IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

Surveillance of Wi-Fi Devices with Privacy Preservation and Evading Anonymous Usage

Gokulnath G

Assistant Professor, Computer Science and Engineering, SAINTGITS College Of Engineering, Kottayam, India

Abstract: Wired networks are increasingly being replaced by wireless networks due to its varied advantages. Open access networks are implemented in most of hotels, libraries and institutions throughout the world. Many security measures and protocols are adopted for securing the wireless personal area networks. The flaws in wireless networks and limitation of security protocols have increased the number of attacks as well. As a result a secure system has to be developed that can prevent attack to the wireless networks by identifying the location of the mobile device with only using the current resources.

Keywords: Wi-Fi system, security protocols, attacks to system , location identification, privacy

I. INTRODUCTION

Internet has been the major mode of communication in this era of computing. It is indefinite to count the applications of it in daily life. The traditional wired networks are replaced by wireless network which are fast and at the same time installed at low cost. This has led many service providers to give lesser cost or free internet to thrive in the competition. Hotels, Inns, Libraries and shopping malls are equipped with wireless networks for free of cost. The tremendous evolution in the wireless network especially WIFI defined by IEEE 802.11b networks attributed too much of low cost networks. Security standards such as WEP, WPA, WPA2 have been developed to neutralise the threats that can arise in wireless networks. Furthermore methods are developed to secure the personal area networks such as RF-shielding and using the IEEE 802.11x restricted access networks.

In spite the various security measures the attacks on wireless systems are increasing. When compared with the wired network the greatest advantage of mobility is often misused to launch attacks to the system. The basic attacks in the wired system such as man in the middle, Identity theft, DOS, etc. apply to wireless system as well. Unintentional threat such as accidental association can also occur in wireless personal area network when one accidentally connects to an access point of the neighbouring building. Intentional threat known as malicious association arises when one manually connects to the neighbouring access point with an intention of causing security threat.

The threats can be resolved by identifying the attacker at time of attack. Often the IP addresses and MAC addresses are used to identify the user. But in wireless networks the use of IP addresses that are free are altered at the time of attack. Also MAC addresses are no longer unchangeable as commands exist in LINUX and Windows operating systems to change the mobile MAC addresses. The solution to such a scenario is to identify the user by the location. In case of mobile devices the location is dynamic and often attackers use specialised techniques to prevent them from being traced. Software defined radios and directional antennas can alter the received signal strength information and change the time of arrival and angle of arrival of the signals to avoid tracking. Temporal and spatial cloaking algorithms are used by specialised attackers to forge the location data. This can be achieved by using a server designed for de-personalisation which transmits only such data that are insufficient to find the location of the attacker. Furthermore an attacker can also send depersonalised data rather than opting for a specialised server [3],[4].

Many position system exists that can trace the location of the users. But none can be used directly to Wi-Fi without external devices. The surveillance systems also need to address the issue of location privacy. [2] Studies that Location is regarded as private data by most of the users.

This paper proposes a system that can be used for positioning of mobile nodes in Wi-Fi networks. The system is built with the present network resources, this helps to easily adapt to system changes and is mobile in nature. Unlike the existing surveillance systems the proposed system is to provide support to the law enforcement agency to bring suspect in front of law. This overcomes the problem of location breaching by evading the misuse of the system in public hands. The location identified by the system can be used as a material for forensic study to reveal the suspect user activities.

Initial process is to design the system with a high gain antenna and low noise amplifier to collect as many as signals in wide area with less interference. The wireless traffic can be analysed to find the access points communicating with the mobile device. Second phase is the localisation process where localisation algorithms are used to calculate the exact position. The algorithms uses the available information about the access points to retrieve the location, if no such information is ready, the system uses war driving technique to collect training data which is used to locate the access point communicable to the mobile device.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

II. SYSTEM

This section briefly describes the idea of the surveillance system.

A. Basic Concepts

While dealing with the positioning of Wi-Fi devices the only information available is communication data between the access points and the mobile devices. So a specialised capturing system can be designed that can collect these data to provide useful information about location. The system can be designed to work in two phases. In the initial phase the low enforcement agencies collects the access point information which are its location and maximum transmission distance with the help of war driving mechanism. These data can also be retrieves from websites showing wireless geographic information and hence this step can be optional. In the second phase the law enforcement agency locates the mobile device based on the information collected. It identifies the set of access points communicating with the mobile device. Then the location of mobile device can be identified with respect to the information related to the communication APs. The location identified so can be proved to be correct by using association and disassociation mechanism with the access points. This evades the possibility of cloaking or transmission of de-personalised information. In order to effectively perform these steps it is required to efficiently capture the whole traffic and this can be done monitoring the channels (minimum of 11 channels for most WI-Fi networks). For this a high gain antenna is used which can collect the traffic at much wider area.

One of the major issues that need to be addressed is- How to collect probing traffic if the device is not sending any. For this the system proposes active and passive mechanism for collecting traffic. The passive mechanism does not interfere with the network protocol whereas the active mechanism alters the protocol in order for the system to transmit packets. The following sections explain these scenarios.

B. System Components

The components of the system vary as per the network. But in common the system is composed of four major parts.

1. Receiver system- It consists of the antenna to capture the wireless packets for wide spread area. This is achieved by increasing the signal strength of distant devices. A low noise amplifier is attached to it which can reduce the unwanted signals creeping in the system. Signal splitters and wireless cards are used to capture the signals to the respected wireless device.

2. Traffic capture system- The low enforcement can monitor the probing traffic in all the 11 channels. Then useful information revealing the identity of suspect such as Service Set Identifier, MAC address etc. can be collected and stored.

3. Localisation System- The data collected through the second step and the training phase is used to locate the suspect device. The information available are fed into the algorithm discussed later which gives the precise location.

4. Web display- The location so obtained from the third stage is shown as web display for ease of viewing by low enforcement agencies. This distinction between the ideal and suspect user is noted.

III. SURVEILLANCE SYSTEM

There are three major challenges while dealing with the design of the system. The first one is the coverage area. The important question while design is- The way to increase the coverage area by using the off the shelf equipments. The system is designed to use a high gain antenna that can boost the signal strength. But this offers high noise in the system, which can be reduced by using low noise amplifier. In order to correctly recognise a wireless signal by the Wireless network interface card, it should have the signal strength higher that the sensitivity of the card which is at receiver side the minimum signal strength received. The following theorem in radio theory supports the fact of using high gain antennas. The proof and explanation of the theorem can be obtained at on the ComputerSocietyDigitalLibraryathttp://doi.ieeecomputersociety.org/10.1109/TMC.2011.70

.Theorem 1. To receive a wireless signal

$$20 \log_{10} D < G_{rx} - NF_{lna} - SNR_{min} + C;$$

where D is the distance between receiver and transmitter (i.e., coverage radius in free space), Grx is the receiver antenna gain, NFlna is the noise factor of the low noise amplifier, SNRmin is the minimum signal noise ratio of the receiver to have acceptable demodulation accuracy for digitalized baseband circuitry, and C is as follows:

$$C = P_{tx} + G_{tx} - 20 \log (4\pi/\lambda) - 10 \log B - (-174);$$

where Ptx is the transmitter power, Gtx is the transmit antenna gain, _ is free space wavelength, -174 (dBm/Hz) is the value of the noise power density of the wireless NIC input impedance (normally 50 Ohm), and B is the receiver's bandwidth in Hz, which is normally defined by the baseband filter bandwidth. The theorem is used for the free space radio transmission but what we are dealing with will be disturbances caused by walls, building etc. So this can be used as model giving the result in worst case with low accuracy of mobile devise location. This decrease in accuracy can be alleviated by the algorithms for localization. Second challenge is for probing traffic collection is the requirement of monitoring a large number of channels. Both 802.11b (DSSS) and 802.11g (OFDM) wireless LANs have 11 channels, each of which has a frequency width of 22 MHz. The only three channels that do not interfere with each other concurrently are channels 1, 6, and 11. It is supposed that, to capture all 802.11b/g signals at 2.4 GHz, there require at least three wireless cards, NIC3, NIC6, and NIC9, to monitor Channels 3, 6, and 9, respectively, such that NIC3 picks up signals for Channels 1 to 5, NIC6 for Channels 4 to 8, and NIC9 for Channels 7 to 11.But this system of based on supposition is not practical. There are chances that the energy of the

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

signal can be leaked to neighbouring channel, but it is difficult for a card to correctly identify the signal on a neighbouring channel because the signal received by the neighbouring channel could possibly be distorted and the card cannot identify it. A possible solution to this is to use as many number of cards as in number of channels. But this new solution not only increases the cost of the system but also make the surveillance system immobile. Moreover, to support 802.11a, an additional 12 cards are required to monitor 12 non-overlapping channels, leading to a total of 11+12=23 cards.

Two approaches are used to solve the problem: the first one is to use statistical information to listen to the most possible channels. In practice, most APs simply use the factory settings, which are limited to a few channels. To obtain the statistical information, law enforcement can perform a field training to identify all existing channels from the received beacon frames or probe response traffic, and only listen on those channels. If the statistical information indicates more existing channels than the number of available network cards, our second approach, frequency hopping, can be used. In particular, law enforcement can hop between a series of channels and dwells on each channel for a period of time to collect wireless traffic.

During data communication, a mobile device communicates with only one AP. In order to obtain the set of APs communicable with a mobile device, a passive and an active approach can be used to collect probing traffic. In a passive approach, law enforcement passively sniffs on wireless traffic and does not interfere with the 802.11 protocol. In an active approach, law enforcement may manipulate the wireless frames, exploit the protocols, and make the suspects send extra frames. As such, the passive approach technique simply listens on those active scanning frames and records the set of APs which respond to a mobile's probe request frames. Law enforcement may also utilize an active approach. It is possible to force a mobile into a position where it will automatically start sending probe request frames by exploiting the 802.11 protocol. The third challenge is discussed below:

A. Forensic localization

There are three scenarios for forensic localisation based on the values collected from the external sources. 1) the location and maximum transmission distance of each AP are known through external knowledge, 2) the location is known but the distance is unknown, and 3) neither the location nor the distance are known.

Scenario 1: Both AP locations and maximum transmission distances are known

When the location and maximum transmission distance of each AP is known through external knowledge, calculate the maximum coverage area for each AP as a disc-centred at the AP's location with radius of the maximum transmission distance. Such a disc is a superset of all locations that can communicate with the AP. For locating a mobile device in this scenario, a simple disc intersection approach can be used. In this approach, firstly, calculate the intersection of the maximum coverage areas of all APs that the mobile device has communicated according to the monitored probing traffic. Then, the intersected area is used as an estimation of the mobile device's location. It is evident from this approach that as long as the APs' locations and maximum transmission distances are accurate, the mobile device's real location is always covered in the intersected area. Thus, the main challenge for this approach is how small the intersected area can be. The smaller the size is, the more accurate the estimation will be.

When a mobile device can only communicate with one AP, the intersected area is the maximum coverage area of the AP, and the disc-intersection approach is essentially reduced to the nearest AP approach. Experiments show that a mobile device can usually communicate with a large number of APs in practice, particularly in urban areas.

If the distribution of APs is not uniform but is pre-known, then one can always use rejection sampling to produce a sub sample that follows the uniform distribution.

The intersected area is roughly inversely proportional with the number of communicable APs. Assume in an area, APs are uniformly distributed. The density is ρ. If there are k APs within the mobile receiver range, then

$$k=\pi r^2\rho$$

Scenario 2: Only AP locations are known

In practice, an important challenge for the disc-intersection approach is that the maximum transmission distance varies between different APs, and may not be known through external knowledge. A simple approach is to set the maximum transmission distance to a predetermined value, such as the theoretical upper or lower bound on the transmission distance. Nonetheless, if the value is set too high, the intersected area may become extremely large. If the value is set too low, the mobile device's real location might not be covered by the intersected area (or the area may even become empty). The following theorem shows the relationship between the performance of the disc-intersection approach and the maximum transmission distance. Theorem: When APs with maximum transmission distance r are uniformly distributed and the disc-intersection approach set estimated distance R, then if R >=r, the expected size of the intersected area for a mobile device communicable with k APs is

$$CA = \frac{1}{\pi^{k-1} r^{2k}} \int_0^{2R} \left(r^2 \cos^{-1}\left(\frac{x^2 + r^2 - R^2}{2xr}\right) + R^2 \cos^{-1}\left(\frac{x^2 + R^2 - r^2}{2xR}\right) - \frac{\sqrt{((r+R)^2 - x^2)(x^2 - (r-R)^2)}}{2} \right)^k dx^2$$

If R < r, the probability that the intersected area covers the real location of the mobile device is

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

$$p=(r/R)^{2k}$$

A linear-programming-based approach can be used to estimate the maximum transmission distance of an AP from the monitored probing traffic. A key point is that if a mobile device can observe two APs within a short period of time, then the maximum transmission distances of the two APs, r1 and r2, must satisfy $r1 + r2 >= d_{12}$, where d_{12} is the distance between the two APs. Also such a distance can be computed from their locations pre-known to law enforcement. On the other hand, if over a sufficient amount of time, the two APs have never been observed by the same mobile device, then it is highly likely that $r1 + r2 < d_{12}$.

Scenario 3: When no AP information is available

When no AP information is available, law enforcement must first collect a set of training data tuples before being able to locate the monitored mobile devices. Each training data tuple consists of two parts: 1) an identifier which consists of the longitude and latitude of a training location, and 2) a set of APs which a mobile device can communicate with at the training location.

After the training data tuples are collected, compute the location of APs by using, again, the disc-intersection approach. In particular, for each AP, derive the intersection of discs centred at the training locations, which can communicate with the AP. Nonetheless, the exact radius of the discs is unknown and cannot be computed using the linear-programming-based approach due to lack of AP location information. Thus, it is suitable to use a theoretical upper bound as the radius, and then estimate the AP's location as the centroid of the intersected area. After the APs' locations are estimated, the APs' maximum transmission distances using the above mentioned linear-programming-based approach can also be estimated. Then the disc-intersection approach can be used to locate the monitored mobile devices.

IV. APCL

The effectiveness of APCL Access point Co-ordinate d Location, to some extent, relies on the existence of multiple APs around the attacker. Given the wide deployment of WLAN, this is very likely to be the case.

The initial estimation of the attacker's position is the coverage region of its home AP, which is the region determined by the maximum communication range between an AP and an attacker. Since the initial estimation region may be too large to locate the attacker, APCL disassociates the attacker from its home AP. This is done by sending a disassociation frame to the attacker to terminate this individual connection. This operation is only targeted at the attacker and does not interfere existing legitimate WLAN users. A specified Reason Code is filled in the Reason Code field of the disassociation frame. Several legitimate reasons can be used to disassociate the attacker [10], e.g., bad link quality, overcrowded wireless access, etc. To continue its attack, the attacker has to reconnect to one of its neighbouring APs. Once the reconnection is established and the attacker starts to send its attacking traffic through this new home AP again, this new home AP can be quickly identified through attack traffic trace back, traffic analysis or wireless device identification techniques. After successful identification, APCL narrows the position estimation region down to the intersection of the previously connected and the new home APs' coverage regions. This process continues until there are no neighbouring APs of the attacker that can help APCL to narrow the position estimation.

The design of APCL is based on the following assumptions. Multiple APs can be coordinated to locate the attacker. Using existing network attack trace back, traffic monitoring, and wireless device identification techniques, the attacker has been successfully identified and traced down to its home AP before APCL is started to locate the attacker. In addition, the system assumes that there is only one attacker in the coverage region of its home AP and the attacker is relatively stationary during the localization process. Each AP is assumed to have correct knowledge about its position. It is also assumed that a rough estimation about the maximum possible communication range between the AP and the attacker can be attained. This assumption is based on the fact that the duration of ACK time-out in the IEEE 802.11 standard limits the maximum distance between an AP and its client to be around 150 m. This limit cannot be changed by larger transmission power or a more sensitive antenna. Instead, long range communications beyond 150 m have to modify the value of ACK time out in both the AP and the attacker, which is impossible since the attacker cannot modify AP's configurations.

In each activation and disassociation step, APCL consumes certain network resources. To minimize the network resource consumption, it is desirable to locate the attacker within the least number of steps. Moreover, to minimize the final localization error, APCL also needs to find the optimal AP coordination process. Therefore, there arises an interesting problem of identifying the optimal AP activation sequence.

The optimal AP activation sequence problem can be modelled as a finite horizon discrete Markov decision process (MDP) based on the following two reasons. Firstly, APCL only has a partial control over the whole process. While it is possible to control which APs to activate in one step, it is impossible to control which activated AP the attacker actually reconnects to. Secondly, the activation process has the Markov property, i.e., given any current state, the transition to the next state is only dependent on the current estimation region and is independent of the previous localization process. Hence, MDP is the appropriate model for computing the optimal AP activation sequence.

V. EXPERIMENTS

The system takes as input the location of AP and maximum transmission area as per the availability. If this data is unavailable it goes for the training phase to collect training data. This is done by the use of wardriving tool,

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

which completes the first phase of the system. Based on this, the system creates an initial topology of the nodes connected to the network.

The next step includes the creation of wireless receiver chain. This is done by defining an omnidirectional antenna and specifying bandwidth for communication to support maximum coverage area to the system. Experiments reveal that most of the nodes use channels 1,6,11 for communication, so three cards were used to monitor the wireless channels. An additional observation is that, when the wireless card transmit through channel 11, the other two cards monitoring the neighbouring channels can barely recognize the packets. This is depicted in the following graph.

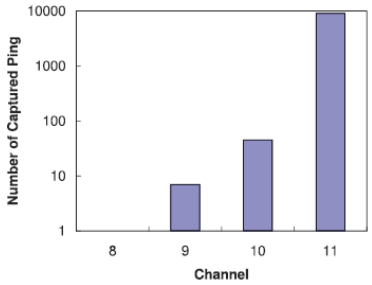

Fig 1 Channel leakage

As claimed earlier the feasibility of the project is valid only when the mobile node is sending packets there by enabling them to be monitored. In case when the mobile node is not transmitting any probe and not disconnected from the network, the system uses active mechanism. This approach is done by disassociation attack which forces the mobile node to send packets. The experiment shows that during this process the mobile node sent packets and all communicable AP's respond to the probing traffic.

The error ratio performance with respect to the network density is shown in Figure. As shown in Figure APCL outperforms the traditional centroid method. The error ratio decreases when the network density increases. The average number of actions with respect to the network density is shown in the next Figure The average number of actions increases when the network density increases. The average localization time is still in the range of several seconds.

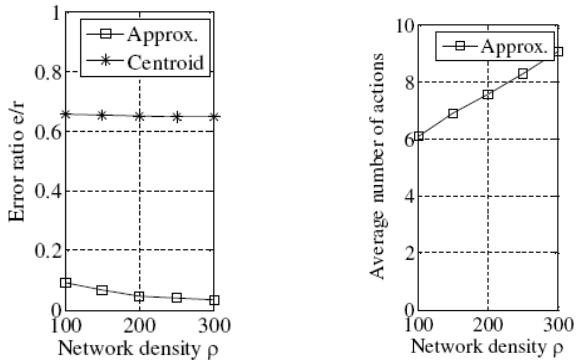

Fig 2 variation of error ratio and average number of actions against network density

VI. CONCLUSION

A novel system that reveals the locations of WiFi-enabled devices in the coverage area of a specialized sniffing system is proposed. The system is proposed for malicious wireless and mobile device tracking. The system is designed to support law enforcement for their forensic application. Thus privacy issue can be neutralized to major extend. This system is used in the attack phase to locate a victim mobile based on AP spatial information from external knowledge or the training phase.

The system is designed with high gain antennas to increase the coverage area. Special wireless packet capture module exists to identify the malicious usage and unauthorized usage of the network. Efficient algorithms are used to accurately track the mobile devices together with mechanisms to identify anonymous usage through cloaking. The system can be used by for forensic application with off the shelf equipments also and features mobile design that enables the system to be quickly deployed to new location. The simple digital display enables the law enforcements ease of use of the system.

REFERENCES

1. M. Grosser and D. Grunwald, "Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking," Proc. MobiSys, May 2003.
2. D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis, "A Study on the Value of Location Privacy," Proc. Fifth ACM Workshop Privacy in Electronic Soc., Oct. 2006.
3. J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, "802.11 User Fingerprinting," Proc. ACM MobiCom, Sept. 2007.
4. B. Gedik and L. Liu, "Location-Privacy in Mobile Systems: A Personalized Anonymization Model," Proc. Int'l Conf. Distributed Computing Systems (ICDCS '05), June 2005
5. K. Ro¨mer, "The Lighthouse Location System for Smart Dust,"Proc. MobiSys, May 2003
6. Roy want, and Jonathan gibbons, "The Active Badge Location System" ACM Trans. Information Systems, vol. 10, no. 1, pp. 91-102, Jan. 1992.
7. D. Singele´e and B. Preneel, "Location Privacy in Wireless Personal Area Networks," Proc. Fifth ACM Workshop Wireless Security (WiSe '06), Aug. 2006.
8. Xiaoguang Xu Hongda Shen, Youzhu LING,"A Novel Method for Indoor Wireless Sensor Network Localization" Journal of Computational Information Systems 7: 12 (2011)
9. P. Bahl and V.N. Padmanabhan, "RADAR: An In-Building RF Based User Location and Tracking System," Proc. IEEE INFOCOM, Mar. 2000
10. N.B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support System," Proc. ACM MobiCom, Aug. 2000.
11. Xinwen Fu,"The Digital Marauder's Map: A Wi-Fi Forensic Positioning Tool" IEEE transactions on mobile computing, vol. 11, no. 3, march 2012
12. Chuan Han, Siyu Zhan, Yaling Yang," Proactive Attacker Localization in Wireless LAN"
13. V. Gupta, "A Characterization of Wireless Network Interface Card Active Scanning Algorithms," master's thesis, Georgia State Univ., 2006.
14. M.Kershaw and J. Wright, "802.11b Firmware-Level attacks,"http://code.google.com/apis/maps, Sept. 2006.
15. J. Bellardo and S. Savage, "802.11 Denial-of-Service Attacks: Real Vulnerabilities and Practical Solutions," Proc. 12th Conf. USENIX Security Symp. (SEC '03), Aug. 2003.
16. "Aireplay," http://www.aircrack-ng.org/doku.php?id= aireplay-ng, 2010.
17. "Aircrack," http://www.aircrack-ng.org, 2010.

18. A.R. Beresford and F. Stajano, "Location Privacy in Pervasive Computing," Proc. Information Hiding Workshop, Mar./Apr. 2003.
19. Y.-C. Hu and H.J. Wang, "Location Privacy in Wireless Networks," Proc. ACM SIGCOMM, Apr. 2005.
20. T. Jiang, H.J. Wang, and Y.-C. Hu, "Location Privacy in Wireless Networks," Proc. MobiSys, June 2007.
21. T. He, C. Huang, B.M. Blum, J.A. Stankovic, and T. Abdelzaher, "Range-Free Localization Schemes for Large Scale Sensor Networks," Proc. ACM MobiCom, Sept. 2003.
22. J. Hwang, Y. Gu, T. He, and Y. Kim, "Realistic Sensing Area Modeling," Proc. IEEE INFOCOM, May 2007.
23. T. Baba and S. Matsuda. Tracing network attacks to their sources. IEEE Internet Computing, vol. 6, 2002.
24. M. Snow and J.-M. Park. Link-layer traceback in ethernet networks. In IEEE Workshop on Local and Metropolitan Area Networks (LANMAN). 2007.
25. S. Northcutt and J. Novak. Network Intrusion Detection. Sams, 2002.
26. V. Brik, et al. Wireless device identification with radiometric signatures. In MobiCom'08. 2008.
27. J. Hightower and G. Borriello. A survey and taxonomy of location sensing systems for ubiquitous computing. UW CSE 01-08-03, Univ. of Washington, Dept. of Computer Sci. and Engineering, Aug. 2001.
28. Y. Zhang, et al. Secure localization and authentication in ultra-wideband sensor networks. IEEE JSAC, vol. 24, no. 4, 2006.

DOI 10.17148/IJARCCE.2016.56169