol involves bidirectional authentication process, namely the base station and terminal will verify the legitimacy of each other, the 3G, 4G pseudo base station is difficult to be achieved. For multi-mode terminals, the attack ways of pseudo base station interfere in the signals of 3G, 4G around mainly by signal suppression, and then they switch the mode of multi-mode terminal to 2G to access to pseudo base stations.

When frequencies of 3G and 4G signals are interfered, multi-mode terminals will monitor frequencies of 2G base stations, and access to 2G base stations. At the same time, as the terminal still adopts one-way authentication to access to 2G base stations, namely base stations need to verify the legitimacy of the terminal, while the terminal does not need to verify the legitimacy of base stations. As long as the transmission power is increased by pseudo base station, terminals that meet the cell reselection conditions will choose pseudo base station’s frequency.

3.2 The characteristics of the pseudo base station

According to the principle of pseudo base station, the process of terminal accessing to pseudo base station is mainly involved in cell reselection conditions. In the process of cell selection, when cell’s reselection parameters indicating (PI) is 1, the channel standard adopted by cell’s reselection is formula C2. The calculation method is as follow:

\[
C2 = C1 + \text{CELL\_RESELECT\_OFFSET} - \text{TEMPORARY\_OFFSET} \times H(\text{PENALT\_TIME} - T)
\]

where \(X < 0, H(X) = 0\); when \(X \geq 0, H(X) = 1\).

Wherein, CELL\_RESELECT\_OFFSET: Man-made correction parameters of cell reselection which can give parameters a positive correction; T: The timer; TEMPORARY\_OFFSET: In PENALT\_TIME, give cell reselection parameters a negative correction.

\[
C1 = \text{RXLEV} - \text{RXLEV\_ACCESS\_MIN} - \text{MAX}((\text{MS\_TXPWR\_MAX\_CCH} - P), 0)
\]

Wherein, RXLEV: Received Signal Level of the terminal, the unit is dBm; RXLEV\_ACCESS\_MIN: Minimum Received Signal Level of terminal base station, the unit is dBm; MS\_TXPWR\_MAX\_CCH: Maximum transmission power level of the terminal access to base station, the unit is dBm.

By giving C2 an approximate derivation, we get a more general formula, as follow:

\[
C2 = \text{RXLEV} - \text{RXLEV\_ACCESS\_MIN} + \text{CELL\_RESELECT\_OFFSET}
\]

Therefore, in order to guarantee the pseudo base station’s C2 value is big enough, we not only increase base station transmission power to make the received signal level of terminal big enough, but also set the minimum received signal level of terminal accessing to base station small enough. When the value of
MS_TXPWR_MAX_CCH is -110dBm, its code is 0. This value of normal base station is not set to -110dBm under normal circumstances. In order to balance traffic, it is generally set at around -100dBm, which is not greater than -90dBm.

Without considering the quality of user communication, pseudo base station will set this code to zero on high probability. When cell’s reselection parameters indicating (PI) is 0, C2=C1, without introducing positive correction, this is unfavorable to trigger cell reselection. In general, the cell reselection parameters indicating (PI) of both normal and pseudo base stations are set to 1, thus C2 is provided a positive correction, the range of correction is from 0 to 126dBm, the default value is set to 0 for 900M cell. 1800M cells are set to 10 (20 dBm), and the correction value of pseudo base station may be set relatively large.

Terminals access to pseudo base station after decoding system messages of pseudo base station’s radio channels, and find pseudo base station’s LAC (Location Area Code) is different from LAC in their storage, and then terminals will send Location updating request to pseudo base station. In the process of Location updating, pseudo base station can send IMSI (International Mobile Subscriber Identity) identity request to terminals to achieve the IMSI numbers of terminals. Using the IMSI, pseudo base station can send a fraud message from SDCCH. After sending text messages, most of pseudo base stations will send Location updating reject to terminals to force them to disconnect with station. In order to get more information, some pseudo base stations will even forge the authentication of mobile communication network to cheat terminals.

Therefore, based on the parameters of base station, pseudo base station has following characteristics: LAC and CI (cell identity) will change a lot even seem abnormal, RXLEV_ACCESS_MIN may be set to 0 or extremely low, CELL_RESELECT_OFFSET may be set extremely high. What’s more, based on terminal parameters, the pseudo base station has following characteristics: 4G or 3G may switch to 2G mode and RSSI is varied. At the same time we will incorporate terminal location for auxiliary criterion, because when the terminal location changes greatly, the above parameters will change a lot, and when the terminal location changes little, the above parameters will not be very different under normal circumstances.

4 THE JOINT DETECTION METHOD

4.1 The abnormal access parameters

The pseudo base station detection method proposed in this paper is a joint detection method based on several abnormal access parameters. According to the differences between pseudo base station and public base station, these parameters can be divided into two classes, one contains base station identification parameters, including location updating area code (LAC), Community Identity (CI), and minimum received signal level of terminal base station and man-made correction parameters of cell reselection. In these parameters, LAC and CI are affected by the change of terminal location. When the terminal position changes little, the normal location updating is not easy to be triggered, so the weight of these parameters is high. When the terminal position changes quickly, the normal location updating is easy to happen because the terminal may be at the edge of the location area or cell. The other class contains terminal parameters, including the mobile mode switch and the received signal strength (RSSI) of terminal. Before terminals access to pseudo base station, they may be suppressed from the high level security network to the lower level security network, so there is a process of 4G or 3G mode switching to 2G mode. Because the pseudo base station’s average power received by the terminal is stronger than normal base station and it changes a lot when terminal access to pseudo base station, the variation of RSSI will also be one of the parameters. Terminal parameters are also affected by the change of terminal location. When the terminal position changes little, the mobile mode switch is not easy to happen, so the weight of these parameters is high.

As the terminal position changes will affect the weights of base station and terminal parameters to identify the pseudo base station, related ways are divided into two categories. One category is when the speed of terminal location change is small (less than 15 km/h, namely the normal movement speed by bike for adults). If the terminal is not in the cell boundary or location boundary, terminals mostly do not induce cell re-selection. This situation contains a large size of samples, which are easy to collect, so the paper will take the Naive Bayesian classifier to train data. On the other occasion, the range of terminal location change is big, so it is difficult to collect sample data. We preset the weight of the above characteristic parameters, and then modify it by existing samples.

4.2 The small changes of terminal location

Considering the small changes of terminal location, we use Naive Bayesian classification methods based on the above identical parameters.

Naive Bayesian classification is defined as follows:

Define: \( x = \{a_1, a_2, a_3, \ldots, a_n\} \) is a classified set, each \( a \) is a feature property of \( x \).

\[ C = \{y_1, y_2, y_3, \ldots, y_n\} \] is a set of category.

If each feature property is conditional independent, according to Bayes’ theorem, the formula is as follows:

\[ P(y_i | x) = \frac{P(x | y_i)P(y_i)}{P(x)} \quad (4) \]
The denominator is a constant, and each feature property is conditional independent, the above equation becomes:

\[ P(x | y_i) P(y_i) = P(a_1 | y_i) P(a_2 | y_i) ... P(a_n | y_i) P(y_i) \]  

(5)

Based on the above analysis, we can conclude process diagram of classification based on naive Bayesian, as shown in Figure 2.

![Figure 2](image)

Figure 2. The process diagram of classification based on Naive Bayesian.

Define: \( C=1 \) is a pseudo base station, while \( C=0 \) is not. First, we determine feature properties and their classifications, and then we select four conditional independence feature properties that can be captured by android application layer: \( L(LAC) \), \( I(CI) \), \( S(MODE\ SWITCH) \), and \( R(CHANGE\ OF\ RSSI) \).

We define: \( L\{l=0\text{ (not change)}, l=1\text{ (change)}\} \), \( I\{i=0\text{ (not change)}, i=1\text{ (change)}\} \), \( S\{s=0\text{ (not change)}, s=1\text{ (change)}\} \), \( R\{r \leq 10\text{ (range is less than 10dBm)}, 10<r\leq 30\text{ (range between 10 and 30dbm), } r>30\text{ (range is greater than 30dBm)}\} \).

Therefore, the process of pseudo base station detection based on Naive Bayes is shown in Figure 3, and we define \( D \) as the change of terminal location in 1 second.

![Figure 3](image)

Figure 3. The process of pseudo base station detection based on Naive Bayes.

We have \( m \) samples that have been identified as pseudo base station and \( n \) samples identified as normal base station.

First, we calculate the probability that training samples are normal base stations:

\[ P(C = 0) = \frac{n}{m + n} \]  

(6)

We also calculate the probability that training samples are pseudo base stations:

\[ P(C = 1) = \frac{m}{m + n} \]  

(7)

Then we calculate probability of each feature property in each category through samples, as shown in Table 1.

<table>
<thead>
<tr>
<th>( p )</th>
<th>( C=0 )</th>
<th>( C=1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L(LAC) )</td>
<td>( l=0 ) ( P(l=0</td>
<td>C=0) )</td>
</tr>
<tr>
<td></td>
<td>( l=1 ) ( P(l=1</td>
<td>C=0) )</td>
</tr>
<tr>
<td>( I(CI) )</td>
<td>( i=0 ) ( P(i=0</td>
<td>C=0) )</td>
</tr>
<tr>
<td></td>
<td>( i=1 ) ( P(i=1</td>
<td>C=0) )</td>
</tr>
<tr>
<td>( S(MODE) )</td>
<td>( s=0 ) ( P(s=0</td>
<td>C=0) )</td>
</tr>
<tr>
<td></td>
<td>( s=1 ) ( P(s=1</td>
<td>C=0) )</td>
</tr>
<tr>
<td>( R(RSSI) )</td>
<td>( r \leq 10 ) ( P(r \leq 10</td>
<td>C=0) )</td>
</tr>
<tr>
<td></td>
<td>( 10&lt;r \leq 30 ) ( P(10&lt;r \leq 30</td>
<td>C=0) )</td>
</tr>
<tr>
<td></td>
<td>( r&gt;30 ) ( P(r&gt;30</td>
<td>C=1) )</td>
</tr>
</tbody>
</table>

The formula for the classifier as follow:

\[ P_0 = P(C = 0)P(L=l | C=0)P(I=i | C=0)P(S=s | C=0)P(R=r | C=0) \]  

(8)

\[ P_1 = P(C = 1)P(L=l | C=1)P(I=i | C=1)P(S=s | C=1)P(R=r | C=1) \]  

(9)

If \( P_0 > P_1 \), it is defined as not pseudo base station; if \( P_0 < P_1 \), it is defined as pseudo base station.

4.3 The large changes of terminal location

Considering the large changes of terminal location, we use the default weights, and then correct weights according to the sample. In this situation, the terminal is likely to reselect another cell, so we don’t choose Cl as the feature parameter. We select \( L(LAC) \), \( S(MODE\ SWITCH) \), \( R(CHANGE\ OF\ RSSI) \) as feature parameters and define the weights of above parameters are the same.

The value of feature parameters is normalized: \( L(LAC) \) is a 2 byte hexadecimal BCD code and its range is from 0000 to FFFF. It can be defined as 65536 different location areas. According to the protocol of GSM 04.08, if the geography of LAC is adjacent, the number of LAC is consecutive, so we set the normal range to 10. We define \( L \) as follows:

\[ L = \frac{\text{value of feature parameter}}{65536} \times 10 \]
L{ \frac{l_2-l_1}{10}, l_1 \text{ is the value before the change; } l_2 \text{ is the value after the change} }

Mode switch is as same as the small change of location.

\[ s=0 \text{(not change), } s=1 \text{(change)} \]

According to the change regular pattern of terminal’s RSSI, we set the normal range of R is 50 dBm,

\[ R\{\frac{r_2-r_1}{50}, r_1 \text{ is the value before the change; } r_2 \text{ is the value after the change} \} \]

We can conclude the formula of general judgment as follow:

\[ a(w_l+w_s+w_r) \geq \frac{1}{2} \quad (10) \]

and \( \sum(w_l, w_s, w_r) = 1 \), \( a \) is the correction value.

We set \( w_l, w_s, w_r \) to 1/3, \( a \) equals to 1, and then we correct the value of \( a \) based on previous samples.

5 EXPERIMENT AND RESULTS

5.1 A evaluation index system for pseudo base station

In this section we will use experimental data to evaluate the classification method. Usually, we use regression tests to assess the accuracy of the method and those test samples that are out of the training data.

The result of identifying pseudo base station is shown in Table 2:

<table>
<thead>
<tr>
<th></th>
<th>The station is pseudo</th>
<th>The station is normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>The system judge station as pseudo station</td>
<td>A times</td>
<td>B times</td>
</tr>
<tr>
<td>The system judge station as normal station</td>
<td>C times</td>
<td>D times</td>
</tr>
</tbody>
</table>

The following definition can be used to evaluate the performance of pseudo base station detection system:

1) Pseudo base station Recall Rate: \( PR = \frac{A}{A+C} \)

It reflects the ability to detect pseudo base station. The higher of the parameter means the lower probability of missing judgment.

2) Pseudo base station Miss Rate: \( PM = \frac{C}{A+C} = 1-PR \)

It reflects the occasion that the detection system does not diagnose the pseudo base station.

3) Pseudo base station False Positive Rate: \( PFP = \frac{B}{B+D} \)

It reflects the occasion that system judges the normal base station as pseudo. The higher of parameters means the lower probability of false judgment.

5.2 The classifier determined by samples

This paper will use the first and third indicators for comparison and evaluation of the algorithm and the second and third indexes to modify \( a \) value.

In the condition of the small changes of terminal location, we use 564 samples which come from the lab of information security for training. The classifier is as follows:

\[ P(C=0)=\frac{223}{564}=0.395 \]

\[ P(C=1)=\frac{341}{564}=0.605 \]

We calculate the probability of each feature property in each category through previous pseudo base station samples. The experimental results are shown in Table 3:

<table>
<thead>
<tr>
<th></th>
<th>C=0</th>
<th>C=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>L(LAC)</td>
<td>I=0</td>
<td>I=1</td>
</tr>
<tr>
<td>l=0</td>
<td>0.780</td>
<td>0.211</td>
</tr>
<tr>
<td>l=1</td>
<td>0.220</td>
<td>0.789</td>
</tr>
<tr>
<td>I(CI)</td>
<td>I=0</td>
<td>I=1</td>
</tr>
<tr>
<td>I=0</td>
<td>0.565</td>
<td>0.387</td>
</tr>
<tr>
<td>I=1</td>
<td>0.435</td>
<td>0.613</td>
</tr>
<tr>
<td>S(MODE)</td>
<td>S=0</td>
<td>S=1</td>
</tr>
<tr>
<td>S=0</td>
<td>0.511</td>
<td>0.440</td>
</tr>
<tr>
<td>S=1</td>
<td>0.489</td>
<td>0.560</td>
</tr>
<tr>
<td>R(RSSI)</td>
<td>r≤10</td>
<td>10&lt;r≤30</td>
</tr>
<tr>
<td>r≤10</td>
<td>0.323</td>
<td>0.282</td>
</tr>
<tr>
<td>10&lt;r≤30</td>
<td>0.350</td>
<td>0.278</td>
</tr>
<tr>
<td>r&gt;30</td>
<td>0.327</td>
<td>0.440</td>
</tr>
</tbody>
</table>

The formula for the classifier as follow:

\[ P(\text{C=0})=0.395 \times P(L=l|C=0)P(I=i|C=0)P(S=s|C=0)P(R=r|C=0) \]

\[ P(\text{C=1})=0.605 \times P(L=l|C=1)P(I=i|C=1)P(S=s|C=1)P(R=r|C=1) \]

5.3 The comparison of PFP and PR

According to the above classification, we classify the 1260 sample data which include training data, and calculate the recall rate and false positive rate. Then we compare the rate with other detection systems which only consider one parameter. We choose two methods, one is based on the change of power, and the other is based on the parameter of base station. These methods are mentioned in section 2.

We can get the recall rate of three methods, as shown in Figure 4. We define BS PAR as the parameter of base station detection, POW as power detection and UNION as the method of this paper. The x coordinate is sample numbers and y coordinate is the recall rate.

As we can see from Figure 4, when the sample size is small, the difference of recall rate is not obvious. As the sample size increases gradually, the joint detection method is better than the base station feature parameter detection and it is also better than the power detec-
The joint detection method considers changes of base station parameters and the terminal parameters at the same time. So the miss rate should be smaller, which is consistent with the theory.

The false positive rate is shown in Figure 5. We define BS PAR as the parameter of base station detection, POW as power detection and UNION as the method of this paper. The x coordinate is sample numbers and y coordinate is the false positive rate.

As we can see, the false positive rate is significantly higher by the power detection. This also verifies the previous: under normal circumstances, the RSSI has large range of change. The false positive rate of Feature parameter and joint detection is basically consistent in the early time, but as the sample size increases, the false positive rate of the joint detection method is slightly lower than the base station parameter detection. This is because the current mainstream pseudo base attacks are only sending rubbish short message and they do not deliberately set the parameters as the same as public network. The scheme of this paper combines several conditions, thus the false positive rate is reduced.

Comprehensively considering the above two indicators, we find that the joint detection method in this paper has a relatively good result on the detection of pseudo base station in the case of small change of terminal position, which has a relatively lower false positive rate and a higher recall rate with a simpler algorithm.

When the location of terminal position changes widely, we fix the parameter $a$ through the 346 existing samples. With the change of parameter $a$, the miss rate curve and the false positive rate curve are shown as Figure 6, in which solid line is for miss rate, dotted line is for false positive rate:

It can be seen from the figure that when terminal position changes relatively widely, with the increase of $1/(2a)$, the false positive rate gradually increases and miss rate decreases. When we modify value $a$ in the intersection of the two curves, the false positive rate and false negative rate are relatively ideal.

Above all, with considering two different occasions of the change range of terminal station, the recall rate of detecting pseudo station is above 70% and the false positive rate is below 30%. The detection method is better than the existing method. When the change of terminal location is smaller, the method is more effective.

6 CONCLUDING REMARKS

In the light of the characteristic parameters and terminal parameters of the base station, this paper proposes a joint detection method of pseudo base station, which is based on the naive Bayes and parameters correlation. Experiment shows that our method achieves better recall rate and false positive rate than the existing method, which means our method is valid. In next stage, the algorithm should be optimized and put into practice. Because the sample size is relatively small, the result should be improved based on large sample size, especially when the location of terminal station changes widely.

ACKNOWLEDGEMENT

This paper is sponsored by National High Technology Research and Development Program of China (863 Program) (2012AA013104).
REFERENCE


