Security Aspects in Emerging Wireless Networks

ZORAN BOJKOVIC, BOJAN BAKMAZ, MIODRAG BAKMAZ
University of Belgrade
SERBIA
z.bojkovic@yahoo.com  http://www.zoranbojkovic.com

Abstract: - Advances in wireless technologies provide both benefits and challenges when it comes to security. Communications over wireless channels are, by nature, insecure and easily susceptible to various kinds of threats. Despite the current efforts from academia and industry, the security paradigms protecting the confidentiality of wireless communications remain open issue. This paper seeks to provide a comprehensive survey of challenging issues and prospective techniques regarding security in emerging wireless networks. Major security aspects are analyzed through physical layer, together with security of cognitive radio and direct communications as promising solutions for an efficient utilization of scarce frequency spectrum.


1 Introduction
Over the past few years, wireless communication has experienced an unprecedented growth in data traffic, spurred by the popularity of various intelligent devices and rich multimedia content, as well as the rapid increase in the points of attachment density. While, academia researches are focused on robust and efficient technologies for the future wireless systems (i.e., heterogeneous dense networking, massive antenna systems, millimeter wave, etc.) [1], at the same time, the industry is undertaking 5G standardization. Despite the current efforts from academia and industry, the security paradigms protecting the confidentiality of wireless communications remain open issue. The increasing demand of users for various wireless multimedia communication services has led to the development and to the coexistence of different and often incompatible technologies with unique applications and characteristics. To integrate several heterogeneous networks into a single architecture, there are a number of challenges that must be addressed. One of the most important challenges is secure interoperability. Service or network providers planning for implementation of security mechanisms need to consider the nature of the security threat, the strength of security needed, the location of security solutions, the cost of available mechanisms, the speed and the practicality of mechanisms, and interoperability.

This paper intends to provide a comprehensive survey of challenging issues and prospective techniques regarding security in emerging wireless networks. Differing from the traditional security approach which protects data through cryptographic techniques, physical layer safeguarding is identified as a promising strategy that provides secure wireless transmissions by smartly exploiting the imperfections of the communications medium. In the second part of the paper, the security aspects in emerging cognitive radio are analyzed together with security in direct communications as promising solutions for an efficient utilization of scarce frequency spectrum.

2 Physical Layer Security
Recently, physical layer security has become an emerging topic in wireless communications [2]. It is identified as a promising strategy that provides secure transmission by smartly exploiting the imperfections of the communications medium. With careful planning and implementation, physical layer security will protect the communication phase while cryptography will protect the processed data after the communication phase. In this way, a well-integrated solution that efficiently safeguards sensitive and confidential data will be obtained.

Physical layer security offers two major advantages compared to cryptography, making it particularly suitable for the emerging wireless networks [3]. First, physical layer security techniques do not depend on computational complexity, which implies that the achieved level of security will not be compromised even if the unauthorized devices have powerful computational
capabilities. Second, physical layer security techniques have a high scalability. In the heterogeneous networks, devices are always connected to the nodes with different powers and computation capabilities at the different levels of the hierarchical architecture. Also, devices always join in or leave the network at random time instants, due to the decentralized architecture. As a consequence, cryptographic key distribution and management become very challenging issues. To cope with this, physical layer security can be used to either provide direct secure data communication or facilitate the distribution of cryptographic keys.

2.1 Physical Layer Security Techniques

The application of physical layer security schemes makes it more difficult for attackers to decipher transmitted information. The existing physical layer security techniques can be classified into major categories [4]: theoretically secure capacity, channel characteristics, coding, power and signal detection approaches.

The secrecy capacity, defined as the maximum transmission rate at which the eavesdropper is unable to decode any information [5], is equal to the difference between the two channel capacities:

\[ C_s = C_M - C_W = \frac{1}{2} \log_2 \left( 1 + \frac{P}{N_M} \right) - \frac{1}{2} \log_2 \left( 1 + \frac{P}{N_W} \right) \]

where \( C_M \) is the capacity of the main channel and \( C_W \) denotes the capacity of the eavesdropper’s channel. Here, \( P \) corresponds to the average transmit signal power, while \( N_M \) and \( N_W \) are power of the noise in the main channel and the eavesdropper’s channel, respectively.

Information-theoretic security is an average-information measure. The system can be designed and tuned for a specific level of security. On the other hand, it may not be able to guarantee security with probability. Furthermore, it requires knowledge about the communication channel that may not be accurate in practice. A few systems (e.g., quantum cryptography) have been deployed, but the technology is not widely available due to its implementation costs.

The perspective techniques that have been proposed to increase security based on the exploitation of channel characteristics include the following:

- Algebraic channel decomposition multiplexing (ACDM) precoding scheme [6], in which the transmitted code vectors are generated by singular value decomposition of the correlation matrix, that describes the channel’s characteristic features between the transmitter and the intended receiver. Because any potential transmitter–eavesdropper channel is going to have a different multipath structure, the eavesdropper’s ability to detect and decode the transmissions can be severely reduced.

- Randomization of multiple-input, multiple-output (MIMO) transmission coefficients [7] is a procedure in which the transmitter generates a diagonal matrix dependent on the impulse response of the transmitter–receiver channel. This diagonal matrix has a unique property that makes the matrix undetectable to the attackers but easily detectable to the intended receiver. This method reduces the signal interception capability of the intruder and leads to a blind deconvolution problem due to the redundancy of MIMO transmissions. The proposed scheme indicates that the physical layer technique can assist upper layer security designs by providing secret key agreement with information-theoretic secrecy. As an illustrative example Fig. 1 shows the BER performance of a legitimate receiver and the eavesdropper with respect to the ratio of the variance of the artificial noise to that of the legitimate receiver’s channel noise (\( \alpha \)) for different ratio of the energy per bit to the artificial noise (\( \beta \)). A random MIMO 4×4 system with BPSK modulation is adopted.

![Fig. 1. BER performance of a legitimate receiver and an eavesdropper when artificial noise is added at the transmitter [4].](image-url)
variance of the artificial noise while simultaneously increasing bit energy such that \( \beta \) stays unchanged. The artificial noise can, thus, be adjusted with the aid of experimental data to choose an operating point that maximizes the performance gain between the legitimate and eavesdropper receivers.

- Radiofrequency fingerprinting system [8] consists of multiple sensor systems that capture and extract corresponding features from each received signal. An intrusion detector processes the feature sets and generates a dynamic fingerprint for each internal source identifier derived from individual packets. This system monitors the temporal evolution of each fingerprint and issues an intrusion alert when a strange fingerprint is detected, thus helping distinguish an intruder from a legitimate user.

The objective of coding approaches is to improve resilience against jamming and eavesdropping [4]. These approaches mainly include the use of error correction coding and spread spectrum coding. In a conventional cryptographic method, a single error in the received cipher text will cause a large number of errors in the plain text after channel decoding. A combination of turbo coding and advanced encryption standard cryptosystem can be used to set up a secure communication session. The main advantages of secure turbo code include higher-speed encryption and decryption with higher security, smaller encoder/decoder size, and greater efficiency.

On the other hand, spread spectrum is a signaling technique in which a signal is spread by a pseudonoise sequence over a wide frequency band with bandwidth much wider than that contained in the frequency ambit of the original information. The main difference between conventional cryptographic systems and spread-spectrum systems lies in their key sizes. Traditional cryptographic systems can have very large key spaces. However, in a spread spectrum system, the key space is limited by the range of carrier frequencies and the number of different sequences.

Data protection can also be facilitated using power and signal detection approaches. The usual schemes in these approaches involve the use of directional antennas and the injection of artificial noise. Directional transmission can improve spatial reuse and enlarge geographical coverage, as beam width is inversely proportional to peak gain in a directional antenna. If an omnidirectional antenna is used, a node in the coverage range of a jammer would not be able to receive data securely. On the other hand, if a directional antenna is used, the node would still be able to receive data from the directions not covered by the jamming signals. Hence, the deployment of directional antennas can improve wireless network capacity, avoid physical jamming attacks, and enhance data availability.

Artificial noise is generated using multiple antennas or the coordination of relaying nodes [9]. This noise is utilized to impair the intruder’s channel, but it does not affect the intended receiver’s channel because the noise is generated in the null-subspace of the legitimate channel. Relying on artificial noise, secret communications can be achieved even if the intruder enjoys a much better channel condition than the intended receiver.

2.2 Physical Layer Security in Heterogeneous Environment

Heterogeneous network (HetNet) is a promising architecture with concurrent operation of different technologies, as well as various base stations (BSs) classes (i.e., macro, pico, and femto) [10]. This approach can provide a flexible coverage and improved spectral and energy efficiency. Overlaying different classes of BSs can also potentially provide a solution for the growing data traffic, especially when the transmission is optimized to take advantage of the HetNets characteristics. It is obvious that HetNet is different paradigm from conventional macrocell-only wireless networks.

In traditional cellular networks, it is typically assumed that mobile users select the strongest BS to connect such that the best channel quality with the highest SINR is obtained. Accordingly, the physical layer security technologies in the open literature are designed based on this assumption. However, in the HetNet environment, such a selection causes a load balancing problem. This is due to the fact that the BSs with high transmit power and large coverage areas are often overloaded, whereas the BSs with low transmit power and small coverage areas are often very lightly loaded. Such an unbalanced load is detrimental to the ubiquitous applications of multimedia services with stringent delay constraints and high power consumption. As such, the unbalanced load should be addressed in the design of physical layer security.

In order to secure transmission and overcome the unbalanced load problem, new security-oriented mobile association policies are required to monitor and balance the instantaneous load of BSs [3]. In designing these policies, the optimization of secrecy performance, e.g. the secrecy rate and the secrecy outage probability, should be prioritized. Under this prioritization, some intelligent mobile association
policies can be developed such that the mobile users are wisely assigned to some BS based on the achievable secrecy performance, the instantaneous load, and other factors such as the transmit power, coverage area, and BSs density. Network security designers should develop new cooperative strategies to allow neighboring heterogeneous BSs to exchange the secure data, the instantaneous load of themselves, and other factors of the network for achieving close-to-maximum secrecy performance.

3 Security Aspects in Cognitive Radio Networks

Cognitive radio is regarded as an emerging technology, which equips wireless devices with the capability to adapt their operating parameters based on the radio environment, in order to utilize the scarce radio frequency spectrum in an efficient and opportunistic manner [11]. However, cognitive radio networks (CRNs) are vulnerable to various attacks because they are usually deployed in unattended environments and use unreliable wireless medium. Moreover, it is not a simple task to organize the implementation of security defenses in CRNs. One of the major obstacles in deploying security in CRNs is that they have limited computational and communication capabilities. Security mechanisms, including trust management, have the ability to secure CRNs against attackers. Specific CRN applications have some unique features and correspondingly, some specific security requirements.

3.1 General Security Requirements in CRNs

CR technology is more susceptible to attack compared with traditional wireless networks due to its intrinsic nature. Although security requirements may vary in different application environments, there are in fact some general requirements that provide basic safety controls such as [12]: access control, confidentiality, authentication, integrity, identification, nonrepudiation, and availability.

Access control is a security requirement for the physical layer that restricts the network’s resources to authorized users. Because different secondary users (SUs) coexist in CRNs, collisions can happen if they simultaneously move to and use the same spectrum band according to their spectrum sensing results. Thus, the access control property should coordinate the spectrum access to avoid collisions.

Confidentiality is closely related to integrity. Although integrity ensures that data is not maliciously modified in transit, confidentiality ensures that the data is transformed in such a way that it is unintelligible to an unauthorized entity. This issue is even more pronounced in CRNs, in which the SU’s access to the network is opportunistic and spectrum availability is not guaranteed.

Authentication has the primary objective of preventing unauthorized users from gaining access to protected systems. It can be considered as one of the basic requirements for CRNs because there is an inherent requirement to distinguish between primary users (PUs) and SUs. An authentication problem occurs in CRNs when a receiver detects a signal at a particular spectrum, that is, how can a receiver be sure that the signal was indeed sent by the primary owner of the spectrum? According to [13], it is practically impossible to conduct authentication in CRNs other than in the physical layer. For example, a CR receiver is able to receive signals from TV stations and process them at the physical layer, but it may lack the component to understand the data in the signals. Therefore, if the authentication depends on the correct understanding of the data, at upper layers, the CR receiver will be unable to authenticate the PU. One way is to allow PUs to add a cryptographic link signature to its signal, so the spectrum usage by PUs can be authenticated.

Integrity is of importance in a wireless environment because, unlike their wired counterparts, the medium is easily accessible to intruders. It is related to the detection of any intentional or unintentional changes to the data occurring during transmission. Data integrity in CRNs can be achieved by applying higher cryptographic techniques.

Identification is one of the basic security requirements for any communication device, whereby a user is associated with his unique identifier. A tamper-proof identification mechanism is built into the SU unlicensed devices. It would be advantageous for a CR to know how many networks exist, how many users are associated with each network, and even certain properties about the devices themselves. To achieve this level of information, it is essential for a CR to gather an accurate notion of the RF environment. Service discovery and device identification provide the necessary building blocks for constructing efficient and trustworthy CRNs.

Nonrepudiation techniques prevent either the sender or receiver from denying a transmitted message. In CRNs, if malicious SUs violating the protocol are identified, nonrepudiation techniques can be used to prove the misbehavior and disassociate/ban the malicious users from the...
secondary network. The proof of an activity that has already happened should be available in CRNs.

Availability refers to the ability of PUs and SUs to access the spectrum in CRNs. For PUs, availability refers to being able to transmit in the licensed band without harmful interference from SUs. On the other hand, for SUs, availability refers to the existence of chunks of spectrum, in which the SU can transmit without causing harmful interference to the PUs. In CR, one of the important functions of this service is to prevent energy starvation and denial of service (DoS) attacks, as well as misbehavior.

3.2 Characteristic Attacks in CRNs

Depending on their target in security requirements, attacks in CRNs can be broadly categorized as [12]:

- **Selfish attacks** occur when an intruder wants to use the spectrum with higher priority. This attack meets its target by misleading other unlicensed users to believe he or she is a licensed user. In that way, the adversarial user can occupy the spectrum resource as long as that user wants. This selfish behavior does not obey the spectrum sharing scheme [14]. Selfish SUs increase their accessing probability by changing the transmission parameters to enhance their own utilities by degrading the performance of other users. Hence, the CRN performance is degraded.

- **Malicious attacks** are related to the cases when the adversary prevents other unlicensed users from using the spectrum and causes DoS. These types of attacks drastically decrease the available bandwidth and break down the whole traffic.

There are other different types of attacks; for example, attacks on spectrum managers [15]. If the spectrum manager is not available, communication between CR nodes is not possible. The spectrum availability should be distributed and replicated in CRNs, whereas the attack can be prevented by specific pilot channels in the licensed band. As for eavesdropping on the transmission range of CR, it is not limited to a short distance because it is using unlicensed bands.

In Table 1, characteristic attacks in CRNs are emphasized and classified depending on various protocol stack layers. Also, these attacks can adversely affect the final layer of the communication stack because protocols that run at the application layer rest on the services provided by lower layers. Most of these attacks are comprehensively analyzed in [16]. However, adversaries can launch attacks targeting multiple layers. These are also known as cross-layer attacks and can affect the entire cognitive cycle because attacks at all layers become feasible [17].

Table 1. Characteristic attacks in CRNs.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Security attack</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport layer</td>
<td>Key depletion attack</td>
<td>Because of a high number of sessions the number of key establishments can increase the probability of the same key being used twice.</td>
</tr>
<tr>
<td></td>
<td>Jellyfish attack</td>
<td>An attacker causes the victim node to switch from one to another frequency band, causing considerable delay.</td>
</tr>
<tr>
<td></td>
<td>Lion attack</td>
<td>A malicious node actually causes the jamming to slow down the throughput of the TCP by forcing handovers frequency.</td>
</tr>
<tr>
<td>Network layer</td>
<td>Network endo-parasite attack</td>
<td>The malicious nodes attempt to increase the interference at a heavily loaded high priority channel.</td>
</tr>
<tr>
<td></td>
<td>Channel endo-parasite attack</td>
<td>A compromised node launches attack by switching all its interfaces to the channel that is being used by the highest priority link.</td>
</tr>
<tr>
<td></td>
<td>Low cost Ripple effect attack</td>
<td>Compromised node can transmit the misleading channel information and forces other nodes to adjust their channel assignments.</td>
</tr>
<tr>
<td>Link layer</td>
<td>Biased utility attack</td>
<td>A malicious SU can intentionally tweak parameters of the utility function to increase its bandwidth.</td>
</tr>
<tr>
<td></td>
<td>Asynchronous sensing attack</td>
<td>A malicious SU can transmit asynchronously instead of synchronizing the sensing activity with other SUs in the network during sensing process.</td>
</tr>
<tr>
<td></td>
<td>False feedback attack</td>
<td>False feedback from one or a group of malicious users can make other SUs to take inappropriate action and violate the protocol terms.</td>
</tr>
<tr>
<td></td>
<td>Intentional jamming attack</td>
<td>The malicious SU jams PUs and other SUs by continuously transmitting in a licensed band.</td>
</tr>
<tr>
<td></td>
<td>Primary receiver jamming attack</td>
<td>A lack of knowledge about the location of primary receivers can be used to intentionally cause harmful interference to a victim primary receiver.</td>
</tr>
<tr>
<td>Physical layer</td>
<td>Sensitivity amplifying attack</td>
<td>Some PU detection techniques have higher sensitivity towards primary transmissions with a view to prevent interference to the primary network.</td>
</tr>
<tr>
<td></td>
<td>Overlapping attack</td>
<td>Transmissions from malicious entities can cause harm to PUs and SUs in CRNs belonging to the same geographical domain.</td>
</tr>
</tbody>
</table>
3.3 Secure Spectrum Management
One of CR’s functions is to detect spectrum holes by spectrum sensing. It keeps monitoring a given band and captures the information. CR users can temporarily use the spectrum holes without creating any harmful interference to the PUs. CR must periodically sense the spectrum to detect the presence of incumbents and quit the band once detected. The detection techniques that are often used in local sensing are energy detection, matched filter, and cyclostationary feature detection [18].

The main benefit of introducing security in the spectrum decision process is a stronger guarantee that the service of PUs will not be significantly disrupted. The resilience of the spectrum decision against malicious attackers protects the secondary network at no additional cost. Many existing dynamic spectrum access protocols make spectrum decisions based under the assumption that all involved parties are honest and there is no malicious outsider that can manipulate the decision process. In [19] it assumed that there is some synchronization among the nodes in the cluster in the network. The time is divided into equal length intervals, whereas the nodes know when each cycle begins and ends. They are also aware of the schedule of the events during a cycle, that is, which node sends its channel availability data, which channels it uses, etc. Three main events are handled in a given cycle: (1) one or more nodes can join the spectrum decision process in a given cluster, (2) the nodes of the cluster send their spectrum sensing data, and (3) the cluster head sends the final channel assignment to the other nodes.

The protocols of different layers of CRNs must be able to adapt to the channel parameters of the operating frequency. Also, they must be apparent to the spectrum handover and related latency. When implementing an algorithm, the best available spectrum should be chosen depending on the channel characteristics of the available spectrum and the QoS requirements of the CR user.

4 Securing Direct Communications
The main goal for direct device to device communications (D2D) is to serve as a means to improve the overall spectral efficiency for mobile systems. Reusing the corresponding spectrum, to D2D user terminals are in a position to form a direct link without the influence of base stations and core networks [20,21]. Together with small cells, D2D communication will form a new underlay tier of low-cost architecture with goal to increase coverage and capacity, offload backhaul, as well as to provide fallback connectivity. For these reasons, D2D communications has become a key topic in both the academic and industrial communities. However, many research works are focused on node discovery, radio resource management, and other aspects, while the issue of security has not yet attract special attention in the open literature.

4.1 Specific Security Threads in D2D Communications
An attacker may break into the Evolved Packet Core (EPC) network and steal or modify the user-specific data (e.g., personal data, users’ privacy, etc.). He can attack the new radio link between users’ devices, since the wireless broadcast nature makes this link vulnerable against the following threads [22]:

- Eavesdropping, where a malicious node passively listens to the radio channel between users’ devices.
- Impersonate attack, where a malicious node pretends to be a legitimate user’s device or BS to get access to the traffic.
- Active attack on traffic data, where a malicious node tries to change the traffic data.
- Active attack on control data, where malicious node tries to change the control data.

4.2 Security Requirements and General Solutions in D2D Communications
Doubtlessly, maintaining data security is an essential task in D2D communications since the transmitted data between connected devices may be overheard by all of the surrounding devices [3]. This task becomes more challenging particularly given the fact that the connected devices may not be able to handle complex signal processing algorithms as network infrastructure do.

One prospective solution known as closed access is proposed in [23]. In closed access the intended device possess a list of "trusted" devices, while the non-listed devices can only communicate with the intended device by getting authenticated in the macro/micro cell tier. Therefore, the establishment of closed access safeguards the data exchange between the intended device and the "trusted" devices against eavesdropping. It is important to notice that closed access may not always be implemented, due to the lack of authentication in the macro/micro cell tiers. In this case, referred to as open access, not only surrounding devices but
geographically close BSs may act as potential eavesdroppers for the connected devices, meaning that they benefit from listening to the transmitted data and pose an acute threat to data security. To address security issues in open access, network designers need to construct new secure data exchange strategies that fully consider the physical characteristics of unintended devices and malicious nodes, e.g. ambiguous location, uncertain mobility, and unknown configuration. In addition, the potential attacks and threats induced by unintended devices and malicious nodes, need to be carefully analyzed.

Besides several security procedures from cellular networks, which can be applied in D2D communications (e.g., authentication and key agreement), network coding also can provide interesting alternatives [24]. For example, if part of the information is coming over the cellular operator and other packets via D2D communication, an attacker may not be able to decode the information. But even if the attacker can overhear all packets, network coding security mechanisms usually rely on encryption of the coding coefficients, requiring the attacker to break the encryption to then decode the data packets. Additionally, the presence of the cellular network might provide a simple mechanism to determine that decoded data is correct (i.e., not corrupted by an active attacker), as the cellular network can provide a small part of the content and also check on the final data, for example, the hash of a given file.

5 Concluding Remarks

Wireless security has been an active and very broad research area since the last decade. Communications over wireless channels is, by nature, insecure and easily susceptible to various kinds of attacks. Regardless of how complex any wireless system becomes, the issue of security should always be approached and managed in a structured and uniform way.

Although physical layers are mainly different considering heterogeneous radio access technologies, security aspects and challenges at this layer are practically common for all wireless systems. Numerous physical layer security approaches have been introduced and evaluated in terms of their abilities and computational complexity. The implementation of physical layer security in a real environment is part of a layered approach, and the design of protocols that combine traditional cryptographic techniques with physical layer techniques is an interesting research direction.

Along with the realization of cognitive radios, new security threats have been raised. Intruders can exploit several vulnerabilities of this new technology and cause severe performance degradation. Security threats are mainly related to two fundamental characteristics of cognitive radios: cognitive capability and reconfigurability. Threats related to cognitive capability include attacks launched by intruders that mimic primary transmitters and transmission of false observations related to spectrum sensing. On the other hand, reconfiguration can be exploited by attackers through the use of malicious code installed in cognitive radios. Furthermore, they face all the classic threats present in traditional wireless networks.

Regarding D2D security, it can be conclude from the above mentioned analysis that the direct radio link is the most vulnerable part of communications. Because this is still a challenging issue, until finding some acceptable solutions, particularly in the domain of physical layer security, reusing the existing LTE-A security architecture as much as possible is mandatory task.

References:


