Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 21 (2013) 430 – 435

Procedia Computer Science

2013 International Workshop on Communications and Sensor Networks (ComSense-2013)

Detecting Primary User Emulation Attacks in Cognitive Radio Networks via Physical Layer Network Coding

Xiongwei Xie[a], Weichao Wang[a]

[a]*Department of SIS*
UNC Charlotte
Charlotte, NC 28223
Email: xxie2@uncc.edu, weichaowang@uncc.edu

Abstract

Primary user emulation (PUE) attacks on cognitive radio networks pose a serious threat to the deployment of this technique. Previous approaches usually depend on individual or combined received signal strength (RSS) measurements to detect emulators. In this paper, we propose a new mechanism based on physical layer network coding to detect the emulators. When two signal sequences interfere at the receiver, the starting point of collision is determined by the distances among the receiver and the senders. Using the signal interference results at multiple receivers and the positions of reference senders, we can determine the position of the 'claimed' primary user. We can then compare this localization result with the known position of the primary user to detect the PUE attack. We design a PUE detection mechanism for wireless networks with trustworthy reference senders. We analyze the overhead of the proposed approach and study its detection accuracy through simulation.

Selection and peer-review under responsibility of Elhadi M. Shakshuki

Keywords: primary user emulation, physical layer network coding, cognitive radio networks

1. Introduction

The dynamic spectrum access technique allows wireless nodes to use spectrum sensing to identify the 'white spaces' in licensed spectrum. The cognitive radios will then opportunistically utilize these white spaces. To avoid any interference with the primary users, a secondary user must leave the occupied channels if it detects a primary user. Therefore, one of the major technical challenges in spectrum sensing is the problem of precisely identifying the signals of the real primary users. The malicious secondary users can mimic the spectral characteristics of primary users to gain priority access to the wireless channels, which is called "primary user emulation" (PUE) attacks.

Existing approaches to detecting the PUE attacks can be divided into two groups: communication oriented and localization oriented. In the first group, the secondary nodes use the spectrum sensing techniques to match the characteristics of the radio signals to those of the primary user. The detection mechanisms include filter and cyclostationary feature detection [1], spectrum decision and channel parameters [2], and shadow senders [3]. In the second group, the researchers use the received signals to estimate the position

1877-0509
Selection and peer-review under responsibility of Elhadi M. Shakshuki
doi:10.1016/j.procs.2013.09.057

of the sender. They have designed different methods to model the communication channels and improve the signal measurement accuracy [4]. Outliers in localization procedures are filtered out to improve the detection accuracy of PUE attacks [5].

In this paper, we propose a PUE attack detection mechanism based on the physical layer network coding (PNC) technique. PNC uses the additive nature of the electromagnetic waves to serve as the coding procedure. In our approach, we estimate the position of a wireless node by letting its radio signals interfere with a reference sender. These interfered sequences will be captured by multiple secondary users. Combining the starting points of signal interference results with their positions, the secondary users will determine a group of hyperbolas on which the wireless sender resides. Then they will compare the intersection point of these hyperbolas with the known position of the primary user to detect the PUE attack.

To turn the approach into a practical solution, research challenges from multiple aspects must be carefully addressed. From the network point of view, we need to verify the authenticity of the received signals and accurately locate the position of the sender. From the security point of view, we need to design mechanisms to identify the false claims of positions and signal interference results provided by malicious nodes.

Our investigation has the following contributions: (1) The research will demonstrate that in addition to improving the bandwidth usage efficiency in wireless networks, physical layer network coding can also be used to detect malicious attacks. (2) The proposed PUE attack detection mechanism does not require the deployment of any special hardware. The assumed trustworthy reference senders already exist in the IEEE standards such as 802.22 and 802.16h. (3) The overhead and detection accuracy of the approach are studied through both theoretical analysis and simulation.

The remainder of this paper is organized as follows. In Section 2, we introduce the basic idea of using PNC for localization. In Section 3, we present the details of the proposed approach. The overhead and detection accuracy of the approach are studied in Section 4. Finally, Section 5 concludes the paper.

2. Localization through Physical Layer Network Coding

2.1. System Assumptions

We assume that the primary user (e.g. a TV station) is located at a fixed position and both the secondary users and the attackers know its position. Since FCC requires all TV towers or radio stations to enforce strict physical security, we assume that the secondary users or the attackers cannot be physically close to the primary user. We assume that all secondary users are equipped with GPS and a secure, lightweight pseudo random bit generator (PRBG) [6]. The authenticity and integrity of the packets are protected by the Message Authentication Code (MAC). The details of packet authentication will be discussed in Section 3.

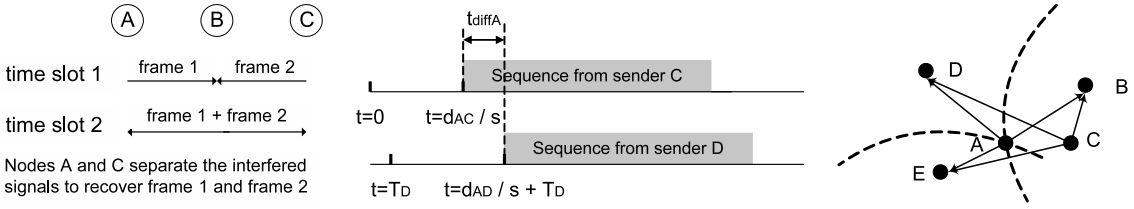
We assume that the attackers also have the GPS device and the PRBG. An attacker can mimic a primary user's radio signals. Multiple attackers can collaborate to conduct a PUE attack. However, the attackers do not have the computation power to compromise the encryption keys or reverse a secure hash function.

2.2. Use PNC to Achieve Localization

Figure 1 illustrates the basic idea of physical layer network coding. In the topology, A and C depend on B to forward the frames between them. In the PNC approach, A and C will send out their packets and B will receive the interference results of the two frames. It will rebroadcast the received signals to both A and C so that they can leverage their knowledge about $frame1$ and $frame2$ respectively to separate the signals and recover the data. Please note that the PNC based mechanism does not require the frames to reach the receiver simultaneously since it can accurately locate the starting point of signal collisions [7].

We can use PNC to calculate the position of a wireless node. We use d_{MN} to represent the distance between two nodes M and N. We use T to represent a specific moment and t to represent a time duration. If radio waves propagate at the speed s, the transmission delay between M and N will be $\frac{d_{MN}}{s}$. In our analysis, we measure the difference between the arriving time of two sequences based on the starting point of signal collisions. We can locate the symbol in the sequence from which the collision starts. Then we can translate this information into a time difference based on the frequency of the radio signals.

Figure 1 illustrates an example of radio signals colliding at wireless receivers. We assume that four nodes A, C, D, and E can receive the signals from each other. We also assume that nodes C, D, and E know their positions. Node A wants to determine its position. Two anchor nodes C and D send out signal sequences that will collide at both A and E. Without losing generality, we assume that C starts sending at $T_C = 0$ and D starts sending at $T_D \geq 0$.



Left: physical layer network coding. Middle: t_{diffA}: difference b/w the arriving time of two sequences at node A. Right: Node A is at the intersection of two hyperbolas.

Fig. 1. Node localization through physical layer network coding.

Node A will receive the sequence from C at $\frac{d_{AC}}{s}$, and the sequence from D at $(T_D + \frac{d_{AD}}{s})$. The difference between the arriving time is $t_{diffA} = (T_D + \frac{d_{AD}-d_{AC}}{s})$. Similarly, we can calculate the difference between the arriving time at node E as $t_{diffE} = (T_D + \frac{d_{ED}-d_{EC}}{s})$. The difference between t_{diffA} and t_{diffE} is:

$$t_{diffE} - t_{diffA} = (T_D + \frac{d_{ED} - d_{EC}}{s}) - (T_D + \frac{d_{AD} - d_{AC}}{s})$$

We simplify this equation and will get:

$$d_{AD} - d_{AC} = (d_{ED} - d_{EC}) + s \times (t_{diffA} - t_{diffE}) \tag{1}$$

Since nodes C, D, and E know their positions, they can calculate $d_{ED} - d_{EC}$. Using the values of t_{diffA} and t_{diffE}, we can calculate $d_{AD} - d_{AC}$. Since nodes C and D know their positions, node A will reside on one wing of the hyperbola that is jointly determined by the positions of C and D and the value of $d_{AD} - d_{AC}$. We need more hyperbolas to determine the position of node A. We can choose other pairs of anchor nodes to determine more hyperbolas. Node A will be positioned at the intersection point of these hyperbolas.

3. Detecting the PUE Attack

3.1. Overview of the Approach

In the remainder of the paper, we call the sender whose position we try to locate as the **interested sender** (it could be the primary user or an emulator), and the sender of the interfered signals as the **reference sender**. The PNC based localization technique provides a very promising approach to distinguish the real primary user from an emulator: when an unknown signal is detected, a legitimate secondary user can intentionally send out a sequence to interfere with the signal. Other secondary users can capture the interference results and determine the hyperbolas. If the intersection point of the hyperbolas is at the known position of the primary user, the secondary users will leave the channel. Otherwise, they will stay there.

The major challenge that we face is the safety of the approach. Since we cannot distinguish an attacker from a legitimate secondary user, the attackers can participate in the localization procedure. They can send out false information about their positions and interference results to mislead the calculation procedures. Therefore, we must design mechanisms to defend against such attacks. In the following scenario, we assume that trustworthy reference senders exist in the network. This scenario matches the application environments of the IEEE 802.22 [8] and 802.16h [9] network standards. The trustworthy nodes can serve as the **reference senders** during PUE detection. We assume that the signal from a trustworthy sender TR can be correctly received by p legitimate secondary users $\{s_1, s_2, \cdots, s_p\}$ and q attackers $\{m_1, m_2, \cdots, m_q\}$. All these nodes can correctly receive the signals from the real primary user P.

When TR senses the communication channel and detects some signals that could have come from the real primary user, it will initiate the PUE detection procedure. TR will choose a random number as the seed for the PRBG to generate a random bit sequence and use the sequence to fill a data packet. When it sends out the packet, the radio waves from TR will interfere with the signals from the primary user (or an emulator). Message authentication codes (MAC) will be attached to the packets to protect their integrity.

Using the mechanism in [7, 10], the wireless nodes can detect the signal collision and record the interference results. Using the MAC code from TR, they can verify the integrity of the packets. They will then use the PRBG to regenerate the random sequence. Combining the interference results with the regenerated sequence, the receivers can recover the packet from the interested sender. The receivers can then calculate the t_{diff} values based on the starting points of interference. Now every receiver (both legitimate secondary users and attackers) will exchange its position, its t_{diff} value, and the hash result of the recovered packet with its neighbors. The broadcast packets will be protected by the MAC codes so that the receivers can verify their contents. The secondary users can combine the t_{diff} values with the node positions to determine the position of the interested sender. Once the position is determined, secondary users can compare it with the known position of the primary user to determine whether or not they are under a PUE attack.

3.2. Construct a Practical Approach

Who are the trustworthy senders: Several IEEE standards using the cognitive radio technique such as 802.22 (CR for Wireless Regional Area Network) [8] and 802.16h (CR for WiMAX) [9] assume the existence of base stations. Many of these base stations are deployed by the cellular phone/network service providers. Therefore, these base stations can serve as the trustworthy senders. The standards such as 802.22 also require the base stations to have GPS devices and loosely synchronized clocks, which can be used for PNC based localization [11] and hash chain based authentication [12], respectively.

Authentication of the Packets: We have assumed that the wireless nodes can attach a MAC code to the packet to protect its authenticity and integrity. We propose to use the same method as in [3] to accomplish the task. We assume that every node can generate a random number y_i and use a secure hash function to construct an l-entry one-way hash chain $hash^j(y_i)$, $(l \geq j \geq 0)$. If you have the knowledge of $hash^{j_1}(y_i)$, it will be very easy for you to authenticate $hash^{j_2}(y_i)$ when we have $l \geq j_1 \geq j_2 \geq 0$. However, the one-way property will prevent an attacker from calculating an earlier entry in the hash chain.

3.3. Safety of the Approach

We assume that an emulator U tries to impersonate the real primary user P. During the PUE detection procedure, for any legitimate secondary user s_i $(i \in 1 \cdots p)$, it will get the positions and the t_{diff} values from $(p-1)$ legitimate secondary users and q attackers. Since the received information is protected by the MAC code of the senders, the attackers cannot impersonate other legitimate users. Using Equation (1), s_i will alternatively combine its own information with information from the other $p+q-1$ nodes to determine $p+q-1$ independent hyperbolas. Since the position information and t_{diff} values from the legitimate secondary users are true, the $p-1$ independent hyperbolas that are determined based on s_i and s_k $(k = 1 \cdots p, k \neq i)$ will all pass through the position of node U.

To assist the emulator U to defeat the detection procedure, the attackers have to lie about their positions and t_{diff} values. Since information from the attackers contains their MAC code, the same attacker cannot send different position and t_{diff} values to different legitimate secondary users. Now we assume that the real position and t_{diff} value of the attacker m_j are $Posi(m_j)$ and $t_{diff}(m_j)$, respectively. m_j will send out the false information $Posi(\overline{m_j})$ and $t_{diff}(\overline{m_j})$ to the legitimate secondary users. In the following description, we use the overline bar $(\overline{m_j})$ to represent the values calculated based on the false information. For the legitimate secondary user s_i, to allow the hyperbola determined by s_i and m_j to pass through the position of the primary user P, m_j must make the false values satisfy:

$$d_{s_iP} - d_{\overline{m_j}P} = (d_{s_iTR} - d_{\overline{m_j}TR}) + s \times (t_{diff}(s_i) - t_{diff}(\overline{m_j})) \tag{2}$$

To fool as many legitimate nodes as possible, the attacker needs to solve the following problem: given the positions of the primary user P, the emulator U, and the legitimate nodes s_i $(i = 1 \cdots p)$, an attacker

needs to calculate the fake information $Posi(\overline{m_j})$ and $t_{diff}(\overline{m_j})$ so that all hyperbolas determined by m_j and s_i $(i = 1 \cdots p)$ will pass through P.

This problem is similar to the GPS spoofing attack in [13]. In their approach, the authors study the relationship between the number of legitimate receivers to be fooled and possible positions of the satellite impersonator. The results are shown in Table 1. In our approach, since emulator U is fixed, the satellite impersonator is replaced by the fake position information $Posi(\overline{m_j})$ of m_j. For example, if we want to mislead the localization results of four secondary users, m_j must be positioned at one of the two points. From Table 1, we find that it is almost impossible to satisfy the requirements when there are more than three legitimate secondary users in the neighborhood since all malicious nodes will be located at those two points.

Table 1. Relationship b/w number of victims and possible positions of the emulator

number of victims	possible positions of the emulator
2	set of hyperboloids
3	set of intersections of two hyperboloids
4	set of two points
≥ 5	set of specific points

Based on the results in [13], we adopt the following scheme to determine the position of the interested sender. We will choose a threshold value *thresh*. For a secondary user s_i, only when there are at least *thresh* independent hyperbolas with s_i as one of the focal points passing through the same point, it will be used as the position of the interested sender. If multiple positions satisfy this requirement, s_i will choose the position with the largest number of hyperbolas as the interested sender.

4. Analysis and Simulation

4.1. Overhead of the Proposed Approach

Since the proposed approach incurs very little storage overhead at the secondary users, our analysis will focus on the computation and communication overhead. The majority of the computation overhead is caused by solving the hyperbolas to determine their intersections. Since a hyperbola can be treated as a second-degree equation in the Cartesian coordinates, determining the intersections of two hyperbolas can be viewed as a procedure to solve two second-degree equations. Several mechanisms to efficiently calculate the intersections of hyperbolas have been proposed [14]. In [15] the authors propose a mechanism that uses only simple add and shift operations in the computation.

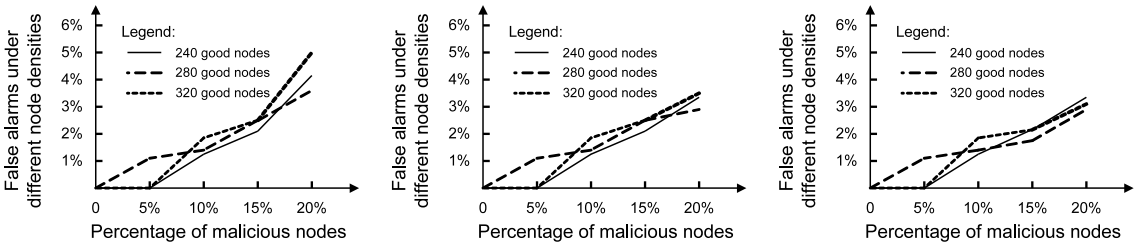
The majority of the communication overhead comes from the exchange of the positions, t_{diff} values, hash results, and MAC codes. Every secondary user needs to send out its own information and receive $p + q - 1$ copies from other nodes. We assume that the packets sent out by the secondary users contain l bytes. Therefore, all users need to send out at most $l \times (p + q - 1 + 1 + 1) = l \times (p + q + 1)$ bytes. If we assume that l =128 Bytes, and $p + q$ has the value of 10, in every round of PUE detection the secondary users need to send out $1.4K$ Bytes altogether, which can be easily handled by modern wireless devices.

4.2. Simulation Results

We assess the detection accuracy of the proposed approach through simulation. We assume a network area of $2000 \times 2000m^2$. Both the legitimate secondary users and the attackers are randomly and uniformly distributed in the network. The radio communication range is $250m$. We assume that the trustworthy senders are also randomly distributed in the network and every secondary user (both legitimate and malicious) is covered by at least one sender. We study the impacts of the legitimate user and attacker densities, and the selected threshold value on the detection accuracy. We focus on false negative alarms, in which an emulator is incorrectly identified as the primary user.

Figure 2 illustrates the false alarm rates under different node densities and threshold values. For each curve, we have a constant number of legitimate users (good nodes) in the network and we introduce different numbers of attackers. From the simulation results, we can see that for different node densities, their curves will stay close to each other when the percentage of attackers is the same. This can be explained as follows.

As the density of the attackers increases, they can cheat more legitimate secondary users under the same threshold value. However, since the density of the legitimate users also increases, their ratio will stay the same. We can also find that when the threshold value increases, the false alarm rate starts to decrease since more attackers are needed to cheat a single legitimate user.



From left to right: the selected threshold values are 4, 5, and 6, respectively. For each curve, we have a constant number of legitimate users in the network and change the number of attackers.

Fig. 2. Detection accuracy under different threshold values and node densities.

5. Conclusion

In this paper we propose a PUE detection mechanism for cognitive radio networks based on physical layer network coding. The analysis shows that the difference between the starting points of interference at two receivers is restricted by the positions of the senders. Using a trustworthy node as the reference sender, we can determine multiple hyperbolas on which the interested sender resides. To turn this mechanism into a practical approach, we study several problems in the network. We design the PUE detection mechanism and study its overhead and the detection accuracy.

Immediate extensions to our approach consist of the following aspects. First, we will implement the proposed approach in software defined radio and test it in real network environments. Second, we will extend our approach to the environments in which the reference senders could be malicious. Finally, we will investigate using physical layer network coding to detect other attacks on wireless networks.

References

[1] D. Pu, Y. Shi, A. Ilyashenko, A. M. Wyglinski, Detecting primary user emulation attack in cognitive radio networks, in: IEEE Global Telecommunications Conference (GLOBECOM), 2011, pp. 1–5.
[2] T. Yang, H. Chen, L. Xie, Cooperative primary user emulation attack and defense in cognitive radio networks, in: International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), 2011, pp. 1–4.
[3] Y. Liu, P. Ning, H. Dai, Authenticating primary users' signals in cognitive radio networks via integrated cryptographic and wireless link signatures, in: Proceedings of IEEE Symposium on Security and Privacy, 2010, pp. 286–301.
[4] Z. Jin, S. Anand, K. P. Subbalakshmi, Mitigating primary user emulation attacks in dynamic spectrum access networks using hypothesis testing, SIGMOBILE Mob. Comput. Commun. Rev. 13 (2) (2009) 74–85.
[5] O. León, J. Hernández-Serrano, M. Soriano, Robust detection of primary user emulation attacks in ieee 802.22 networks, in: Proceedings of the International Conference on Cognitive Radio and Advanced Spectrum Management, 2011, pp. 51:1–51:5.
[6] R. Jenkins, Isaac, in: International Workshop on Fast Software Encryption, 1996, pp. 41–49.
[7] S. Katti, S. Gollakota, D. Katabi, Embracing wireless interference: analog network coding, in: SigComm, 2007, pp. 397–408.
[8] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, W. Caldwell, IEEE 802.22: the first cognitive radio wireless regional area network standard, Comm. Mag. 47 (1) (2009) 130–138.
[9] IEEE 802.16 License-Exempt (LE) Task Group, IEEE 802.16 Draft Ver 15 (2010).
[10] W. Wang, D. Pu, A. Wyglinski, Detecting sybil nodes in wireless networks with physical layer network coding, in: IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2010, pp. 21–30.
[11] Z. Li, D. Pu, W. Wang, A. Wyglinski, Node localization in wireless networks through physical layer network coding, in: Proceedings of IEEE Globol Communications Conference (GLOBECOM), 2010, pp. 1–5.
[12] A. Perrig, R. Canetti, J. D. Tygar, D. Song, The tesla broadcast authentication protocol, RSA CryptoBytes 5 (2) (2002) 2–13.
[13] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, S. Capkun, On the requirements for successful gps spoofing attacks, in: Proceedings of the ACM conference on Computer and communications security (CCS), 2011, pp. 75–86.
[14] M. Leonardi, A. Mathias, G. Galati, Two efficient localization algorithms for multilateration, International Journal of Microwave and Wireless Technologies 1 (2009) 223–229.
[15] E. Doukhnitch, M. Salamah, General approach to simple algorithms for 2-d positioning techniques in cellular networks, Computer Communications 31 (10) (2008) 2185–2194.