An Optimal Sensor Architecture for Wi-Fi Intrusion Detection

Zhiqi Tao† Baikunth Nath (Sr. Member IEEE) †† Andrew Lonie†

†Department of Information Systems,
††Department of Computer Science and Software Engineering,
The University of Melbourne,
Victoria 3010 Australia

Summary
This paper presents a novel approach to identifying illegitimate nodes in wireless networks. We demonstrate that, given a sufficiently dense sensor network, we are able to discriminate all network nodes based solely on signal strength datasets. Further, we demonstrate that within the dense sensor network, only a small subset of available sensors are required to effectively discriminate between nodes, and that this subset can be readily identified. Finally, we propose that duplicated network nodes may be identified by unsupervised, real-time analysis of signal strength datasets.

Key words:
Wi-Fi network security, IEEE 802.11 network, Network Monitoring, intrusion detection, clustering

1. Introduction

The development of wireless (Wi-Fi) network technologies has been characterized by a succession of security problems [1-3]. Although the upcoming security enhancement standard, IEEE 802.11i [4-7], is widely expected to overcome weaknesses of cryptographic mechanisms in early versions of Wi-Fi products [8], the nature of wireless infrastructure ensures that attacks based on impersonation (spoofing attacks) in Wi-Fi networks will remain a major threat to Wi-Fi security [9]. As Wi-Fi network infrastructure inherently lacks reliable positional knowledge of the origin of individual network packets – as compared to, for instance, a wired network in which the physical wires provide strong evidence of a packet's origin - intruders are potentially able to impersonate legitimate access points (APs) and clients, and such impersonated nodes are extremely difficult to detect.

We propose to use radio signal strength data from multiple sensors to differentiate Wi-Fi network nodes at different positions. We demonstrate that a densely distributed sensor network is capable of collecting a rich signal-strength data set for each monitored node, which can be exploited to eliminate "blind spots" and effectively discriminate duplicated nodes.

We also demonstrate that, for each monitored node, the 'discriminatory ability' of the signal strength dataset is heavily weighted towards a small number of the available sensors for each node; generally the important sensors are very close to one of the node positions, and there is a strong correlation between the discriminatory effectiveness of a sensor and the absolute strength of the signal received by that sensor, providing a simple mechanism for identification of critical sensors. We provide evidence that this subset of critical sensors is sufficient to discriminate between duplicate nodes in all of our test cases.

Lastly, we propose that, using this approach to identifying critical sensors, it may be possible to detect duplicated network nodes by unsupervised, real-time analysis of the signal strength dataset, suggesting a new approach to intrusion detection in wireless networks.

The remaining sections of this paper are organized as follows: Section 2 describes background of this research. Section 3 discusses our approach to discriminating Wi-Fi network nodes via signal strength data from a densely distributed sensor network. Section 4 describes the generation of a real-world dataset based on a dense sensor network. Section 5 presents a comprehensive analysis of this dataset, demonstrating the discriminatory ability of the sensor network and correlation between absolute signal strength and discriminatory ability of individual sensors. Section 6 presents our conclusions and future work.

2. Background

2.1 Security threats in Wi-Fi infrastructure

From the network architecture perspective, Wi-Fi networking is an alternative implementation of the MAC sub layer and physical layer as compared to fixed network

Manuscript received February 5, 2008
Manuscript revised February 20, 2008

technologies – in other words, both wireless and wired networks share the same high level protocols but use different low level (data-link layer and physical layer) protocols. Wi-Fi network technologies bring a number of new challenges to security and privacy, due to the nature of the medium used to transmit information [10]. The challenges can be categorized via the classic CIA triad (Confidentiality, Integrity and Availability) [11]:

Confidentiality

Confidentiality is at risk if individuals other than the intended recipients are able to read a communication (confidentiality is not restricted to a particular network layer). Wi-Fi nodes broadcast omnidirectionally, subject to physical restraints of the environment such as radio signal absorption. So any Wi-Fi device operating on the same frequency band and within range is able to capture the signal at the data-link layer. Given the signal itself cannot be made confidential, encryption is an effective mechanism to protect the confidentiality of a communication over a Wi-Fi connection, and indeed, one of the most widespread security measures for wireless networks is the use of encryption, providing both confidentiality and potentially authentication through the use of shared secrets between client node and infrastructure (or more complex authentication mechanisms such as RADIUS). Well administered, strong encryption provides a very strong defense against confidentiality attacks; unfortunately, encryption standards employed in Wi-Fi have a poor record, and weak cryptographic mechanisms have been responsible for a number of security incidents [12-16]. Newer standards such as IEEE 802.11i [17] are more promising, but encryption systems are hard to administer unless the environment is very well controlled, and are often faulty or absent in practice. WEP (Wired Equivalency Privacy) in particular is a weak form of wireless security with fundamental and well-publicized flaws, but is extremely widespread. Many free tools exist for analyzing collected Wi-Fi traffic, including tools for breaking encrypted streams across various wireless encryption implementations [18-20].

Assuming a Wi-Fi network is either unencrypted, or a tool/method for decrypting the network stream exists, an attacker may be able to monitor not only confidential information but potentially the authentication information associated with the network clients, leading to the greater risk of a impersonation attack – where a legitimate network device is impersonated at the packet/frame level, either to gain direct access to the network infrastructure, or to manipulate a legitimate client into connecting to a rogue access point to monitor that client’s activity, thus (in both cases) compromising both integrity and availability.

Integrity

Integrity is achieved when a communication is authentic and complete. Wi-Fi networks are especially susceptible to integrity attacks, because of the lack of evidence to match received data to a particular source. There is a level of basic packet-level integrity support in commonly used Wi-Fi standards: for instance, the WEP standard uses a very simple Cyclic Redundancy Check (CRC) to check if the received Wi-Fi network frame was altered, and in IEEE 802.11i and WPA 2, a Pair wise Transient Key method is included to replace CRC to ensure the Wi-Fi frame unimpaired [17]. However, the fundamental integrity issue in Wi-Fi is the difficulty of exclusively identifying the origin of Wi-Fi network frames. This is discussed in more detail below.

Availability

Availability is compromised when legitimate users are unable to access a computer network service in a reliable and timely manner. Wi-Fi networks are also especially susceptible to availability attacks, again because of the difficulty of authenticating the source of a data packet. For instance, in a Wi-Fi network, all of the nodes generally share the same transmission channel on one open medium, so in order to make sure that only one Wi-Fi device transmits at one time, IEEE 802.11 Standard builds in Carrier Sense Multiple Access Collision Avoidance (CSMA/CA) [21]: before a Wi-Fi network node transmits, it sends out a request to occupy channel for a certain amount of time and all other devices which received this request will not transmit during this period. However, this service assurance feature creates a number of availability-based security vulnerabilities, including such scenarios as a Carrier Sense attack [3], in which an attacker could continuously requests for large intervals of transmission time, so that other clients on the same channel would be able to transmit at a very low throughput or even fail to transmit. Another example of an availability vulnerability would be de-authentication/de-authorization attacks [3], in which, by impersonating either a party of an established connection, an attacker can send de-authentication frames or de-authorization frames to force the other party to exit the authenticated state or to exit the associated state.

In our opinion, the fundamental reason that Wi-Fi networks infrastructures are vulnerable to Integrity and Availability attacks is that, lacking the ability to control or identify the position of a node, Wi-Fi network nodes must rely on (at a minimum) the MAC address in the header of each Wi-Fi network frame to identify traffic associated with a particular Wi-Fi network node. Although a MAC address is designed to be a unique identifier of a network interface, a number of operating systems support arbitrary network packet generation, including the data-link layer header; hence a Linux user can, for instance, generate a correctly formed network packet, with arbitrary headers including a modified MAC address, using freely available libraries [22]. This obviously lowers the identification 'value' of a MAC address in a wireless, or indeed wired, network. However, wired network infrastructures have physical wire-based links between sender and receiver nodes, which provide strong (but not irrefutable) identification evidence of a packet sourced from that link, particularly if supported by a monitored node: port list within the infrastructure. So, although packets can be impersonated (spoofed) in a wired network, it is not an insurmountable security issue, as the physical link between nodes can be used as a second identifying factor of end-nodes: if a particular network packet comes from the 'wrong' physical link – generally identified by a lookup table matching MAC address of the network node interface to port in the network switch/router - then it must have been spoofed and can be discarded (and, potentially, an alarm raised). The reliance on 'softer' measures for node identification, such as MAC address and shared secrets, makes WLAN inherently more insecure in node identification, and given that the lack of physical link is the major benefit of a wireless network, this is unlikely to change.

Clearly, it would be beneficial to Wi-Fi security if more positional evidence was available for node identification. In particular, intrusions based on impersonation attacks, which are the basis for many integrity and availability vulnerabilities, would be much more reliably detected if duplicated nodes could be discriminated on position.

2.2 Related work

Both academic researchers and industry vendors have proposed numerous ways to detect impersonation attacks in wireless (and wired) networks. For instance, a popular approach to detecting illegitimate APs in commercial systems is to physically patrol the area with a handheld "network analyzer" and match the identified nodes and APs against the known authorized points [23], [24]. This approach is likely to be effective in sparse networks covering a small area, or when restricted to detecting simple AP masqueraders, but is unlikely to detect a sophisticated fake AP setup. A skillful intruder would know to turn off the response of continuous beacon frames to avoid detection [9].

A variety of traffic-based intrusion detection methods have been proposed. Wright [25], Dasgupta et al. [26] and Guo et al. [27] suggested the use of packet sequence number to detect MAC address impersonation. Wright's approach was based on analysis of the sequence number gap between temporally subsequent network packets; if the gap exceeds a threshold, it is evidence that the packets may not have originated from the same source. Dasgupta et al. presented an anomaly based approach based on Fuzzy Logic in which a system trained on artificially generated sequence number traces containing mock impersonation attacks was used to identify anomalous events based on sequence number variation. Guo et al also investigated the regular and abnormal patterns of sequence number changes but limited to the scenario that victim node and attacking node are connecting to same access point simultaneously. Nevertheless the fundamental limitation of sequence number based approach is that sequence number is a plain-text field in Wi-Fi network frame header and is subject to compromise.

Tao et al. [28] proposed an approach based on deployment of dedicated wireless network "snoopers" close to the access points in their WLAN. They physically moved a transmitting laptop through the environment hosting the network, while recording both the position information of the laptop, and the signal strength measurements on packets transmitted by the laptop as received by the 'snoopers'. This data was used to train a Bayesian network model which was capable of estimating the location of network nodes (and thus by extension, unauthorized nodes). This technique is similar to commercial WLAN localization system Ekahau Position Engine [29]. Such approaches have a drawback: the training data is a snapshot which may not be completely representative of the environment in which the derived model is working. Generally, for a trained system to be extended to a new environment, another training phase would have to be conducted. Even in the same environment, if there are any substantial physical changes in the environment, training data may not accurately reflect actual radio propagation afterwards.

Gill et al. [30], [31] proposed intrusion detection techniques based on meta-information from wireless networks, i.e. RSSI and RTT. By extending the Snort-Wireless RSSI and RTT plug-ins, they developed a correlation engine based on observations of these metrics from a single sensor and used these metrics to profile Wi-Fi network clients. If RSSI and RTT change beyond defined thresholds, it is indicative that the packets may not have originated from a single source. However, some potential issues with this approach may be:

(i) Generally, the system clock of a Wi-Fi interface card only supports the hardware timestamp with microsecond accuracy [32]. As a Wi-Fi signal transmits at ~3×10^8 meter/second, the level of accuracy of discriminating a position change may be limited to ~300 meters. So RTT is largely decided by chipset's processing capacity in Wi-Fi network interface card but not the transmission distance of a Wi-Fi network frame.

(ii) One sensor's RSSI observation does not provide reliable 'absolute' information on a transmitting WLAN node because RSSI is designed to be used as a relative metric and does not guarantee an accurate absolute signal strength measurement. In the IEEE 802.11 standard [21], RSSI has the following definition:

> *"14.2.3.2 RXVECTOR RSSI*
> *The received signal strength indicator (RSSI) is an optional parameter that has a value of 0 through RSSI Max. This parameter is a measure by the PHY sublayer of the energy observed at the antenna used to receive the current PPDU. RSSI shall be measured between the beginning of the start frame delimiter (SFD) and the end of the PLCP header error check (HEC). RSSI is intended to be used in a relative manner. Absolute accuracy of the RSSI reading is not specified."*

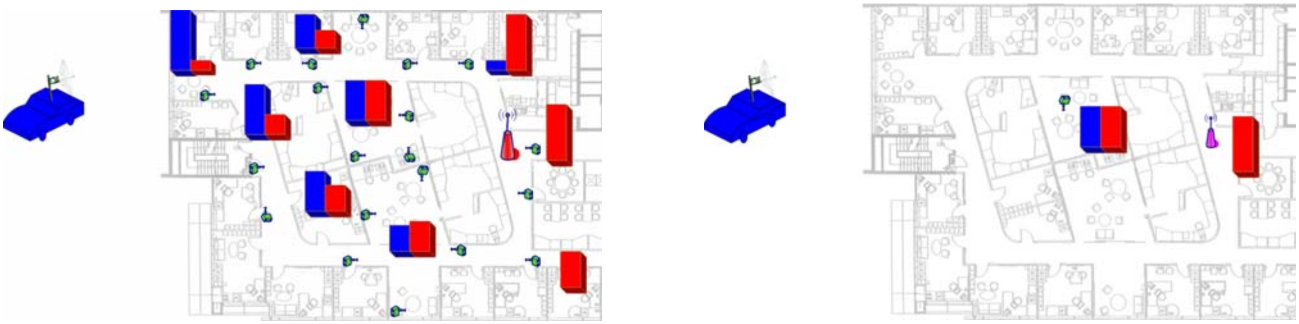
As reported in [33], different chipset vendors have implemented RSSI in different ways, so even WLAN network cards from same vendor might use different chipsets and therefore would report different RSSI value for the same signal. RSSI is also known to be dependent on the type of antenna equipped.

3. A novel approach to node discrimination

In order to exclusively discriminate Wi-Fi network nodes, we need to consider what 'unfakeable' identification information is available for each Wi-Fi network node.

We propose a method of using multiple and multi-angle measurements of a Wi-Fi network node's signal strength. As Wi-Fi network is a wireless technology based on radio wave, its signal propagation is subject to the length, medium and obstructions on its transit path. When a Wi-Fi network node emits network traffic, sensors at different locations receive and report different levels of signal strength. Although an impersonating device can disguise itself with fake identity at the packet level, the received signal strength (RSS) of its traffic is unlikely to be the same as that of the legitimate node because of different routes of radio propagation, particularly when RSS is measured by multiple sensors from different angles. As a result, the signal strength measurements from a sensor network may provide adequate information of discriminating Wi-Fi network nodes at physical different positions even if they appeared to have exact same identity.

We also suggest that, the denser and more evenly distributed such a sensor net is, the richer data set of RSSI value for each monitored node it can collect, potentially resulting in more effective discrimination. In contrast, many current approaches to using RSSI as an informational metric in Wi-Fi network monitoring tend to be based on RSS values derived from transmissions to or from the access points or few dedicated sensors within the network [34-37], which are normally located in geometrically central area (for the very good reason of providing maximal coverage with minimal investment) and may not provide sufficient discriminatory ability – as depicted in Figure 1, the smaller the number of RSS values, the greater chance for coincident value sets for nodes in different positions. Another concern is that centralized sensors may not be able to gather sufficient RSSI data



(a) Wi-Fi network nodes form a dense sensor network which has excellent coverage in the concerned area

(b) Access Points only reside in the central area but might not detect intrusions at the fringe

Fig. 1 Comparison between Access Point oriented approach and our sensor network architecture

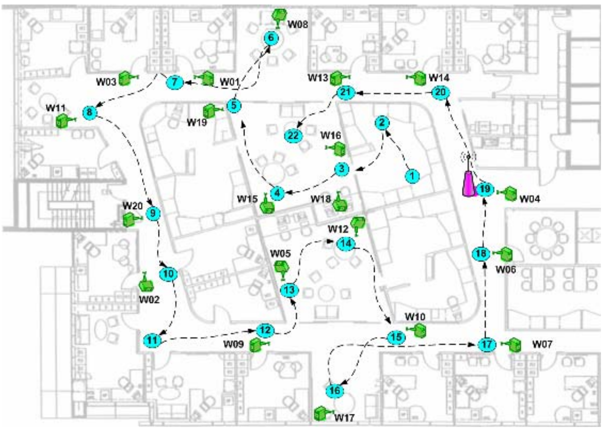


Fig. 2 Experimental Network

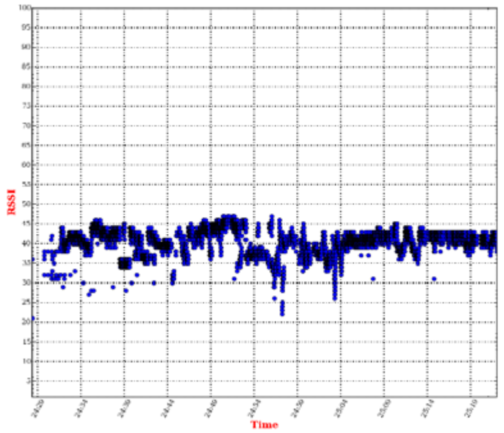


Fig. 3 Sensor w01's RSSI measurement for position 1

from nodes which are on the periphery of the monitored area. In fact as discussed in section 5, in this paper we demonstrate that when only few sensors are available, their RSSI measurements are often similar for a particular node in different locations.

4. Experimental setup and data generation

We constructed an experimental network to generate a data set that we considered would be sufficiently rich to explore the feasibility and effectiveness of our approach of discriminating between network nodes based on RSSI. Figure 2 illustrates our network setup at south side of Level three in ICT building in the University of Melbourne. It consists of twenty sensors and one moving monitored node. These sensors were Pentium III workstations equipped with a PCI Wi-Fi network interface card. The number of sensors is roughly equivalent to the regular number of client computers in this area, and the sensors were placed at locations that we considered reasonably simulated the distribution of client computers in the environment. On each of the sensors, the wireless interface was configured into passive monitoring mode. The moving monitored node was a laptop with a wireless interface, which was constantly transmitting to a central access point (via an automated script that generated web traffic). Sensors were configured to record all wireless traffic on the channel that the monitored node was transmitting, using an open source network monitoring program, WireShark. The monitored node was left at each position for 1 minute, which was sufficient time to collect a large amount (~15000) of network frames at each sensor. The node was then moved to the next position as depicted in Figure 2. When data at all positions were recorded, the Received Signal Strength Indicator (RSSI) value in the PRISMHEADER, along with other relevant information such as sensor identification index and node position index, was extracted and saved into a MySQL database. We also recorded the absolute position of each position, but have not used that data in this study.

During the experiment, the monitored node was moved through positions <1> to <22> as shown in Figure 2. In this way we collected a dataset which consists of RSSI measurements of the monitored node at 20 sensors over 22 positions. By choosing subsets of sensors out of the full 20 sensors, up to 1,048,575 different sensor configurations are possible (20!/20! + 20!/19!(1!) + 20!/18!(2!) … + 20!/1!(19!)), with 231 possible 'position pairs' ((22!/2!(20!)). Each position pair may be considered to represent one possible 'impersonation' attack, as for each pair there are two network nodes appearing to have the same frame-based identification information (MAC address). Because we have a range of distances and environmental obstructions represented by the set of 231 position pairs, we suggest that this data set covers a realistic and diverse wireless office environment.

5. Analysis

We conducted a comprehensive analysis of our dataset. In order to focus on examining the discriminatory ability of the sensor network to illegitimate nodes, we used a statistical method to epitomize the complex dataset and designed a comparative method. Not only does our analysis demonstrate that our approach can effectively cope with and various scenarios but also reveals correlation between individual sensors' received signal strength and their discriminatory ability, which suggests a simple mechanism to select sensible data for the real time intrusion detection discussed in section 6.

5.1 Statistical Analysis

The experiment was designed such that each sensor collected a large number of Wi-Fi network frames from the target node at each position. For instance, when the target node was at position <1>, sensor w01 collected 15079 frames, sensor w02 collected 14842 frames and sensor w03 collected 14744 frames. Fig 3 is an illustration of how sensor w01's RSSI measurements vary during this period. Darker regions in Fig 3 indicate that more frames were measured at these values. Intuitively we see that RSSI values for a stationary node as recorded by a single sensor demonstrate considerable variation - RSSI values at some points are between 20 and 25 while others are greater than 45. In order to reduce the dataset complexity for further analysis, some statistical derivative of the data collected is required. Table 1 summarizes general statistical descriptors of this dataset. It is clear that there is little difference between mean, median and mode of this dataset. This suggests that, although RSSI measurements for a particular node/position/sensor instance might vary, a statistical representation of the full RSSI dataset is valid. In all our further analyses, we chose mode as the representative function; subsequent references to RSSI actually refer to the mode derivative of the set of RSSI values for a particular node/position/sensor.

Table 1: Statistical Analysis of RSSI measurement of

Mean	Median	Mode	Standard Deviation
40.49009	41	42	3.104027

5.2 Level of Discrimination

In order to quantitatively compare how well sensor configurations can discriminate between 'identical' nodes at different positions using sets of RSSI values, we introduce a metric called "Level of Discrimination (LoD)". For any set of sensors, the set of (statistically simplified) RSSI measurements of a monitored node's transmitted signal (assuming the node is at a single stationary position) may be considered as an n-dimensional vector, in which n is equal to the number of sensors in the set. If the node is moved, a second vector would be expected to be created from the new set of RSSI measurements; the second vector will differ from the first as a function of the propagation of the radio signal from the node and it's interaction with the environment. So, multiple vectors will be defined for different positions of a monitored node, and the mathematical difference between the vectors represents the difference between the node 'environments' as represented in the recorded RSSI dataset.

For a specific set of sensors, LoD is defined as the Euclidean distance between two vectors representing the RSSI data sets recorded by that set of sensors for two different nodes (or one node at two different positions), and is calculated through formula (1).

$$LoD = \sqrt{\sum_{i=1}^{n}(s_i - r_i)^2} \qquad (1)$$

s and r represent a sensor's RSSI measurement in one sensor configuration regarding to position pair <s, r>. n represents the number of sensors in the configuration.

Vectors s and r may represent the same node in different positions, or two different nodes. In effect the LoD measures how well a set of sensors can discriminate between nodes in different positions based solely on RSSI. For the same position pair, a larger LoD indicates that a sensor configuration can better differentiate a position pair than the configuration with smaller LoD.

As an example, consider a subset of the experimental data described in the previous section. The RSSI dataset gathered on position pair <8, 17> by sensors {2, 7, 10} forms two RSSI vectors: |36, 35, 32| and |48, 86, 54|. The LoD (Euclidean distance between the two vectors) is calculated as.

$$\sqrt{(48-36)^2+(86-35)^2+(54-32)^2} = 56.82$$

Similarly, for the same position pair <8, 17>, we can consider a second set of sensors {4, 13, 14, 15}, with the resultant RSSI vectors |33, 42, 43, 42| and |50, 40, 46, 35|. The calculated LoD is then

$$\sqrt{(50-33)^2+(40-42)^2+(46-43)^2+(35-42)^2}$$
$$=18.73$$

Sensor configuration {2, 7, 10} has a much higher LoD than configuration {4, 13, 14, 15} for discrimination of position pair <8, 17>, even though more sensors are present in the second data set. Therefore, configuration {2, 7, 10} is a better choice of discrimination for this case.

We realize that this statistical treatment has limited practical value for impersonation detection, as LoD is derived from a simplified representation of RSSI datasets; this is only possible because the experimental data is pre-categorised according to node position and so data from different nodes can be statistically simplified independently. Neither the statistical simplification of RSSI measurements, nor the calculation of LoD would be

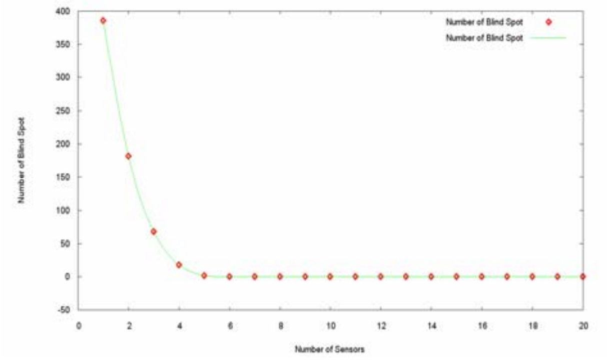


Fig. 4 Elimination of Blind Spots

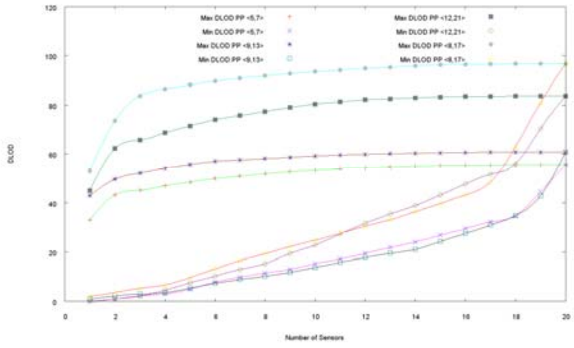


Fig. 5 Max and Min LoD in Position Pair <5,7>, Position Pair <9,13>, Position Pair <12, 21> and Position Pair <8,17>

possible in a real world impersonation attack because RSSI measurements would be received as a single spectrum of values for a particular MAC address, containing both genuine and impersonated packets, as there is no simple categorization feature available outside of MAC address (this is of course the main issue with impersonation attacks in wireless networks). However, our assumption is that the statistical analysis of pre-categorised data will help us to understand the information available in an RSSI dataset, its distribution between sensors, and how different sensor configurations affect discriminatory behavior. These analyses will inform further research in effective real time, naïve node discrimination; we discuss this further in Section 6.

5.3 Elimination of ‘blind spots’

For a number of instances in our experimental dataset, the LoD of a particular sensor configuration regarding a position pair is so small that it is equal to zero. This indicates that such a sensor configuration fails to discriminate between identical nodes at these two positions because sensors in those configurations measured insufficient differences in RSSI values when the monitored node was in different positions. For example, configuration {1, 3, 15, 19} reported same RSSI measurements at position 1 and position 13: |41, 44, 48, 43|. It is obvious that LoD for this case is zero. We term this kind of scenario a “blind spot”.

We calculated all possible LoD values from our experimental dataset, for all sensors sets (1,048,575) over all 231 positions pairs. Figure 4 plots the number of blind spots against the number of sensors in the sensor set. We notice that, for configurations consisting of few sensors, there are a considerable number of blind spots. For example, for sets of single sensors (of which there are 20) there are nearly 386 blind spots, which are 8.3% of total position pairs multiplying configurations (231 * 20 = 4620). However, as the number of sensors increase, the number of blind spots declines dramatically. Data from more sensors make RSSI measurements regarding to certain position pair more diversified, which results in eliminating blind spots. In our dataset, as long as we use sensor configurations consisting of more than 6 sensors, RSSI measurements are diversified enough to completely eliminate blind spots from our system, no matter which sensor configuration we choose – i.e. in any set of 6 sensors, there is enough information to demonstrate a LoD > 0.

5.4 Critical sensors

Although maximum LoD regarding all of position pairs can be achieved by taking data from all of sensors into consideration, we notice that the improvement of LoD declines after the first few ‘best’ sensors were taken. This indicates that, generally, a few of critical sensors supply much more important information in terms of discrimination than others. Fig 5 is an illustration. In Fig 5 the X-axis represents the number of sensors we used in a configuration. Y-axis represents LoD of this configuration

Table 2: RSSI measurements of Position Pair (8, 17)

	1	2	3	4	5	6	**7**	8	9	10	**11**	12	13	14	15	16	17	18	19	20
P8	64	36	48	33	43	38	**35**	45	50	32	**85**	39	42	43	42	47	27	38	52	32
P17	24	48	38	50	47	49	**86**	34	37	54	**32**	48	40	46	35	37	40	34	36	17

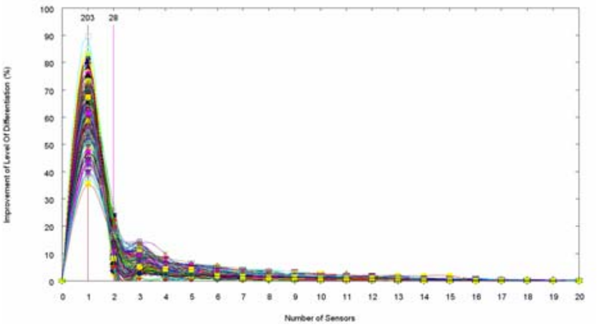


Fig.6 The number of Critical Sensors needed to achieve 50% of max LoD

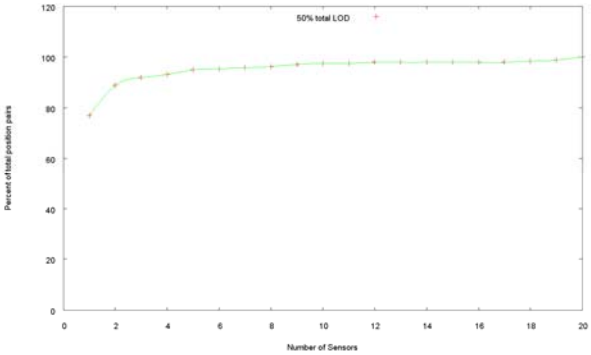


Fig. 7 The number of highest RSSI sensors vs. percentage of position pairs

regarding this position pairs. For simplification purpose we only plot the maximum LoD and minimum LoD for configurations consisting of certain number of sensors. LoD of all other configurations would then fall into the area between two edges. We plotted four position pairs, <5, 7>, <9, 13>, <12, 21> and <8, 17>. Position Pair <5, 7> has the shortest physical distance and Position Pair <8, 17> were selected for plotting as these two pairs have the shortest and longest absolute physical distance between node positions respectively, so might reasonably be expected to be 'hard' and 'easy' to discriminate (this is of course a simplification of the real case, as distance is only one of the factors that affects RSSI). Position Pair <9, 13> and Position Pair <12, 21> are two random examples from the rest of the 231 position pairs.

For all of four positions pairs, Maximum LODs improve dramatically in the first a few sensors and after that its improvements are trivial. On the other hand, minimum LOD continuously improves at similar rate as the number of sensors increase and suddenly accelerate as approaching the maximum number of sensors. This indicates that there exists a small set of critical sensors. We took position pair <8, 17> as an example and highlighted the three most critical sensors for discrimination of this position pair in Fig 6. Sensor 11 and 7 are two most critical sensors for this position pair as highlighted in Table 2. A configuration consisting of only these two sensors can achieve 76% of the maximum possible LoD (maximum LoD is assumed to be that which takes into account all of 20 sensors). In fact sensor 11 and sensor 7 are "mirror" sensors in that they each physical reside close to one end of position pair <8, 17> (refer Figure 2), and Table 2 demonstrate that their RSSI values are nearly opposite.

In order to find out if critical sensors exist in all of 231 position pairs, we analysed the contribution of each individual sensor to the discriminatory power of the complete set of sensors. In Figure 6, the X-axis represents the number of sensors in the configurations maximum LoD. The Y-axis represents the percentage contribution of each sensor, in decreasing order of contribution (calculated by ranking the sensors based on absolute difference in RSSI values between the two positions), to the total discriminatory power of the complete set of 20 sensors. So, the total of all plotted contributions for any position pair is 100% LoD. Intuitively, we see that for all position pairs, there is a clear bias to critical sensors that provide a large percentage of the overall discriminatory power. Vertical lines are marked on Figure 6 to represent the number of critical sensors needed to exceed 50% of maximum LoD, and the numbers above the lines represent how many position pairs such a number of critical sensors can achieve 50% of total LoD. It shows that, for 88% of 231 position pair, one critical sensor is enough to achieve 50% of maximum LoD; only 12% of them need two critical sensors to achieve 50% of maximum LoD.

5.5 Identifying critical sensors

We find a strong correlation between the discriminatory effectiveness of a sensor and the absolute strength of the signal received by that sensor, providing a simple method for naïvely identifying critical sensors. As an example, critical sensors 7 and 11 for position pair <8, 17> are sensors with the highest RSSI values at each end. Figure 7 demonstrates the generalisability of this approach across all 231 position pairs. This figure plots the number of highest RSSI sensors needed to include critical sensors contributing to 50% of maximum LoD against the percentage of position pairs that fall within this criteria. The figure demonstrates that, for over 90% of position pairs, the highest three RSSI sensors achieve over 50% of maximum LoD. Further, although we have not quantitatively analysed it, the sensor/node location spread from Figure 2 suggests a very strong correlation between physical distance between sensor and node, and discriminatory contribution of that sensor. This suggests further possibilities in identifying the location of intrusions.

Several questions arise from the results we have presented here, the most obvious being: what is a sufficient LoD for practical node discrimination? As presented, LoD is an artificial metric, as it is derived from RSSI sets that have been pre-categorised according to node position, reduced via the statistical mode function. In a real scenario, RSSI data is not pre-categorised, and the statistical derivative that our LoD metric is based on (mode), will not allow for discrimination of nodes based on RSSI alone, as it is an averaging metric across all RSSI data. We address these points in the next section.

6. Conclusions and future work

In this paper we investigated an important issue in Wi-Fi network security - node discrimination. We critically examined previous research in impersonation attacks in Wi-Fi networks, and suggested that the difficulty of discriminating Wi-Fi network nodes is a product of the inherent nature of current Wi-Fi network infrastructure. We proposed a distributed dense sensor network to address this issue. Our research demonstrates that such a sensor network can achieve a high level of discrimination for various scenarios without any pre-knowledge of operating environment, and because data from multiple sensors is used, the approach is effective in addressing problems such as blind spots which are generally encountered when only limited sensors participate in node monitoring. Our results also suggest that, among all of the participating sensors, only a small set of sensors play a critical role in discrimination. For the majority of position pairs it is sufficient to select the sensors reporting the highest RSSI in order to achieve half of the maximum discriminatory values. It is important to note that the set of critical sensors is different for each position pair; so although only a few sensors are required in each discrimination instance, it is necessary to have a dense sensor network from which to identify the critical sensors in each case. Additionally, it is necessary to have a well distributed sensor network, as critical sensors tend to be the sensors that are closest to one of the monitored node(s).

Our results have demonstrated the feasibility of using rich sensor data sets for discriminating duplicate nodes in a Wi-Fi network. The analysis methods presented are not, however, directly applicable to real world discrimination as the statistical treatments we have used depend on the data being precategorised by node position. To this end, we are currently engaged in developing a approach to analysis in which, under the assumption that there is sufficient discriminatory power available in a multi-sensor RSSI dataset from which critical sensors have been selected, we would identify the concurrent RSSI value ‘peaks’ which would indicate duplicate nodes in different positions. We are currently exploring the development of a naive clustering algorithm based on mode which will exploit the large separation of concurrent RSSI clusters from critical nodes.

We also plan to investigate how to discriminate traffic from two identical nodes in real time. Because security applications, and intrusion detection systems in particular, require both fast response times and low false positive and false negative rates, we will explore trade offs between discrimination accuracy, processing time, amount of data required, thresholds of detectable distance between sensors and number of sensors.

Finally, a potential criticism of the dense sensor network approach is that it is economically unfeasible to deploy such a network. However, we suggest that a densely distributed sensor network can be constructed with minimal extra hardware investment by recruiting existing wireless clients in the network as sensors. Most Wi-Fi network interface cards are capable of being configured to passively monitor Signal Strength (RSS) for any Wi-Fi network frames they receive, and as the nature of a wireless client base is to be well distributed, they can be recruited to create a densely distributed sensor network, and, in fact, our experimental dataset was generated from a client network that had been configured in such a way. There are of course other concerns with using non-fixed sensors for intrusion detection; we intend to address these and other issues in our subsequent reports.

References

[1] N. Borisov, I. Goldberg, and D. Wagner. Intercepting mobile communications: the insecurity of 802.11. Proceedings of the 7th annual international conference on Mobile computing and networking, pages 180–189, 2001.

[2] W. Arbaugh, N. Shankar, and Y. Wan. Your 80211 wireless network has no clothes. Wireless Communications, IEEE Personal Communications, 9(6):44–51, 2002.

[3] J. Bellardo and S. Savage. 802.11 denial-of-service attacks: Real vulnerabilities and practical solutions. In Proceedings of the USENIX Security Symposium, 2003.

[4] K. Baek, S. Smith, and D. Kotz. A survey of wpa and 802.11 i rsn authentication protocols. Technical report, Technical report, Dartmouth College, 2004. In preparation, 2004.

[5] J. Chen, M. Jiang, and Y. Liu. Wireless LAN security and IEEE 802.11 i. Wireless Communications, IEEE [see also IEEE Personal Communications], 12(1):27–36, 2005.

[6] J. Edney and W. Arbaugh. Real 802.11 Security: Wi-Fi Protected Access and 802.11 i. Addison-Wesley, 2004.

[7] C. He and J. Mitchell. Analysis of the 802.11 i 4-way handshake. Proceedings of the 2004 ACM workshop on Wireless security, pages 43–50, 2004.

[8] J. Edney and W. Arbaugh. Real 802.11 Security: Wi-Fi Protected Access and 802.11 i. Addison-Wesley, 2004.

[9] Z. Tao and A. Ruighaver. Detecting rogue access points that endanger the maginot line of wireless authentication. In Proceedings of the 3rd Australian Information Security Management Conference, pages 103–110, 2005.

[10] Z. Tao and A. Ruighaver. Wireless intrusion detection: Not as easy as traditional network intrusion detection. Proceedings of Tencon 2005: 2005 IEEE Region 10, 2005.

[11] Solomon, M.G. and Chapple, M., "Information Security Illuminated", pp 2-5, Sudbury, MA, Jones & Bartlett Publishers, c2005

[12] B. Brewin, D. Verton, and J. Disabatino. Wireless LANs: Trouble in the air. CNN Technology Online, 1 2002. (accessed on 20 Nov 2007).

[13] D. Sieberg. 'off-the-shelf' hack breaks wireless encryption. CNN Technology Online, 8 2001. (accessed on 20 Nov 2007).

[14] W. Slavin. Security overhaul for wireless networks. BBC Sci/Tech Online, news.bbc.co.uk/2/hi/science/nature/1723171.stm, 12 2001. (accessed on 20 Nov 2007).

[15] Wireless networks easy to hack. CNN Technology Online, 8 2001. (accessed on 20 Nov 2007).

[16] M. Ward. Welcome to the era of drive-by hacking. BBC Sci/Tech Online, 11 2001. (accessed on 20 Nov 2007).

[17] IEEE Standard 802.11i-2004, 2004.

[18] J. Ellch. jc-wepcrack. http://www.802.11mercenary.net/jc-wepcrack/, 2001. (accessed on 20 Dec 2007).

[19] WEPCrack - An 802.11 Key breaker, http://wepcrack.sourceforge.net/ (accessed on 20 Dec 2007).

[20] Aircrack-ng, http://www.aircrack-ng.org (accessed on 20 Dec 2007)

[21] IEEE Standard 802.11 1999, 1999.

[22] MadWifi, http://madwifi.org, (accessed on 20 Dec 2007)

[23] M. Milner. Netstumbler v4. 0. www.netstumbler.com, 2005. (accessed on 20 Nov 2007)

[24] AirMagnet. Laptop analyzer & handheld analyzer, 2007. (accessed on 22 Nov 2007).

[25] J. Wright. Detecting wireless LAN MAC address spoofing. White Paper, January, 2003. (accessed on 20 Nov 2007).

[26] D. Dasgupta, J. Gomez, F. Gonzalez, M. Kaniganti, K. Yallapu, and R. Yarramsetti. MMDS: Multilevel Monitoring and Detection System. Proceedings of the 15 the Annual Computer Security Incident Handling Conference, Ottawa, Canada, June, pages 22–27, 2003.

[27] F. Guo and T. Chiueh. Sequence number-based MAC address spoof detection. In Proceedings of the 8th International Symposium on Recent Advances in Intrusion Detection (RAID 2005), pages 309–329, 2005

[28] P. Tao, A. Rudys, A. Ladd, and D. Wallach. Wireless LAN location-sensing for security applications. Proceedings of the 2003 ACM workshop on Wireless security, pages 11–20, 2003.

[29] M. Johnson, J. King, R. Shryock, and T. Kiviniemi, J.; Heinonen. Analysis of a signal strength based positioning system for commercial environments. Consumer Communications and Networking Conference, 2005. CCNC., pages 533– 538, 2005.

[30] R. Gill, J. Smith, and A. Clark. Experiences in passively detecting session hijacking attacks in IEEE 802.11 networks. Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54, pages 221–230, 2006.

[31] R. Gill, J. Smith, and A. Clark. Specification-based intrusion detection in WLANs. In Proceedings Annual Computer Security Applications Conference, 2006.

[32] Velayos, H. and Karlsson, G., "Limitations in range estimation for wireless LAN", Proc. 1st Workshop on Positioning, Navigation and Communication (WPNC'04), Hannover, Germany, 2004

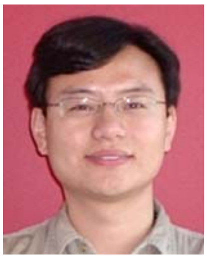
[33] J. Bardwell. Converting signal strength percentage to dbm values. Technical report, WildPackets, Nov 2002.

[34] AirDefense. Airdefense enterprise. http://www.airdefense.net. (accessed on 20 Nov 2007).

[35] AirMagnet. Airmagnet enterprises suite. http://www.airmagnet.com. (accessed on 20 Nov 2007).

[36] HighWall. Technologies. Highwall technologies mobilesecure. http://www.highwalltech.com. (accessed on 20 Nov 2007).

[37] WildPackets. Wildpackets airopeek. http://www.wildpackets.com (accessed on 20 Nov 2007).



Zhiqi Tao is pursuing his PhD degree at the Department of Information Systems in the University of Melbourne. He received the B.S. degree in Information and Telecommunication Engineering from Xi'an Jiaotong University in 2000. His research interests are Wireless network architecture and information security.

Dr Baikunth Nath received his MA at Punjab University, India and PhD at the University of Queensland, Australia. He served Monash University for over 25 years in various senior positions including the Director of Research at the Gippsland School of IT. In 2001, he joined the Department of Computer Science and Software Engineering as Associate Professor and the Director of Postgraduate Studies at the University of Melbourne, Australia. His research interests include Image Processing, Intrusion Detection, Scheduling, Optimization, Data Mining, Evolutionary Computing, Neural Networks, Financial Forecasting and Operations Research. He is a senior member of IEEE and is author of numerous research publications in various well-reputed international journals and conferences.

Dr Andrew Lonie lectures at the Department of Information Systems in the University of Melbourne. He received his PhD degree from Adelaide University in 1994. Prior to joining the university he was Security Architect at ANZ bank. His research interests include: information systems organisation & management, conceptual modeling, bio-informatics and information security.