Research Seminar in Cryptography

Report on 'Hiding Information in Noise: Fundamental Limits of Covert Wireless Communication'

Mart Simisker
Supervised by: V. Skachek

May 30, 2019

Abstract

The aim of this report is to give an overview of the work done by B.A. Bash, D. Goeckel, S. Guha and D. Towsley on the topic of covert communication in wireless networks. The work studies information theoretic limits in radio frequency channel drawing parallels to digital steganography. It is the first step in building "shadow networks".

1 Introduction

This report will give an overview of the article 'Hiding Information in Noise: Fundamental Limits of Covert Wireless Communication' by B.A. Bash, D. Goeckel, S. Guha and D. Towsley [1]. The report first begins with an introduction to the topic, covering also the introduction section of the article. It provides some general definitions and explanations.

1.1 Introduction to the Topic

As is stated in [1], "Security and privacy are critical in modern-day wireless communication." Cryptography uses computational problems to protect the content of a message. In this case, however, the knowledge of the existence of the message is a threat. As connection metadata is widely collected, transmitting encrypted data can arouse suspicion. Many cryptographic schemes can be defeated by using non-computational attacks such as a side-channel analysis [1]. Current anonymous communication tools such as Tor [2] only work in digital networks with enough nodes. However, they are useless in wireless schemes and do not work if the user is being monitored. The authors emphasize the needs of a secure communication system to provide covert, stealth or low probability of detection/intercept (LPD/LPI) communications. Such a system will protect the contents of the message and prevents the adversary from detecting the transmission.

Next, the concept of *shadow networks* is introduced – the establishment of which is the major research goal of the covert wireless communication field. The shadow network consists of following components:

- relays that generate, transmit, receive and consume data;
- wardens that try to detect communication;
- jammers that generate artificial noise and distract wardens from detecting the presence of communications.

Figure 1: The explored area of a "shadow newtork" better illustrated in [1, Fig. 1].

To create such a network, first it is necessary to connect components by stealthy communication links. The article focuses on the fundamental limits of such point-to-point links and addresses the following question: how much information can a sender Alice reliably transmit to the intended recipient Bob while hiding it from the adversary, warden Willie? The limited case is displayed in figure 1.

In [1], first a brief overview of the field of steganography is given. Then, the fundamental limits of covert communication over analog radio-frequency (RF) channels, where information is hidden in the channel artifacts such as additive white Gaussian noise (AWGN), as well as over digital communication channels, are examined. A covert broadcast scenario is briefly covered. The paper concludes with a discussion on shadow networks and an overview of ongoing research in jammer-assisted covert communication.

1.2 Used Terminology

In this part, we bring some explanations and definitions which will be used in the latter sections.

A communication channel is called *zero-rate*, if the average number of bits that can be covertly transmitted per channel use tends to zero as the number of channel uses n grows large.

A channel is called *positive-rate*, if the average number of covertly transmittable bits is a positive number, when the number of channel uses n grows large.

A key in this context is similar to cryptographic keys, which are a bitstring of a certain length.

Let *channel use* [1] denote the unit of communication resource - a fixed time period that is used to transmit a fixed-bandwidth signal. Let n denote the total number of channel uses available to Alice and Bob.

A *discrete memoryless channel* (DMC) [1] model assumes discrete input and output, which allows it to be represented by a bipartite graph, where the vertices correspond to input and output symbols, and edges correspond to the stochastic transitions from input to output symbols. The memoryless nature means that the output symbol is statistically dependent only on the current input symbol.

A *binary symmetric channel* (BSC) [3] with crossover probability p is a channel with binary input and binary output and probability of error p; if X is the transmitted random variable and Y is the received variable, then $Pr[X \neq Y] = p$ and $Pr[X = Y] = 1 - p$. BSC is a popular example of DMC.

Additive white Gaussian noise (AWGN) [4] channel model describes a channel, where noise with normal distribution and equal power on all frequencies is added to the signal. Let Z denote the channel input – discrete domain containing values 1 and -1; let N denote the noise value, a real value given by the normal distribution. Then the output of the channel is $Y = Z + N$, a real value from the continuous domain.

Channel resolvability [1] is the minimum input entropy needed to generate a channel output that is close to the channel output distribution for a given input.

Channel capacity [1] is the maximum rate of reliable communication that is unconstrained by the security requirements. Example, the channel capacity of AWGN channel is

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right),$$

where P represents the maximum channel power and N is the variance of the normal distribution describing the noise.

2 Overview of Steganography

Steganography is the practice of hiding sensitive messages in innocous objects. According to *The Histories*, an account of the Greco-Persian Wars chapter 5, steganography can be traced back to circa 440 BCE, where Histiaeus shaved the head of a slave, tattooed a message on the slave's scalp, waited for the hair to grow back, and then sent the slave to Aristagoras. In another case, described in chapter 7, a message was hidden under the wax of a wax tablet.

In digital steganography, the following objects are defined:

- covertext – a finite-length, finite-alphabet object, such as images or software binary codes, used to conceal message in [1].
  Usually, it is not available to other parties besides the communicators.
  Its statistical distribution is known to the warden, but not to all parties.
- stegotext – obtained through embedding a hidden message in a covertext object. The statistical distribution of it does not differ much from that of its covertext.

Similarly to fields such as cryptography, the process of looking for differences between covertext and stegotext is called steganalysis (analogous to cryptanalysis).

The Effects of the Availability of the Statistical Model of the Covertext bounds the amount of information that can be secretly embedded. According to [5], if the model is available, then *positive-rate steganography* is achievable. Given an $\mathcal{O}(n)$-bit secret "key" shared between the sender and the recipient, $\mathcal{O}(n)$ bits can be embedded in an n-symbol covertext without being detected by the warden [5]. If the model is not available to the sender, up to $\mathcal{O}(\sqrt{n}\log n)$ bits can be safely embedded by modifying $\mathcal{O}(\sqrt{n})$ bits out of n in the covertext, at the cost of pre-shared $\mathcal{O}(\sqrt{n}\log n)$ secret bits. The *square root law of digital steganography* yields *zero-rate steganography* since $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n}\log n)}{n} = 0$ [1, 5]. According to [6], an *empirical* model of covertext suffices to break the square root law and achieve positive-rate steganography.

When embedding at a positive rate lets Willie obtain $\mathcal{O}(n)$ stegotext observations, the increasing size n of the covertext allows improvement of the covertext model to produce statistically matching stegotext.

Steganography is an application layer technique and thus the results have limited use in physical layer covert communication. Generally there are three assumptions:

- the stegotext is not corrupted by a noisy channel,
- the generalization of the results for steganographic systems is limited because of their finite-alphabet discrete nature

- the process of embedding a message replaces a part of the covertext.

Although this enables positive rate steganography methods, it cannot be done, unless the sender controls the wardens' noise source. Since the transmission of stegotext is the biggest obstacle in covert communication, the plan is to use channel artifacts such as noise to hide transmission.

3 Covert Communication in Physical Layer

This section begins with investigating the physical layer covert communications with respect to the RF wireless communications. Protecting communications from detection, jamming and eavesdropping has been of concern. *Spread spectrum* techniques devised between the two world wars were the earliest and most enduring form of the physical layer security. After looking at spread spectrum techniques, the square root law is used to analyse covert communication over AWGN channel. The square root law of physical layer can also give some insights to digital covert communication, which will be discussed in section 3.3.

3.1 Spread Spectrum Communications

Spread spectrum [7] techniques involves transmitting a signal on a much wider bandwidth than it is normally required, thereby suppressing the power spectral density of the transmission below the noise floor. Such systems provide covert communication and are resistant to jamming, fading and other forms of interference.

It is mentioned in [1] that *direct sequence* spread spectrum (DSSS), *frequency-hopping* spread spectrum (FHSS), and their combination are some of the most typical techniques. The following is a short explanation.

DSSS is defined as follows – the signal waveform is multiplied by a *spreading sequence*, a randomly generated binary waveform with a substantially higher bandwidth than the original signal. The resulting waveform is "spread" over a wider bandwidth, and thereby the power spectral density of the transmitted signal is reduced. The spreading sequence is a shared secret between the sender and the recipient. It is required in the process of de-spreading the signal.

Outside of the field of security, the use of public uncorrelated spreading sequences between transmitter/receiver pairs enables multiple access; DSSS thus forms the code-division multiple access (CDMA) protocols used in cellular telephony.

When using the FHSS, the carrier frequency is re-turned for each transmitted symbol. In this case, the frequency hopping pattern is a randomly generated and previously shared secret between the communicators. By combining FHSS with orthogonal frequency-division multiplying (OFDM), the use of multiple carrier frequencies can be enabled.

By combining FHSS with the time-hopping techniques, the average transmitted symbol power can be further reduced.

The authors point out, that even though the spread spectrum architectures are well-developed, the analytical evaluation on this topic has been sparse. Secrecy and undetectability in a multiple-input multiple-output setting with focus on the signal processing aspects has been studied by A. Hero in [8]. The results show that covert communication systems are constrained by average power, and there is also a need to explore the fundamental information theoretic limits.

The authors emphasize the importance of the knowledge of limits of a communication system, particularly since modern coding techniques allow 3G/4G cellular systems to operate near their theoretical channel capacity.

Although the secrecy part of [8] has received a lot of attention, the covert communication part had been largely overlooked until the authors' work.

They note, that the following fundamental results apply to both the classical spread-spectrum systems as well as to the modern covert communication proposals which rely on the channel noise and equipment imperfections to hide communication. One such example is discussed in [9].

3.2 Square Root Law for Covert Communication Over the AWGN Channels

Spread spectrum systems allow communication in prohibited area because the signal power is spread over a large time-frequency space, reducing the wardens' signal-to-noise ratio (SNR). This impairs the wardens' ability to discriminate between the noise and the information-carrying signal corrupted by noise. The authors are interested with how small the power has to be to be fundamentally undetectable, and how much covert information can be transmitted reliably [1].

The following example is given: consider an additive white Gaussian noise (AWGN) channel model where the signaling sequence is corrupted by the addition of sequence of independent and identically distributed zero-mean Gaussian random variables with variance σ^2. This is the standard model for a free-space RF channel. The channels from Alice to Bob and to Willie (the recipient and the warden) are subject to AWGN with respective variables $\sigma_b^2 > 0$ and $\sigma_w^2 > 0$. (If $\sigma_b^2 = 0$, Alice can transmit infinite number of bits without being detected. If $\sigma_w^2 = 0$, covert communication is impossible).

Next, they look at what the warden can observe. When Alice is not transmitting, the warden observes the AWGN with the total power $\sigma_w^2 n$ over n channel uses on average. By the standard statistical arguments the observations lie within $\pm c\sigma_w^2\sqrt{n}$ of the average, where c is a constant. If Alice wants to transmit, the total power that this transmission can emit over n channel uses is limited to $\mathcal{O}(\sigma_w^2\sqrt{n})$. Otherwise the transmission will be detected. The amount of bits that Alice can transmit is bounded from above by $\mathcal{O}\left(\sigma_w^2\sqrt{n}/\sigma_b^2\right)$ [10]. Similarly to the square root law, this yields a zero-rate channel. The similarity between the square root law for covert communications and the steganographic square root law is attributable to mathematics of statistical hypothesis testing. In the steganographic square root law, the channel to Bob is noiseless, and therefore the additional factor $\log n$ is added.

Next consider an example scheme, where Alice and Bob can reliably transmit $\mathcal{O}\left(\sigma_w^2\sqrt{n}/\sigma_b^2\right)$ covert bits using binary amplitude modulation, any error-correction code (which can be publicly known), and $\mathcal{O}\left(\sqrt{n}\log n\right)$ pre-shared secret bits. In this scheme, the secret contains a random subset $\mathcal{S}$ of n available channel uses and a random one-time pad of size $|\mathcal{S}|$.

Subset $\mathcal{S}$ is generated by flipping a random biased coin with the probability of heads being $\mathcal{O}(1/\sqrt{n})$ – a channel use is used to transmit, if the i-th flip is heads. The one-time pad is of length $|\mathcal{S}| = \mathcal{O}(\sqrt{n})$ on average (this is dependent on the number of 1's in $\mathcal{S}$). Bob will only observe the channel uses as guided by $\mathcal{S}$. Since the warden does not know $\mathcal{S}$, all of the channel uses must be observer, most of which is noise. The intent of the one-time pad is to prevent the warden from using the structure of the error correction code to detect a transmission.

The authors of [1] emphasize that even though the size of the pre-shared secret is asymptotically larger than the size of the transmitted message, this trade-off is acceptable as opposed to being detected. The extension of [10] to digital covert communication suggests that in some cases the pre-shared secret can be eliminated.

3.3 Digital Covert Communication

It is said in [1] that "*discrete memoryless channel* (DMC) model describing digital communication often sheds light on what is feasible in practical communication systems."

First, consider the *binary symmetric channel* (BSC), which restricts the DMC to binary input output alphabet {0,1}. Denote by $p_b > 0$ and $p_w > 0$ the respective crossover probabilities of Bob's and the warden's BSC. In [11], it has been shown that if $p_w > p_b$, then although no more than $\mathcal{O}(\sqrt{n})$ bits can be reliably transmitted in n BSC uses, the pre-shared secret is unnecessary.

In [11] channel *resolvability* is used to generalize the square root law to DMCs. The authors say that in [12], resolvability was used to obtain new, stronger results for the information-theoretic secrecy capacity. According to [13], if the channels from the transmitter are DMCs, and the channel between the transmitter and the warden is worse (according to some parameter specific to the channel model type) than the one between the transmitter and Bob, then techniques from [11] and [14] can be used to demonstrate the square root law without the pre-shared secret.

Furthermore, as long as the transmitter knows the channel to the warden, $\mathcal{O}(\sqrt{n})$ pre-shared secret bits are sufficient for covert communication, with the assumption that the wardens' channel capacity is greater-than-or-equal to that of Bob's channel capacity [13]. For AWGN channels, the paper mentions the existence a covert communication scheme, which uses $\mathcal{O}(\sqrt{n})$ pre-shared secret bits. If the noise power at the warden's receiver is greater than that at Bob's receiver, then secret-less covert communication is achievable.

3.4 The Effects of Wardens Ignorance of Transmission Time

The square root laws are derived with assumption that the warden knows when the transmission takes place. However that is not always the case - the transmission window could also be a very short time of the day meaning that the warden has to monitor a larger timeframe for transmissions. Additional technical details are available in [15].

3.5 Positive-rate Covert Communication

The previously described communication channels are zero rate, since the average number of bits that can be covertly transmitted per channel use tends to zero as n, the number of channel uses, grows large. However, positive-rate communications offers more interest, as it would ensure the transmission of a certain number of bits. The general circumstances when the transmitter can covertly communicate are when the warden either allows them to or when the warden is ignorant of the probabilistic structure of the noise on his channel.

When the warden allows communications, the covert capacity is the same as information-theoretic secrecy capacity.

Incompleteness of the wardens noise model can also allow for positive-rate covert communications. In the noisy digital channel setting, the wardens ignorance is a special case of the scenario in [14].

It has been show in [16], that in the AWGN channel setting, random noise power fluctuations yield positive-rate covert communications. This holds, even when the noise power is bounded, because the warden does not have a constant baseline of noise for the comparison.

3.6 Covert Broadcast

Some of the results of the point-to-point covert communication in the presence of a single warden that were discussed up to this point can be extended to scenarios with multiple independently-controlled receivers. For example in the case of the AWGN channel, there is an effective power constraint on the transmitter. If the secrets have been previously shared, the transmitter can use standard techniques to encode covert messages to multiple recipients.

4 Conclusion

The main objective of this work is to enable a wireless "shadow network", comprised of transmitters, receivers and friendly jammers that generate artificial noise, impairing wardens' ability to detect transmissions. In [17], the positive use of jammers on the information-theoretical secrecy has been shown. In covert networks the jammers act independently of the transmitters, thereby the jammers have a parasitic effect on the wardens' SNR. It is important to characterize how the system scales, similarly to what was done with secure multipath unicast communication in large wireless networks in [18].

The next step is extending the scenario of this article to point-to-point jammer-assisted covert communication in the presence of multiple wardens. The authors of [1] stress that "Preliminary results [19] assume that jammers operate at a constant power, and the signal propagation model accounts only for the path loss and AWGN." They also note that [16] has demonstrated that the uncertainty of noise on wardens' channel is beneficial to Alice. This is why the authors suggest to incorporate a variable jamming power and multipath fading into the covert communication model, with the hopes of achieving positive-rate covert communication.

Acknowledgement

I would like to thank Vitaly Skachek for his guidance.

References

[1] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: fundamental limits of covert wireless communication," *IEEE Communications Magazine*, vol. 53, pp. 26–31, Dec 2015.

[2] "Tor project webpage." `https://www.torproject.org`. Last accessed 2019.05.07.

[3] "Binary symmetric channel." `https://en.wikipedia.org/wiki/Binary_symmetric_channel`. Last accessed 2019.05.07.

[4] "Additive white gaussian noise." `https://en.wikipedia.org/wiki/Additive_white_Gaussian_noise`. Last accessed 2019.05.07.

[5] J. Fridrich, *Steganography in digital media: principles, algorithms, and applications*. Cambridge University Press, 2009.

[6] S. Craver and J. Yu, "Subset selection circumvents the square root law," in *Media Forensics and Security II*, vol. 7541, p. 754103, International Society for Optics and Photonics, 2010.

[7] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread spectrum communications handbook*, vol. 2. 1994.

[8] A. O. Hero, "Secure space-time communication," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3235–3249, 2003.

[9] A. Dutta, D. Saha, D. Grunwald, and D. Sicker, "Secret agent radio: Covert communication through dirty constellations," in *International Workshop on Information Hiding*, pp. 160–175, Springer, 2012.

[10] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[11] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *IEEE International Symposium on Information Theory*, pp. 2945–2949, 2013.

[12] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8077–8098, 2013.

[13] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, pp. 2334–2354, May 2016.

[14] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *IEEE International Symposium on Information Theory*, pp. 601–605, 2014.

[15] B. A. Bash, D. Goeckel, and D. Towsley, "Lpd communication when the warden does not know when," in *IEEE International Symposium on Information Theory*, pp. 606–610, June 2014.

[16] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, pp. 1195–1205, Oct 2015.

[17] L. Lai and H. El Gamal, "The relay–eavesdropper channel: Cooperation for secrecy," *IEEE Transactions on Information Theory*, vol. 54, pp. 4005–4019, Sep. 2008.

[18] C. Çapar, D. Goeckel, B. Liu, and D. Towsley, "Secret communication in large wireless networks without eavesdropper location information," in *Proceedings IEEE INFOCOM*, pp. 1152–1160, March 2012.

[19] R. Soltani, B. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert single-hop communication in a wireless network with distributed artificial noise generation," in *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1078–1085, Sep. 2014.