4

Vulnerabilities in the Telephony

The physical and logical isolation of cellular networks from other systems has long provided many of the advantages of so called “air-gap” security. Even still, these networks have historically exhibited significant security vulnerabilities. In this chapter, we explore many of the most critical of such weaknesses. Our examination ranges from the collapse of proprietary cryptographic algorithms and the lack of authenticated messages between elements in the network core to a growing pool of smart phone-based malware. In combination with the attacks discussed in the coming chapters, the information in this chapter can help the reader understand the current state of security in cellular networks.

4.1 Weak Cryptographic Algorithms

For much of the history of cryptography, ensuring that a cipher remained secret was an integral part of protecting the confidentiality of data. However, the passage of time has repeatedly demonstrated that secret algorithms as simple as the Caesar Cipher to the complex Content Scramble System (CSS) used to protect DVDs frequently contain fatal weaknesses. Starting around the turn of the twentieth century, however, a movement toward the creation of strong, publicly vetted algorithms arose. Spurred by Kerckhoff’s principle, which states that strong algorithms must assume that adversaries possess complete knowledge of the system, the majority of modern cryptography is instead based around the idea that secrecy lies in the key and not the algorithm.

While history provides us with countless examples of proprietary cryptographic algorithms failing once examined by adversaries, cellular providers and equipment manufacturers have long resisted analysis of the algorithms used to provide authentication and confidentiality in their networks. We offer a brief overview of how the most critical current and next-generation algorithms have been broken through cryptanalysis.

From the perspective of the network, the most important function of cryptography is the ability to vet users as they attempt to login to a network. Unique identification allows a provider to associate the use of specific services with a specific account, thereby providing a means of accurate, unforgeable billing. In GSM networks, this functionality is embodied generically as A3/A8, which is specifically implemented by variants of the COMP128 algorithm. Specifically, the A3 algorithm is tasked with authenticating a user with the network using a 128-bit key K_i (See Chapter 3). The A8 algorithm creates a 64-bit session key K_c, which consists of the last 54-bits generated by the COMP128 algorithm and 10 zeros appended to the end, for communication with the network. While the use of COMP128 for authentication was believed to be strong (given the key length), a number of attacks have been demonstrated by the academic community. For instance, Goldberg et al [75] noted that the secret key K_i could be recovered by querying a GSM SIM card approximately 2^{19} times, which can take between six and eight hours to accomplish. Soon after this work was published, the code for COMP128 was leaked and published on multiple sites on the Internet. Rao et al [144] used this information later to dramatically improve the performance of attacks against COMP128, creating a method through which K_i could be recovered in under one minute. Accordingly, the COMP128 is widely viewed, even by the industry, as a vulnerable algorithm.

A number of mitigations have been proposed to protect against the weakness of COMP128. For example, many smart card developers limited the functional lifetimes of their SIM cards to under 2^{16} operations. While this mechanism certainly prevents the original attack by Goldberg et al, it does not prevent more modern attacks such as the work by Rao et al. More critically, it allows malicious code to potentially deny a legitimate user the ability to use their phone (See Chapter 5). Most networks, however, now rely on modified versions of the COMP128 algorithm, known as COMP128-2 and COMP128-3. While the GSM trade organization, GSM World, actively supports the adoption of one of these algorithms [81], it is unclear how many networks have fully upgraded beyond the original COMP128 algorithm. Specifically, full substitution of A3/A8 algorithms requires that all SIM devices are replaced. Because the details of these new algorithms remain private[1], little is known of their strength.

From the perspective of the user, the most important security feature provided by the network is the confidentiality of conversations. Unlike the Internet, the vast majority of people believe that their communications via cellular networks can not be intercepted by third parties. The network, in fact, does provide encryption for voice calls between the user's phone and the base station. In GSM networks, this algorithm is known generically as A5 and comes in a number of variants. A5/1, for instance, is used through-

[1] Publicly, it is claimed the COMP128-2 and COMP128-3 fix the problems of COMP128, with COMP128-3 addressing the appended zeros in K_c.

out the United States and Europe. Due to the political environment at the time of their development, the purposefully weakened A5/2 algorithm was developed for export to other parts of the world. Like A3/A8 before them, however, both such algorithms were found to be critically flawed when examined by external cryptographers. Golic [77] published the first cryptanalysis of the algorithm believed to be A5/1 and discovered that the key could be recovered in approximately 2^{40} operations. Biryukov et al [43] developed the first such attack in 2000, in which they demonstrated the ability to recover an A5/1 key (K_c) by processing the first two minutes of a conversation using a standard desktop PC with 128MB or RAM and two 73 GB hard drives. This attack was able to successfully able to recover the key in under one second. Biham and Dunkelman successfully reduced the complexity of the attack and decreased the storage to execute it to 38 GB [41]. Barkan et al [148] then developed an attack requiring no preprocessing and only a short sample of known text. Petrovic et al [138] instead examined the weaker A5/2 algorithm and demonstrated that only approximately 2^{16} operations, requiring no more than 10 milliseconds, were necessary to recover the key. Barkhan et al [38] later proposed a cipher-text only attack against A5/2 requiring no more than a few milliseconds to recover the key. Most recently, Hulton and Miller used FPGAs and rainbow tables to demonstrate the ability to passively recover an A5/1 key in approximately 30 seconds [111].

The alleged evolution of both encryption algorithms is also interesting. Ross Anderson, for instance, has argued that insider information revealed that politics ultimately decided the cipher used to secure voice communications in GSM systems [30]. During a meeting of the NATO signals agencies during the mid 1980's, nations argued over whether or not GSM should be protected by a strong algorithm. Countries bordering Warsaw Pact nations, especially Germany, argued that strong encryption should be used to protect communications against eavesdropping. However, the majority of other nations disagreed with this assertion and ultimately settled on a less-strong, French designed algorithm. Whether or not the details are true, the story reflects the atmosphere of fear of the widespread use of strong cryptography in the late part of the twentieth century.

Next generation cellular networks have made great leaps forward in their application of modern, publicly vetted cryptography. The descendant of GSM networks, or so-called WCDMA networks[2], use the new A5/3 algorithm. Also known as Kasumi, A5/3 is significantly stronger than its predecessors. However, research by Biham et al [42] demonstrated that the key can theoretically be recovered using $2^{54.6}$ chosen plaintexts (with a time complexity of $2^{76.1}$ Kasumi operations). Accordingly, such an attack is currently beyond practical implementation. The work by Barkhan et al [38] demonstrated that weaknesses in the protocol used by GSM networks and not Kasumi itself allow for the recovery of the key. Networks based on the CDMA2000 standard in-

[2] See Chapter 3 for details on the evolution of these networks.

stead now apply a derivative of AES, for which no weaknesses have yet been reported.

4.2 Vulnerabilities in the Network Core

Logical and physical isolation have been especially important to the security of the core of telecommunications networks. However, as the capabilities of and access available to the general public have expanded, so too has the ability of adversaries to gain unfettered access to critical services. In this section, we investigate some of the better known vulnerabilities in traditional and cellular telecommunications networks. These models ultimately point to a changing reality - that the security measures taken by such systems must change to reflect the reality of threats to the system.

One of the best known historical examples of a telecommunications vulnerability came as the result of the end of an era in long distance communication. Beginning in the 1940's, the AT&T network began offering many of its customers so called "direct distance dialing", or the ability to place long distance calls without the assistance of an operator. Because the technology allowing operator-free local calls was not capable of supporting long distance calling, AT&T developed a new series of protocols for carrying signaling information over the existing infrastructure. Specifically, by transmitting audio signals over the same line as voice calls, the network was able to accurately route and bill calls without human intervention. The new system spread quickly and, by the late 1950's, was deployed nearly universally across the United States. The ability to manipulate such in-band signaling from the end-points of the network was soon noticed. Nicknamed "phone phreaks", hackers with access to signal generator equipment were able to enable free long-distance calling. For example, by briefly emitting a pitch at 2600 hertz, which is the tone used by the network itself to indicated that a trunk line is not in use, an adversary could cause the network to stop billing for a call while maintaining the connection end-to-end. As the tools evolved from a cereal box toy (i.e., the Cap'n Crunch "Bosun Whistle") to the more sophisticated "Blue Box" [156], providers recognized that such an exploit could only successfully be prevented by removing in-band signaling from the network. While the vast majority of such systems were replaced digital switching systems by the mid 1980's, the vulnerabilities remained in related products. For instance, Sherr et al [165] noted in 2005 that even modern legal wire-tapping equipment could be disabled through a similar in-band audio attack. Such vulnerabilities speak directly to the evolving state of the perimeter of telecommunications networks. Specifically, the increasing availability of advanced technology to end users and vastly expanded interconnection between provider networks invalidate the many of the assumptions of an entirely controlled environment.

Increased access and interconnection with external networks creates opportunities for a much wider array of attacks than billing fraud by an individual.

The belief that modern networks retained their formerly isolated status led the designers of the current Signaling System Number 7 (SS7) network (see Chapter 3 for an in-depth overview) to design signaling protocols without any real system of authentication. Because nodes trust that all values contained in packets exchanged between network nodes, from source address to request type, an adversary with even minimal access to the SS7 infrastructure can cause significant damage to telecommunications networks nation- and even world-wide. For example, a lack of origin authentication allows messages indicating a link failure to be injected into a system, thereby forcing the network to route around (or toward) a target. Accordingly, an adversary with access to an ISDN connection or a cable can easily cause portions of the network to become unreachable or reroute all traffic toward a single victim node [117, 124]. Moreover, the cryptographic algorithms discussed in the previous section offer no protection of confidentiality in the core. Because traffic is only encrypted between a user's phone and a base station, anyone able to gain access to a cable or switch can easily eavesdrop. Attempts to mitigate such problems including MAPsec [24], which is discussed in Chapter 3, have largely been unsuccessful.

Like much of the software used on the Internet, weaknesses in the implementation of protocols in the network core have also been discovered. One of the more significant such vulnerabilities was discovered in parsers of the *Abstract Syntax Notation* (ASN.1) language. ASN.1 is used widely throughout the core of SS7 networks for tasks including call routing. However, many implementations of parsing functions fail to do sufficient input checking, thereby making a number of buffer overflow vulnerabilities exploitable. Work by researchers at Oulu University revealed that nearly every device running the SNMP protocol, which uses ASN.1 to communicate, was vulnerable to compromise. Such vulnerabilities were not simply due to the code itself, but also the compiler used to generate the parsers themselves. Many within the industry used the occasion to note that such development should leverage the knowledge of the security community [164]; however, the vast majority of systems used by telecommunications networks remain opaque to independent security researchers.

Telecommunications networks simply do not have the isolation upon which their security once relied. As the Telecommunications Act of 1996 requires [173, 161], an individual or group can connect to the SS7 infrastructure simply by paying a relatively small fee.[3] Accordingly, the security of such networks must evolve to address the changing nature of user access and behavior.

The damage possible by such attacks has already been observed in such networks, except that the source of the harm was not malicious. In 1990, AT&T updated the software in all of its 4ESS "long haul" switches. When the software on a single switch noticed that it could not correct a fault, it alerted its neighbors that it was going out of commission for four to six seconds

[3] In 1999, this fee was only $10,000.

to fix the problem [133]. The first communication from the problem switch to its neighbors after such a message indicated that the switch was able to process messages again. However, the new version of the software caused the neighboring switches to go into their fault mode when they received the next signal from the now functional switch. This behavior caused a cascade effect of outage in all of the switches in the network, thereby shutting down the entire long distance network in the country. Because AT&T was aware that the software caused the problem, they were able to restore service within a few hours by reinstalling an older version of the switching software. However, had an adversary instead exploited a vulnerability, similar problems in the future may not be so easily fixed.

4.3 Wireless Eavesdropping

Protecting the content of conversations sent over the air has not always been possible. Limitations in the capabilities of user devices and tight government controls on cryptography left first generation analog cellular systems such as AMPS (see Chapter 3) vulnerable to eavesdropping. Second generation cellular networks, such as GSM, were able to provide protection against direct interception by all but the most sophisticated adversaries.[4] However, poor protocol design in some modern networks can render the cryptographic protections offered by the network useless.

One of the great lessons in underestimating the capabilities of an adversary comes from the wireless portion of GSM networks. Users attempting to authenticate themselves to the system use the combination of A3 and A8 algorithms to generate the correct response and session key based on a challenge transmitted by the base station. However, the base station and therefore the network itself do not authenticate themselves to the user, allowing anyone with equipment capable of interacting with a phone the ability to pretend to be the network. Worse still, because encryption is only enabled in a negotiation after the user correctly responds to the challenge, the false base station can force the user to use the A5/0, or null, encryption algorithm. By acting as a man-in-the-middle, the presence of a false base station allows the conversation of any user to potentially be illegally monitored. While most providers initially downplayed this attack given the expense of base station equipment, an adversary could easily assemble the necessary equipment for under $10,000 [159]. While this issue has been addressed in third generation cellular networks, it should serve as a reminder that communications over the majority of current networks are in fact vulnerable.

Breaking well vetted cryptographic algorithms may not be required to gain access to conversations in reality. A number of networks can in fact

[4] See Section 4.1 for an examination of the weaknesses in the cryptographic algorithms used in GSM.

operate with encryption turned off entirely. For example, the A5/0 cipher in GSM networks alerts the user during connection setup that no encryption at all will be applied. Unlike the above false base station attack, the network itself may allow conversations to be exposed. Countries that legislate strict controls on cryptography, such as France, explicitly require the use of A5/0 in all GSM networks within their borders [56]. It is unknown how widespread such practices are, but rumors of the disabling encryption during periods of elevated traffic have been previously suggested. While a number of end-to-end encryption mechanisms for phones are available [202, 39], such solutions are generally limited to VoIP devices and not currently readily available to most consumers.

4.4 Jamming

When moving from a wired to a wireless transmission medium, the attack surface of a network inherently expands. In particular, the ability of an attacker to interfere with or "jam" communications increases significantly simply by being within the transmission range of a base station. Most of the jamming technology operates by making the network's control channels, for either uplink, dowlink or both directions, unusable by phones in an area. Because the frequency of the control channel is always known to the adversary (unlike the traffic channel used in CDMA communications - see Chapter 3 for more details), all cellular networks are susceptible to attacks by an adversary with the correct radio.

Jamming technology varies significantly in its intended uses and effective range. For example, many "personal" jamming units can deny service to all users within 10 meters of the device [139, 176]. Devices used in military or counter-terrorism operations, such as the Tactical Response Jammer (TRJ) Series [32], can have effective jamming radius of up to five miles. Units with ranges somewhere between these two extremes are being used in some parts of the world to stifle calling in movie theaters, subways and other enclosed spaces.

The legality of such devices, of course, varies wildly. In the United States, for instance, the "manufacture, importation, sale or offer for sale, including advertising, of devices designed to block or jam wireless transmissions is prohibited" [69], as specified by the amended Telecommunications Act of 1934. Because providers are required to pay for spectrum, violation of this law is considered property theft and is punishable by fines of as much as $11,000 for the first offense and up to a year in prison for additional infractions. In spite of these restrictions, such devices are available and popular throughout the country [151]. France, alternatively, permits blocking in many of the enclosed spaces mentioned above [83]. Accordingly, jamming represents a real threat in most environments.

4.5 User Tracking and Privacy

In many circumstances, an adversary may find their target's location information more valuable than the content of their conversations. By tracking an individual's location, an adversary may be able to inflict any number of physical attacks. For instance, developing knowledge of a user's usual behaviors may allow their actions to be predicted and privacy violated. From a technological perspective, tracking a user to their bank may allow targeted attacks such as "spear phishing" to gather more effective information. Accordingly, protecting a user's identity from eavesdroppers is an important consideration for cellular networks.

Early cellular networks such as AMPS provided no protection against user tracking. To verify their identity to the network, users in the AMPS system transmitted a unique 32-bit identifier known as the *Electronic Serial Number* (ESN). However, because of the limited bandwidth of this analog system and the lack of cryptographic mechanisms on user devices, this number was broadcast in the clear. In addition to tracking users, this allowed adversaries with listening equipment to intercept ESNs and make fraudulent phone calls. This attack is said to have cost providers at least a million dollars a day in the 1990's[5] [143].

Modern digital cellular networks do not have quite the same privacy concerns. Increased device capability, in combination with increased bandwidth and the use of cryptography, make tracking individual users nearly impossible. To augment these protections further, networks such as GSM even create temporary identifiers for use over the air such that even unencrypted transmissions leak nothing about a user's identity. The growing move toward cellular phones becoming general purpose computing devices, however, presents new risks to privacy and security. Viruses and malware, as will be discussed in greater detail in Section 4.7, may purposefully leak such information to eavesdroppers. Accordingly, while the problem of user tracking has evolved, it has not evaporated from cellular networks.

4.6 Overload

Like the Internet, telecommunications are susceptible to overload conditions. Such conditions occur for a very simple reason - resources for networks simply are not allocated for worst-case scenarios in which all users want to use the service at the same time. Instead, providers typically allocate enough resources to operate within a reasonable tolerance expected load. For instance, a provider may allocate enough spectrum to handle normal rush-hour volumes plus enough extra resources to handle usage spikes as large as 20%. The

[5] As a note, the author's father was a victim of such an attack in New York City, and was made aware of it when the United States Secret Service contacted him in regards to phone calls made to Iraq during the first Gulf War.

cost of extra equipment and spectrum coupled with the somewhat infrequent occurrence of overload make under-utilized resources simply too expensive to maintain.

Tighter integration of these networks with the larger Internet and an expanding set of services are rapidly redefining "typical" traffic in cellular networks. Outside of the work presented in this book, a number of others have also noticed the increasing potential of this phenomenon. Soon after the attacks of September 11th, for instance, the National Communications System (NCS) issued a report on the potential of using SMS during an emergency [131]. While SMS performed admirably during the attacks in 2001, the service was largely unused by most cellular subscribers. Predicting an upswing, however, this report noted that the current infrastructure would have to be expanded by nearly 100 times to reliably support wide-scale usage of SMS during such elevated periods.

After the initial publication of our attacks, a number of other authors investigated malicious overload conditions in cellular networks. Serror et al [163], for example, noted that the effort needed address a linearly increasing set of request on the paging channels in CDMA networks grew extremely quickly. Racic et al [142] noted that phones themselves could easily be overloaded with requests targeted at draining their batteries. As we discuss in Chapter 6, the fundamental tension caused by the network's handling of data traffic will continue to reveal similar such attacks.

4.7 Malware

Viruses, worms and spyware continue to plague traditional data networks. It should therefore come as little surprise that increasing connectivity with the Internet and dramatically improved device capability have raised the specter of malware polluting cellular networks. While the impact of such malware has not yet been felt to any great extent in this environment, there is little evidence to suggest that this environment is any less susceptible.

The vast majority of viruses targeting cellular phones have not been observed "in the wild". Specifically, most such malware has been created as a proof of concept and attack a variety of services. Skulls and its variants, for example, pose as a benign Flash application and disable SMS and the *Multimedia Messaging Service* (MMS) functionality on infected Symbian Series 60 phones [66]. The Cabir worm, alternative, spreads via Bluetooth. The functionality of the worm, however, is extremely limited. For example, it requires the user to accept it being downloaded before it can infect a targeted device. Once infection occurs, the worms simply attempts to replicate itself again [62]. Commwarrior-A because the first virus to spread via MMS [64]. Mabir, a worm related to Cabir, uses the propagation methods of the previously mentioned viruses (SMS and MMS) to infect new hosts [65]. Viruses targeting other phone operating systems, such as Duts [63], have also been identified. While a number of companies currently offer antivirus software for such devices, the vast majority of users do not use such services.

Internet-based malware may also present a threat to cellular networks. Ricciato [149] surmised that the traffic generated by Internet-scale worms such as Slammer may be sufficient to accidentally deny service to users of cellular data networks. Off the record conversations with providers and equipment manufacturers have indicated that significant activity has already been observed in the other direction. Specifically, laptops equipped with GPRS modems and infected with spyware and malware cause significant and unexpected spikes in network usage and behavior. Such behavior is expected to become even more commonplace as so-called “smart” phones become the norm.

The potential for such malware to rapidly expand is greatly assisted by the number and systematic nature of the vulnerabilities in cellular-capable devices. Mulliner and Vigna [126] investigated the security of MMS agents implemented on PocketPC-based phones and discovered a number of previously unknown vulnerabilities. Through the use of fuzzing techniques, this team found a number of buffer overflow vulnerabilities in this software without gaining direct access to the source code itself. By exploiting these weaknesses, an adversary may be able to gain full control over such devices. As we discuss in Chapter 5, our own work on Symbian phones noted that few internal controls exist at all. Most critically, because the device only has a single user, all commands issued to the kernel are executed without question. Such weaknesses allowed us to develop exploits including keylogging software and will ultimately allow more powerful malware to gain complete control over a mobile phone.