Chapter 2
Wireless Threats and Key Management Issues

We have discussed how the mobile Internet involves the ongoing convergence of wireless next generation networks (NGNs) with IP-based core networks, and some of the more typical security threats to both components of the infrastructure. The evolution of the mobile Internet also involves the ongoing convergence of different wireless technologies such as third generation (3G) mobile networks, wireless local area networks (WLANs), wireless wide area networks (WWANs), mobile WiMAX, and wireless sensor networks (WSNs). Although WSNs were motivated by military applications for surveillance and national defense, they are increasingly being deployed in a variety of civilian applications such as medical devices, home healthcare, autonomous vehicles, smart structures, supervisory control and data acquisition (SCADA) systems, disaster management, and cybersecurity using real-time distributed control systems (RTDCSs). Similarly, although RTDCSs were motivated by the need to monitor national critical infrastructures such as electric power grids, oil and natural gas supplies, water and wastewater distribution, and transportation systems, they are increasingly being deployed in cybersecurity to protect portions of critical information infrastructures such as wireless networks due to their ability to monitor the activity in a localized area or region.

Many of the privacy and security issues of WSNs and RTDCSs also hold for wireless NGNs in general. Despite their increasing popularity, wireless NGNs remain susceptible to several security threats due to their distributed nature and to the limited computational capabilities and energy supply of mobile nodes and sensor nodes. Typical attack vectors in all types of wireless technologies include threats that are common to both wired and wireless networks, jamming attacks against wireless communication channels, and attacks aimed at key distribution in mobile networks with roaming agreements that have different network access technologies such as WLANs and WWANs. Any trust model supporting the convergence of fixed and mobile networks needs to address these attack vectors by being able to respond to different security policies for each type of network and to different security domains in a flexible fashion. Such a trust model also needs to address the problem of key management for mobility in wireless NGNs to control entities as they roam within the same network or between different types of networks. In particular, methods for secure key derivation and distribution should

M.G. Harvey, *Wireless Next Generation Networks*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-11903-8_2

avoid approaches that have large computation and communication overhead to expedite fast and seamless handovers to minimize data leakage and prevent network service disruptions for authorized users.

2.1 Attack Vectors in Wireless NGNs

In general, whereas traditional wired networks can have distant attackers but adequate perimeter defenses, wireless networks need to be within range of an attacker's wireless device and can be accessed without a perimeter defense. Kisner et al. [1] argue that the security requirements for WSNs can be divided into primary goals and secondary goals. Primary goals include confidentiality, integrity, authentication, and availability. Secondary goals include data freshness, self-organization or distributed collaboration between nodes, node synchronization, and secure localization of secret information such as keys. Similarly, typical attacks against wired and wireless networks can be divided into passive attacks carried out by selfish nodes and active attacks carried out by malicious nodes. Passive attacks include monitoring of network traffic and wireless communication channels for information that can be used to execute active attacks. These attacks typically involve eavesdropping and traffic analysis. In eavesdropping, an attacker acquires data passively by intercepting data exchanges. This type of attack includes decrypting any encrypted data. In traffic analysis, an attacker deduces properties of a data exchange based on personal knowledge of the interacting entities, the duration of the data exchange, timing, bandwidth, and other technical characteristics that are difficult to disguise in packet networks.

Active attacks include masquerading, replay attacks, message modification, denial-of-service (DoS), jamming, and routing attacks. In masquerading attacks, an attacker impersonates an authorized entity to gain access to network applications, resources, or services. Man-in-the-middle attacks involve a double masquerade, where the attacker convinces the sender that she is the authorized recipient of a message on one hand, and convinces the recipient that she is the authorized sender of the message on the other. In replay attacks, an attacker injects malicious packets into the network to disrupt it. Typically, an attacker rebroadcasts a previous message to cause the network to reset, thereby placing it in a vulnerable state, and to gather information for further attacks. Replay attacks most often compromise integrity, but can also compromise authentication, access control or authorization, and non-repudiation. In addition, selective replay attacks can negatively impact both availability and confidentiality. In message modification, an attacker alters packets by inserting changes into them, deleting information from them, reordering them, or delaying them. In wireless networks, message modification is typically accomplished through man-in-the-middle attacks. This type of attack compromises integrity, but can potentially affect all aspects of security. In DoS attacks, the availability of network applications, resources, or services is compromised. In wireless networks, this type of attack is typically accomplished by disabling one of

the interacting entities in the data exchange or by jamming the wireless communication channel.

Jamming refers to the disruption of a communications system such as a wireless network through the intentional use of electromagnetic interference. Jamming blocks a signal or message between two interacting entities by keeping the communications medium busy. An attacker sends a signal with a significantly greater signal strength relative to normal signal levels in the system to flood the channel. Thus, jamming is effectively a form of DoS attack. Jamming can be performed by a single attacker or multiple attackers working together, and can target a specific sender or receiver as well as the entire shared medium. A more sophisticated form of jamming involves the violation of the network protocol, where many more packets than normal are transmitted to increase the number of packet collisions. Wireless networks are more susceptible to jamming than wired networks because of their potential access from covert locations, and because these locations can be easily changed through mobility.

Routing attacks involve interference with the correct routing of packets through a network. Several different types of routing attacks can be carried out at the network layer, including spoofed, altered, or replayed routing information. These attacks can create routing loops, extend or shorten intended routing paths, generate bogus error messages, and increase end-to-end latency, thereby compromising availability. Selective forwarding attacks subvert a node in a network in order to drop selected packets. Similarly, sinkhole attacks subvert a node in a network in order to attract packets to it. Wormhole attacks record packets from one location in a network and retransmit them in another location in the network to disrupt its overall functionality.

Table 2.1 shows how Kisner et al. have summarized these attack vectors in a cybersecurity attack-vulnerability-damage (AVD) model. The AVD model expresses attack vectors as malicious actions that target a vulnerability in a network whose impact affects some aspect of its performance. The model also shows how these attack vectors can be organized according to the level of their severity. Wireless network designers can use the AVD model to assist them in developing a defense-in-depth or layered approach to security.

Current protection strategies against these attack vectors are inadequate. In the case of WSNs, where sensor nodes are often deployed in adversarial or hostile environments, the sensors themselves need to be physically protected to prevent them from being replaced with malicious nodes. In the case of the wireless sensor network itself, several different trust models for secure routing have been proposed to minimize the impact of these attack vectors. Vasilache et al. [2] have compared two popular trust models for WSNs, μRacer and k-FTM. The μRacer trust model, proposed by Rezgui and Eltoweissy, is an adaptive and efficient routing protocol suite designed for sensor-actuator networks as shown in Fig. 2.1. In a sensor-actuator network, distributed controllers, sensors, and actuators are connected in a wireless sensor network to monitor and control the correct operation of a physical system. Information about the environment of the physical system is collected by sensors and serve as inputs to the controllers. Controllers perform corrective actions

Table 2.1 Cybersecurity attack-vulnerability-damage model [1]

Attack			Vulnerability	Damage		
Origin	Action	Target		State effect	Performance effect	Severity
Local	Probe	Network	Configuration	None	None	None
Remote	Scan	Process	Specification	Availability	Timeliness	Low
	Flood	System	Implementation	Integrity	Precision	Medium
	Authenticate	Data		Confidentiality	Accuracy	High
	Bypass	User				
	Spoof					
	Eavesdrop					
	Misdirect					
	Read/Copy					
	Terminate					
	Execute					
	Modify					
	Delete					

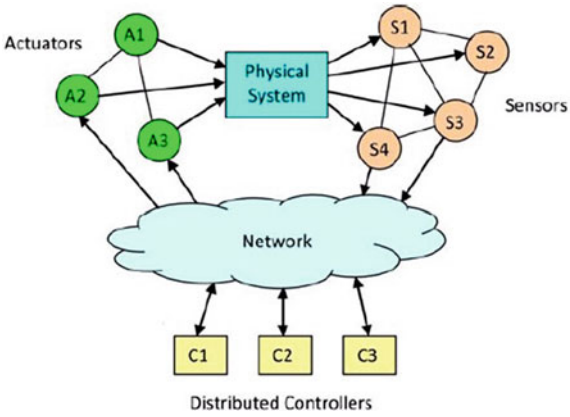

Fig. 2.1 General architecture of a sensor-actuator network [1]

on the physical system through actuators connected to the system based on the values of sensor inputs.

The μRacer protocol suite consists of a trust-aware routing protocol (TARP), context-aware routing protocol (CARP), and a service-aware routing protocol (SARP). The TARP routing protocol aims to prevent packets from being routed through malicious or malfunctioning nodes using two concurrent sub-functions, reputation assessment and path reliability evaluation respectively. The reputation assessment sub-function assesses the direct reputation of a node based on its most recent interaction with another node, and the indirect reputation of the node based on its routing behavior or past interactions with its neighbor nodes. When an

authorized node broadcasts a reputation request message regarding an unknown node, each node concerned broadcasts a reputation report on the unknown node. Upon receiving the reputation report, the requesting node updates the indirect reputation of the unknown node. If the reputation of an unknown node falls below a minimum acceptable threshold, any node in the network that is able to detect this situation broadcasts an unsolicited reputation report. The aggregate reputation of the unknown node is then calculated based on both its direct reputation and indirect reputation.

The k-FTM trust model, proposed by Srnivasan and Wu, is a k-parent flooding tree for secure and reliable broadcasting in wireless sensor networks. Based on the flooding tree model (FTM) for communications, where messages are sent from the root of the tree toward its leaves, this trust model aims to prevent denial-of-broadcast message (DoBM) attacks which are similar to DoS attacks. The wireless network topology of this trust model is well suited to broadcast communications such as the IP multimedia subsystem (IMS) which will serve as the key enabling technology for service convergence in the all-IP core network of the mobile Internet. Except for nodes that are within transmission range of the base station, each node in the tree has exactly k parents. Blind flooding, where each node rebroadcasts a message when it is received for the first time, is carried out once to generate the initial structure of the flooding tree. Each node in the flooding tree retains a reputation value for each of its neighbors, namely, its parent and child nodes. Messages are broadcast in two phases which include the broadcast phase and the acknowledgment phase. Both phases are susceptible to routing attacks, however, which can employ a compromised node to block a message, thereby preventing it from reaching or flooding the entire sub-tree. Thus, both tree height and node degree are important in preventing DoBM attacks.

As usual in any engineering field, these trust models involve tradeoffs. When computing the trust value for an unknown sensor node, μRacer takes into account the indirect reputation of the node as reported by its neighbor nodes in addition to its direct reputation. In contrast, k-FTM imposes an extra topology characteristic to ensure that messages can flow in the tree in case of a malicious or malfunctioning node. Whereas trust values computed by μRacer have the least fluctuation over time, thereby enhancing security, k-FTM is able to determine more quickly that a node is compromised by malicious or malfunctioning behavior, thereby enhancing availability. Thus, while k-FTM is faster at detecting compromised nodes, its trust values for unknown nodes are more uncertain given greater fluctuations due to taking into account only the direct reputation of the node. Conversely, while μRacer is slower at detecting compromised nodes due to the additional computation time required to calculate the indirect reputation of an unknown node, its trust values for the node are more certain. This comparison suggests that μRacer is more suitable for WSNs where routing attacks are of primary concern, whereas k-FTM is more suitable for WSNs that are deployed in harsh environmental conditions where nodes may fail. Both trust models, however, illustrate limitations that motivate the need for a new trust model for wireless next generation networks. Whereas μRacer has a large trust value computation overhead, k-FTM has a large communication

overhead. To achieve a better balance between security and availability, we need a trust model that reduces both the computation overhead in the calculation of trust values for nodes and the communication overhead in the number of messages that need to be exchanged between nodes.

2.2 Key Management for Mobility in Wireless NGNs

Barker et al. [3] argue that cryptography is used to protect information from unauthorized disclosure, to detect unauthorized message modification, and to authenticate the identities of network entities. A network entity can be a human agent, including an individual or an organization, or it can be an artificial agent such as a network device, process, or autonomous rational agent acting on behalf of a human agent. A cryptographic key management system (CKMS) protects keys and metadata from being stolen and decrypted. In addition to managing and protecting cryptographic keys throughout their lifecycles, a CKMS needs to protect certain metadata about the key such as the cryptographic algorithm it uses, the authorized uses of the key, and the security services that are provided by the key.

Since keys can be stolen by an attacker when they are generated by a serving network and distributed to a target network, key management in wireless next generation networks has become an attack vector in its own right. Barker et al. [4] argue that the proper management of cryptographic keys is essential for the effective use of cryptography. Poor key management can easily compromise the strongest cryptographic algorithms. Secure information protected by cryptography depends on the strength of the keys, the effectiveness of mechanisms and protocols that support the keys, and the protection of the keys themselves. All keys need to be protected from modification, and secret and private keys need to be protected from unauthorized disclosure or leakage. Effective key management provides the foundation for secure key generation, distribution, and storage. In the case of secure key generation and distribution, Bergstra and Burgess [5] argue that there is often a high level of uncertainty in knowing the true source of a cryptographic key. Beyond a certain threshold of evidence, one needs to trust the assumption of ownership. The higher the level of uncertainty regarding the source, the more risk or vulnerability one entity in a trust relation needs to accept. This can destabilize the trust relation between two interacting entities, based on the assumption that the risk or vulnerability accepted by both entities should be ideally shared or symmetric.

This problem has motivated the need for new approaches concerning how to securely handle key exchanges in mobility applications. When an entity switches from one point of attachment (PoA) to another in the same network or in a different network with a roaming agreement, the wireless network connection needs to be secure to prevent connecting to a rogue PoA or transmitting confidential data over an unprotected link. Hoeper et al. [6] describe the problem and propose three possible solutions. They begin by describing the requirements for full network

authentication, which involve the establishment of a secure connection between a mobile node and PoA in a wireless network through an authentication and key exchange protocol. The authentication process involves the sharing or exchange of secret keys or passwords between two interacting entities during the protocol execution. At each security level or trust level in the wireless network, new keys are derived from the exchanged keys to protect subsequent communications between the mobile node and PoA.

The term "handover" originates from cellular networks, and refers to the process of changing the current PoA to a target PoA through the use of switches. According to Chen and Gong [7], a communications system is dynamic in that some nodes may move from one location to another as in the ubiquitous case of cellular phones. Such nodes are called mobile nodes which are terminals that can be connected to networks through different fixed network nodes or PoA as shown in Fig. 2.2.

Initially, the mobile node *m* is connected to the network through fixed node 2. As *m* moves closer to fixed node 5 in the network, the mobile node switches its connection from fixed node 2 to fixed node 5. The process of switching from one fixed node to another in the network is called a handover. If executed efficiently, handovers provide a continuous connection to the network through different base stations without disrupting communications. Service mobility refers to the protocols and mechanisms responsible for facilitating collaborative communication between mobile nodes. In order to handover key and service information from one base station to another, both entities in the data exchange must share the same knowledge about a cellular subscriber's roaming. The term "roaming" originates from the global system for mobile (GSM) communications standard used by mobile phones. Roaming agreements between networks help ensure that a traveling wireless device is kept connected to a network without breaking the connection. If a cellular subscriber travels beyond the transmitter range of his or her cell phone company, the cell phone should automatically hop onto another phone company's service, if available, using the subscriber's identity in the visited network. A handover is said

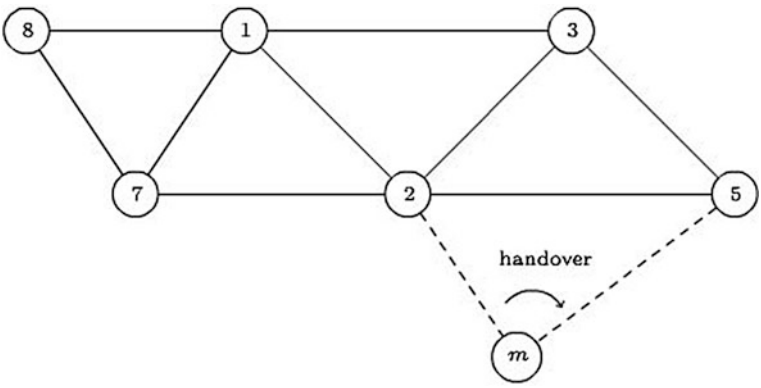

Fig. 2.2 A network with mobile nodes [7]

to be "seamless" when a new connection is established before the old one is broken or torn-down.

Hoeper et al. argue that the problem is how to execute the handover as quickly as possible to minimize security threats both to the network and to the mobile terminal, and to prevent network services from being disrupted while an entity is roaming. A common approach, called pre-authentication, establishes a full network authentication with a target PoA based on the current network connection before the handover is executed. A more efficient approach, called re-authentication, utilizes the information obtained from the establishment of keys in a previous authentication in the same network or in a different network with a roaming agreement to expedite the handover. The solutions proposed by Hoeper et al. are based on a re-authentication approach. Two different handover scenarios are considered, one involving a handover within a single security domain and the other involving a handover between two different security domains with a roaming agreement.

Figure 2.3 shows the architecture for a handover within a single security domain. A single security domain consists of one authentication server (AS) and several key holders connected to the authentication server. A mobile node (MN), which is currently attached to the serving PoA, plans to attach to a target PoA in the same security domain. Whereas the serving PoA is connected to the authentication server through $n - 1$ key holders, the target PoA is connected to the authentication server through a different path of $m - 1$ key holders. $L(x)[S]$-key holders and $L(x)[T]$-key holders refer to network entities in the serving and target networks respectively, which can serve as key holders in the authentication process. The key holder level in the network is represented by x. Lc-key holder is defined as the lowest common

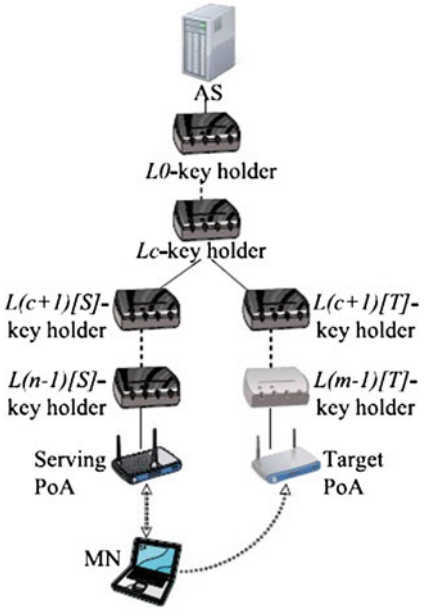

Fig. 2.3 A handover within a single security domain [6]

key holder before the paths from the authentication server to the serving and target PoAs split into two different branches.

Figure 2.4 shows the architecture for a handover between two different security domains that have a roaming agreement. Each security domain has its own authentication server, AS1 and AS2 respectively. In this case, the serving and target networks and their corresponding key holder paths are different. Typically, a key holder in the serving network can communicate with a key holder in the target network only through the authentication servers. A shortcut is defined as a direct communication path between a key holder in the serving network and a key holder in the target network that does not have to go through the authentication servers. Figure 2.4 illustrates a shortcut between the lowest common key holders, $Lc[S]$-key holder and $Lc[T]$-key holder.

In both handover scenarios, the handover assumes that there is a key hierarchy that has been established between a mobile node and the serving PoA in a previous authentication. Figure 2.5 shows the key hierarchy for the serving network S. The key hierarchy depends on the key holder path in the serving network S or the target network T, the wireless access technology i, and the number of key holder levels in the serving or target networks, denoted by n and m respectively. If the network access is successful, the mobile node and the authentication server derive the root key $RK_i[S]$. The authentication server then derives the $L0K_i[S]$ key for level 0 and sends it to the $L0$-key holder. The $L0$-key holder then derives the $L1K_i[S]$ key for level 1 from the $L0K_i[S]$ key and sends it to the $L1$-key holder. The key derivation and distribution process is continued until the lowest key holder in the chain, the PoA in the network, receives the $LnK_i[S]$ key. The PoA then derives the transient protection key $TPK_i[S]$ which is used by the mobile node

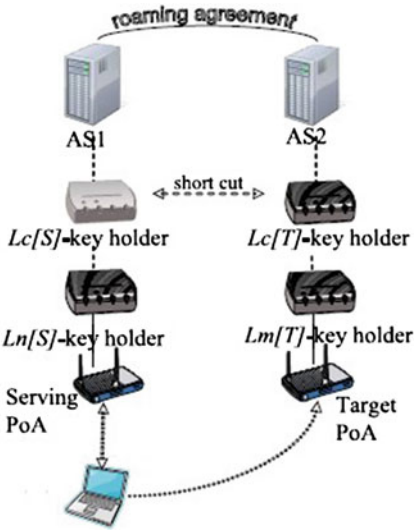

Fig. 2.4 A handover between two different security domains [6]

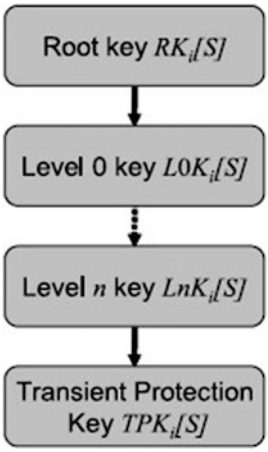

Fig. 2.5 A key hierarchy for a wireless technology i [6]

to derive all the keys in the hierarchy necessary to protect the wireless link between itself and the PoA.

The security and availability issues associated with seamless handovers in wireless next generation networks arise from two unresolved problems. First, unlike cellular networks, other types of wireless networks lack a dedicated key management infrastructure to support the derivation of handover key hierarchies from previous network connections and their distribution to a target network. Consequently, no dedicated network entities are available to trigger and manage key distribution. Roaming information is necessary to trigger key distribution to the correct target network. But in some wireless networks entities are unable to exchange information about the roaming behavior of mobile nodes. Thus, it is unclear how key distribution can be triggered. Once it is triggered, a network entity needs to derive handover keys and distribute them to a target network. Hoeper et al. point out that this network entity could be the serving authentication server, a common key holder, or an entity with shortcut access to the target network. There are other related problems that any suitable handover scheme needs to address, including the fact that not all wireless networks share the same trust model. Two entities operating in different networks or security domains at the same key holder level may not have the same physical protection. Moreover, if the serving and target networks have a different number of key holder levels, a given key holder level in one security domain may not correspond to the key holder level in another security domain.

Secondly, the key distribution protocol needs to be able to perform handover functions in a timely fashion to prevent security threats and avoid network disconnections. As we saw in the comparison of two secure routing protocols, μRacer and k-FTM, this means that these handover functions should not have a large computation and communication overhead. At one extreme, if the network entity that performs the handover functions is the serving authentication server, the handover will be more secure but slower. At the other extreme, if the network entity

that performs the handover functions is an entity with a shortcut access to the target network, the handover will be faster but less secure. Many security features can be provided only through an authentication server such as network-wide key synchronization, homogeneous trust models, channel binding, and the prevention of replay attacks using sequence numbers or timestamps. Such security features increase the computation and communication overhead of the key distribution protocol which results in delays that could adversely affect availability.

Thus, as in the case of our analysis of secure routing protocols, there is a tradeoff between security and availability in handovers, depending on the key holder level at which the handover is executed. In the light of this tradeoff, Hoeper et al. suggest the following strategy. In the best case, where some security requirements can be limited or even suspended, a fast handover can be performed by using the lowest common key holder or a shortcut. Once the handover is complete, a full network authentication can be forced by limiting the key lifetime. This releases the serving network from liability and assures the target network that the new connection is secure. In the worst case, if no security requirements can be suspended or limited for any duration, both the serving and target authentication servers in the backbone need to be accessed during the handover. In this case, a timely initiation of the re-authentication protocol is required to minimize performance degradation.

2.3 Current Approaches to Seamless Handovers

Although mobile communication systems were originally developed for military applications, Forsberg et al. [8] discuss how the concept of a cellular network was extended to commercial applications in the early 1980s. The advanced mobile phone system (AMPS) and the Nordic mobile telephone system (NMT) were the first cellular networks developed in the U.S. and northern Europe respectively. These first generation systems utilized analog transmission techniques and frequency division multiple access (FDMA). These systems supported handovers between different cells in a network such as a phone call from a car. Second generation (2G) mobile systems appeared in the early 1990s, and were predominantly based on the global system for mobile (GSM) communications standard. These second generation systems utilized digital transmission techniques over the radio interface between the mobile phone and the base station and time division multiple access (TDMA). 2G systems provided increased network capacity due to the efficient use of radio resources, and supported improved audio quality and new types of security features due to digital coding techniques. Third generation (3G) mobile systems appeared in the early 2000s, and introduced the concept of fully mobile roaming. This concept makes it possible for users to access mobile services from anywhere in the world through a collaboration of standards bodies from Europe, Asia, and North America called the 3G partnership project (3GPP). 3G systems provided large increases in data rates up to 2 Mbps using wideband code division multiple access (WCDMA). GSM systems and 3G systems are divided into

two different domains, the circuit switched (CS) domain for carrying voice and short messages and the packet switched (PS) domain for carrying data.

Long-term evolution (LTE) of radio technologies, together with system architecture evolution (SAE), were initiated by 3GPP a decade later as the next evolutionary step in mobile communication systems. The new system is called evolved packet system (EPS), and its most important component is the radio network called evolved universal terrestrial radio access network (E-UTRAN). The EPS system contains only a PS domain and provides large increases in data rates up to more than 100 Mbps by using orthogonal frequency division multiple access (OFDMA) for downlink traffic from the network to the mobile terminal and single carrier frequency division multiple access (SC-FDMA) for uplink traffic from the mobile terminal to the network. LTE is based on a radio interface specification standardized by 3GPP. The ongoing LTE standards development is progressing toward an enhanced LTE radio interface called LTE-advanced (LTE-A). The new radio interface is motivated by the need for higher communications system capability in the light of the growth of mobile data traffic due to the proliferation of smartphones and new mobile devices. Since the spectrum in lower frequency bands of the original LTE radio interface is becoming scarce, the new LTE-A radio interface requires the efficient utilization of higher frequency bands to sustain future growth and support further network densification.

Kishiyama et al. [9] argue that this requirement will involve the integration of wide and local area enhancements through multicell cooperation between macrocells and small cells. Figure 2.6 shows the future development of LTE. In LTE-A, standardized technologies for multicell cooperation include coordinated multipoint (CoMP) transmission/reception and enhanced intercell interference coordination (eICIC). To enable the upgrade from the original LTE radio interface to the new LTE-A radio interface, enhanced local area (eLA) specifications and technologies for the mobile

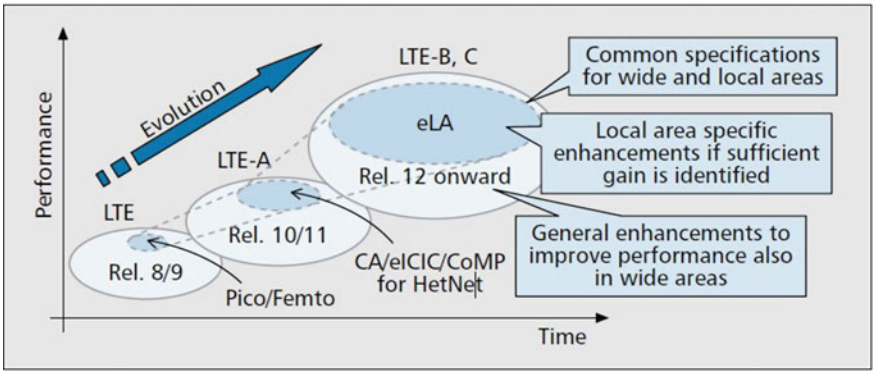

Fig. 2.6 Future development of LTE [9]

communications system need to be developed and added, while retaining common specifications for the LTE radio interface between wide and local areas.

Techniques are still needed to reduce the potential impact of the increased volume of signaling traffic on both the mobile network side and the mobile handheld device side. As network density is increased through the deployment of more small cells, these techniques need to optimize spectrum utilization which ranges from lower and narrower frequency bands to higher and wider frequency bands. Higher frequency bands, however, cannot be optimally utilized in wide areas of deployed macrocells due to space limitations on the evolved node B (eNB) side, which serves as a base station for the LTE radio access communications system. In addition, not only are higher frequency bands subject to higher path loss, but the cost of altering the existing network infrastructure to support higher frequency bands is prohibitive. Thus, the approach recommended by most researchers is to use lower frequency bands to provide basic coverage and mobility for macrocells in wide areas, and to use separate higher frequency bands to support small cells and high-speed data transmission in local areas. In contrast to wide areas consisting of macrocells, local areas refer to outdoor dense deployments and hotspots of small cells. Figure 2.7 shows how wide areas can be supported by lower and narrower frequency bands, while local areas can be supported by separate higher and wider frequency bands.

Kishiyama et al. argue that the integration of wide and local areas can be seen as a new form of cooperation between conventional macrocells in lower frequency bands and small cells in higher frequency bands. To support the integration, the authors introduce the novel concept of a phantom cell as a macro-assisted small cell that can extend the spectrum into higher frequency bands. Multicell cooperation based on macrocell assistance of small cells aims to make efficient use of higher and wider frequency bands for small cells. In the phantom cell solution, the protocols for the control (C)-plane and the user (U)-plane are separated. Whereas the C-plane

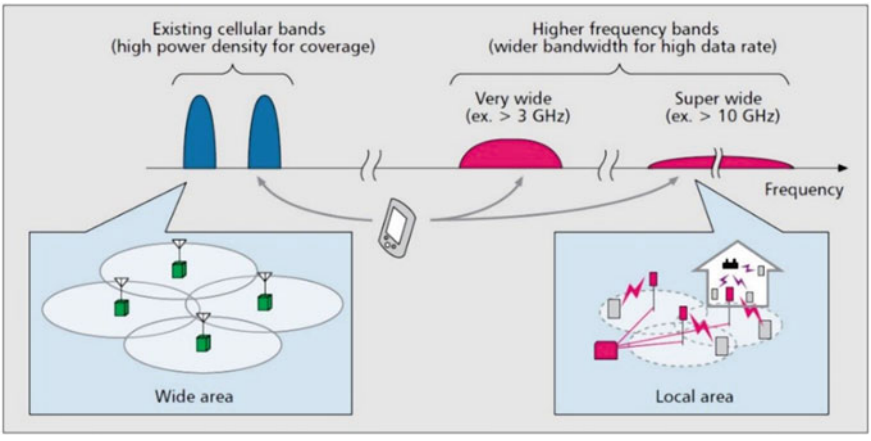

Fig. 2.7 Wide area and local area bandwidth utilization [9]

is used to transmit control signals, the U-plane is used to transmit user data. In the case of wide area deployments of macrocells, both the C-plane and the U-plane for a mobile terminal are provided by the serving macrocell as in conventional mobile communication systems. In the case of local area deployments of small cells, however, the C-plane for a mobile terminal is provided by a macrocell in a lower frequency band, while the U-plane for the mobile terminal is provided by a small cell in a higher frequency band. For this reason, macro-assisted small cells are called phantom cells because they transmit only UE-specific signals. This scheme reduces the amount of control signaling or communication overhead involved in frequent handovers between small cells on one hand, and between small cells and macrocells on the other. Thus, using small cells in higher frequency bands helps to ensure network connectivity and availability.

3G mobile networks currently provide protection for data confidentiality, authentication, C-plane and U-plane confidentiality, and C-plane integrity. The LTE next generation mobile communications system will require additional security functions, including a key hierarchy, separate security functions for access stratum (AS) and non-access stratum (NAS), and expanded forward security functions to protect handovers. The AS and NAS refer to different functional layers in the universal mobile telecommunications system (UMTS) protocol stack. Whereas the NAS specifies communication between a mobile terminal and a node in the core network, the AS specifies communication between a mobile terminal and an eNB node or base station at the network edge for the LTE radio access communications system.

Zugenmaier and Aono [10] have proposed a handover scheme that utilizes forward security to limit the scope of damage when a compromised key is used. The handover scheme is based on the key hierarchy shown in Fig. 2.8. Two keys are generated by the core network and mobile terminal (UE) during the execution of the authentication and key agreement (AKA) mechanism for mutual authentication. This includes the *CK* key for encryption and the *IK* key for integrity protection. These keys are passed from the universal subscriber identity module (USIM) to the mobile terminal, and from the authentication center (AuC) to the home subscriber server (HSS). The HSS and mobile terminal use both keys to generate the K_{ASME} key using a key generator function based on the ID of the visited mobile network to ensure that the key can be used only by that network. This key is passed from the HSS to the mobility management entity (MME) of the visited mobile network to serve as the root key in the key hierarchy.

Two additional keys are generated from the root key, the K_{NASenc} key for NAS protocol encryption between the MME and the mobile terminal, and the K_{NASint} key for NAS integrity protection. Once the mobile terminal is connected to the visited mobile network, the MME generates the K_{eNB} key and passes it to the eNB node or base station at the network edge for the LTE radio access communications system. Finally, three more keys are generated from this key, the K_{UPenc} key for U-plane encryption, the K_{RRCenc} key for radio resource control (RRC) encryption, and the K_{RRCint} key for RRC integrity protection.

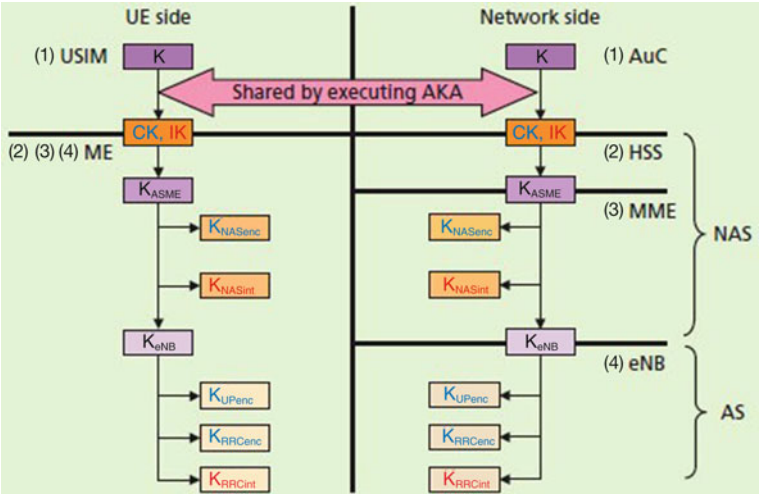

Fig. 2.8 Key hierarchy and generation in LTE [10]

The concept of forward security was introduced to provide protection from unauthorized network access in the event that an eNB node at the network edge is subverted. If an attacker gains access to an encrypted key through a compromised eNB node, forward security leverages computational complexity to ensure that the key cannot be decrypted. Figure 2.9 shows two different types of key delivery or distribution mechanisms, including horizontal and vertical handovers. When a mobile terminal is connected to an eNB node at the edge of a visited network through a shared initial AS security context, the mobility management entity

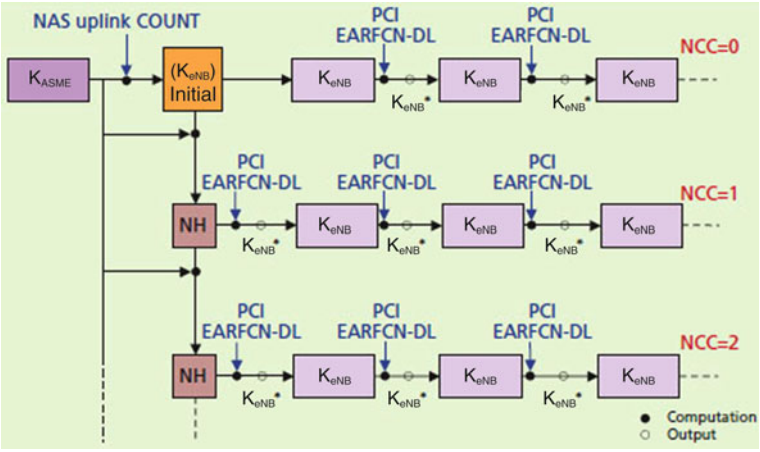

Fig. 2.9 Horizontal and vertical handovers [10]

(MME) of the visited network generates the K_{eNB} key. In addition, the MME and the mobile terminal generate another key used for forward security called the next-hop (*NH*) parameter. The initial value of the K_{eNB} key is generated from the K_{ASME} root key in the key hierarchy with additional input from the NAS uplink COUNT. The initial value of the *NH* key is generated from the K_{ASME} root key and the initial value of the K_{eNB} key. In the case of horizontal handovers, a new K_{eNB}* key is generated from the current K_{eNB} key with additional inputs from the connection's E-UTRAN absolute radio frequency channel number-down link (EARFCN-DL) and its target physical cell identity (PCI). In the case of vertical handovers, the value of the *NH* key is changed when the *NH* chaining counter (NCC) is incremented. Thus, for each value greater than NCC = 0, a new K_{eNB}* key is generated from the current value of the *NH* key with additional inputs from the connection's EARFCN-DL and its target PCI.

In both cases, the new K_{eNB}* key serves as the base key for securing communication between the mobile terminal and the eNB node at the edge of the visited network. Since the *NH* key can be generated only by the MME and the mobile terminal, the *NH* keys provide a method for implementing forward security in handovers across multiple eNB nodes. Thus, at the time of vertical key delivery, the next-hop forward security mechanism ensures that the future K_{eNB} key used to connect the mobile terminal to another eNB node after n or more handovers (where $n = 1$ or 2) cannot be guessed due to computational complexity. Even if the current K_{eNB} key is leaked, threats to the network are limited because future keys are generated without using the current K_{eNB} key. Unlike horizontal key distribution, however, vertical key distribution has a large computation overhead which can delay handovers. Thus, like other security mechanisms we have discussed, forward security involves a tradeoff between enhanced security and availability.

Our discussion of trust models for secure routing in wireless sensor networks and key management schemes for secure key derivation and distribution in non-cellular wireless and mobile networks underscores the following point. Traditional security mechanisms with large computation and communication overhead are not feasible for wireless next generation networks. In the case of wireless NGNs, secure routing, key derivation and distribution, and data aggregation and storage need to be implemented using distributed schemes and collaborative nodes rather than centralized mechanisms. We have seen two examples of distributed schemes for handovers in mobile networks that require some form of cooperation and collaboration between nodes or cells. In one case, we have seen how LTE-A might employ phantom cells to reduce the communication overhead in handovers. In another case, we have seen how handovers in LTE-A might be made more secure through forward security, but at the cost of both a large computation and communication overhead.

New paradigms for distributed cooperation and collaboration and information diffusion will require a new trust model for wireless NGNs. Di Pietro and Guarino [11] have proposed a novel scheme for data distribution in mobile unattended WSNs that might be applicable to key management schemes in LTE-A. The scheme is motivated by epidemic theory. Analogous to an infected person, a node in a

network can leak data if it has been infected by a malicious entity or compromised by malfunctioning behavior. Epidemic studies on how healing and infection rates affect the probability that a disease will become extinct or spread can be used to determine the replication rate necessary to prevent data leakage. Although replication and distributed dissemination provide the most straightforward approach to reliable data storage in WSNs, this approach compromises confidentiality and source location privacy. Similarly, although key derivation and distribution provide the most straightforward approach to handovers, this approach compromises the same cybersecurity principles since keys can be intercepted in the handover process and subsequently decrypted. Instead, Di Pietro and Guarino argue that local information sharing schemes are the best solution to security in a mobile context.

To implement a local information sharing scheme, a WSN is used to monitor a specified area with a variable number of sensors depending on the requirements of the security domain. After acquiring data, sensor nodes collaborate to securely route the data to one or more sinks for storage. The network topology helps to ensure availability by reducing data leakage due to the malicious capture or failure of a sensor node. The network topology helps to ensure confidentiality by preventing exposure of stored data such as secret keys to unauthorized entities. Instead of protecting individual pieces of information sensed by each node in the network, the network topology protects the information sensed by the network as a whole. By making sure that the pieces of information stored by the sensor nodes in the network are uncorrelated with their location, the quantity and content of information that can be recovered or inferred from a given fraction of the network depends on the size of the fraction rather than on the identity of the nodes. To limit the bandwidth requirements and energy consumption of sensor nodes, the sensed data needs to be locally transmitted. Here mobility, which is often seen as a liability in a security context, is exploited as an asset to spatially diffuse or spread the sensed data throughout the monitored area of the network in an energy efficient fashion.

This can be accomplished by establishing a local information sharing scheme over the monitored area, where one share of the secret information is randomly distributed to each of k neighbor nodes of the distributing node. The property of mobility is then leveraged to randomly move these nodes around until the secret information is spatially spread out over the localized area. The shares of secret information are diffused over a localized area through the mobility of sensor movements. Subsequently, a sink can be used to explore the monitored area to retrieve all the shares of information stored by the nodes within the local communication range of the sink. Such a local information sharing scheme provides availability, which is the main benefit of data replication, without the cost of potential data leakage and loss of confidentiality. The primary goal of confidentiality is enforced by requiring access to at least k nodes in the network to recover a single piece of secret information. Information diffusion effectively renders the content of the data stored in a node independent of its location, thereby providing both source and location confidentiality. The remaining chapters will develop a virtue-based trust model that can be used to support the cooperation and collaboration required between network entities in local information sharing schemes.

References

1. Kisner RA et al (2010) Cybersecurity through real-time distributed control systems. Oak Ridge National Laboratory, ORNL/TM-2010/30, pp 1–14
2. Vasilache RA et al (2013) Comparative evaluation of trust mechanisms for wireless sensor networks. In: 11th RoEduNet international conference, Sinaia, Romania, 17–19 Jan 2013, pp 1–5
3. Barker E et al (2013) A framework for designing cryptographic key management systems. NIST, Special Publication 800–130. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-130.pdf. Accessed 15 Jul 2014
4. Barker E et al (2012) Recommendations for key management. NIST, Special Publication 800–57, Part 1, rev 3. http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf. Accessed 15 Jul 2014
5. Bergstra J, Burgess M (2009) Local and global trust based on the concept of promises. Computing Research Repository (CORR), paper 0912.4637. http://arxiv.org/pdf/0912.4637.pdf. Accessed 15 Jul 2014
6. Hoeper K et al (2008) Security challenges in seamless mobility–how to 'handover' the keys? In: 4th international ACM wireless internet conference (WICON '08), Maui, HI, 17–19 Nov 2008, pp 2–3
7. Chen L, Gong G (2012) Wireless security: security for mobility. In: Communication system security. CRC Press, New York
8. Forsberg D et al (2013) LTE security, Chap 2, 2nd edn. Wiley, Hoboken
9. Kishiyama Y et al (2013) Future steps of LTE-A: evolution toward integration of local area and wide area systems. IEEE Wirel Commun 20(1):12–18. doi:10.1109/MWC.2013.6472194
10. Zugenmaier A, Aono H (2013) Security technology for SAE/LTE (system architecture evolution 2/LTE). NTT DOCOMO Tech J 11(3):28–30. http://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/rd/technical_journal/bn/vol11_3/vol11_3_027en.pdf. Accessed 15 Jul 2014
11. Di Pietro R, Guarino S (2013) Confidentiality and availability issues in mobile unattended wireless sensor networks. In: 14th IEEE international symposium and workshops on a world of wireless, mobile and multimedia networks (WoWMoM), Madrid, Spain, 4–7 Jun 2013, pp 1–6

http://www.springer.com/978-3-319-11902-1

Wireless Next Generation Networks
A Virtue-Based Trust Model
Harvey, M.G.
2014, XIX, 117 p. 17 illus., Softcover
ISBN: 978-3-319-11902-1