Covert Communication with Finite Blocklength in AWGN Channels

Shihao Yan[†], Biao He[‡], Yirui Cong[†], and Xiangyun Zhou[†]
[†]Research School of Engineering, The Australian National University, Canberra, ACT, Australia
[‡]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong
Emails: {shihao.yan, yirui.cong, xiangyun.zhou}@anu.edu.au, eebiaohe@ust.hk

***Abstract*—Covert communication is to achieve a reliable transmission from a transmitter to a receiver while guaranteeing an arbitrarily small probability of this transmission being detected by a warden. In this work, we study the covert communication in AWGN channels with finite blocklength, in which the number of channel uses is finite. Specifically, we analytically prove that the entire block (all available channel uses) should be utilized to maximize the effective throughput of the transmission subject to a predetermined covert requirement. This is a nontrivial result because more channel uses results in more observations at the warden for detecting the transmission. We also determine the maximum allowable transmit power per channel use, which is shown to decrease as the blocklength increases. Despite the decrease in the maximum allowable transmit power per channel use, the maximum allowable total power over the entire block is proved to increase with the blocklength, which leads to the fact that the effective throughput increases with the blocklength.**

I. Introduction

In future wireless networks, the demand for wireless data is growing at such a rate that requires 1000x today's capacity in the next five to ten years. Against this background, crucial concerns on the security and privacy of wireless communications are emerging since a large amount of confidential information (e.g., email/bank account information and password, credit card details) is transferred over wireless networks. In addition to the secrecy and integrity of the transmitted information, in some scenarios a user may wish to transmit messages over wireless networks without being detected. This is due to the fact that (for example) the exposure of this transmission may disclose the user's location information, which probably violates the privacy of the user. Therefore, covert communication is attracting an increasing amount of research interests recently (e.g., [1–3]). In covert communication, a transmitter (Alice) intends to communicate with a legitimate receiver (Bob) without being detected by a warden (Willie), who is observing this communication.

In fact, covert communication was addressed by spread spectrum techniques in the early 20th century and a review on spread spectrum techniques can be found in [4]. However, the performance limit of covert communication has not been fully examined in the literature and recently attracts much research attention. Considering additive white Gaussian noise (AWGN) channels, a square root law has been derived in [5], which states that Alice can transmit no more than $\mathcal{O}(\sqrt{n})$ bits in n channel uses covertly and reliably to Bob. Following [5], the scaling constant of the amount of information with respect to the square root of n was characterized for a broad class of discrete memoryless channels (DMCs) and AWGN channels in [6]. We note that this square root law requires a pre-shared secret to be established between Alice and Bob prior to Alice's transmission. This pre-shared secret is proved to be unnecessary for the square root law when the channel quality from Alice to Bob is higher than that from Alice to Willie, for binary symmetric channel (BSC) [7], DMC [8], and AWGN channel [8].

In the square root law we have $\mathcal{O}(\sqrt{n})/n \rightarrow 0$ as $n \rightarrow \infty$, which states that the rate is asymptotically zero (i.e., the average number of bits that can be covertly and reliably transmitted per channel use asymptotically approaches zero). However, in some scenarios a positive rate has been proved to be achievable (e.g., [7, 9–13]). For example, it is proved that a positive rate can be obtained when Willie has uncertainty about the receiver noise variance in AWGN channels [11, 13], when Willie does not exactly know the receiver noise model in BSC channels [7], or when Willie lacks knowledge of his channel characteristics in AWGN and block fading channels [12, 13]. In addition to the noise or channel uncertainty, as proved in [10] a positive rate can also be achieved when Willie has uncertainty on the time instant of the communication.

In the literature as seen in the aforementioned works, only [11] mentioned the impact of finite samples (i.e., finite n) on the detection performance at Willie. It is numerically shown that with noise uncertainty at Willie there may exist an optimal number of samples that maximizes the communication rate subject to $\xi \geq 1-\epsilon$, where ξ is the sum of false positive and miss detection rates at Willie and $0 < \epsilon \leq 1$ is an arbitrarily small number. Besides the detection performance at Willie, finite n also has significant impact on the maximal achievable rate R of the channel from Alice to Bob (i.e., the maximal achievable rate decreases as n decreases for a fixed decoding error probability δ) [14], which has not been considered in the literature of covert communication (including [11]). Therefore, the impact of finite n on covert communication has not been well examined. This leaves a significant gap in our understanding of the performance limit of practical covert communication, since in practice the length of a codeword is always finite. For example, to achieve transmission efficiency (e.g., short delay) we may require the codeword to be short (e.g., in the order of 100 channel uses) for vehicle-to-vehicle communication or real-time video processing [15].

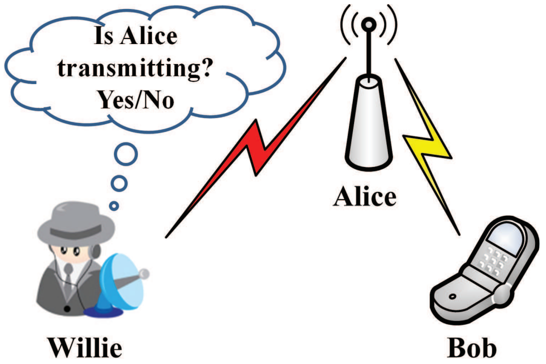

Fig. 1. Illustration of the system model of interest for covert communication.

A. Our Contributions

Considering AWGN channels, we study the impact of finite n on both the maximal achievable rate at Bob and detection performance at Willie in covert communication. To this end, noting that the decoding error probability δ is not negligible when n is finite, we first propose to adopt the effective throughput η (i.e., $\eta = nR(1-\delta)$) subject to $\xi \geq 1-\epsilon$, as the performance metric to evaluate covert communication. As can be seen from the definition of η, it explicitly captures the tradeoff among n, R, and δ for a given covert requirement.

We consider a maximum blocklength of N channel uses, in which the covert information needs to be transmitted. Hence, the actual number of channel uses n is constrained by $n \leq N$. Although a larger n offers more observations to Willie for detecting the transmission, we analytically prove that the optimal value of n that maximizes η subject to the given covert requirement is N (i.e., the entire block with all available channel uses). We also determine the maximum allowable transmit power per channel use (denoted by P^*) that achieves the maximum η. Our examination shows that P^* decreases as N increases, which is due to the fact that increasing N forces Alice to allocate less power for each channel use to meet the covert requirement. Nevertheless, we show that the maximum allowable total transmit power (i.e., NP^*) increases as N increases, which leads to the fact that the effective throughput of the communication from Alice to Bob increases as N increases. The results in this paper, for the first time, provide important insights on the design of covert communication with a finite blocklength.

Notations: Scalar variables are denoted by italic symbols. Vectors and matrices are denoted by lower-case and upper-case boldface symbols, respectively. Given a vector $\mathbf{x}$, $x[i]$ denotes the i-th element of $\mathbf{x}$. The expectation is denoted by $\mathbb{E}\{\cdot\}$ and $\mathcal{CN}(0,\sigma^2)$ denotes the circularly-symmetric complex normal distribution with zero mean and variance σ^2.

II. System Model

A. Channel Model

The system model of interest for covert communication is illustrated in Fig. 1, where each of Alice, Bob, and Willie is equipped with a single antenna. We assume the channel from Alice to Bob and the channel from Alice to Willie are only subject to AWGN. In the covert communication, Alice transmits n complex-valued symbols $x[i]$ $(i = 1, 2, \ldots, n)$ in each codeword to Bob, while Willie is passively collecting n observations on Alice's transmission to detect her presence (i.e., whether Alice is transmitting). In this work, we consider that the length of a codeword is constrained by a maximum blocklength denoted by N. Thus, we have $n \leq N$ as a constraint on n. We denote the AWGN at Bob and Willie as $r_b[i]$ and $r_w[i]$, respectively, where $r_b[i] \sim \mathcal{CN}(0,\sigma_b^2)$, $r_w[i] \sim \mathcal{CN}(0,\sigma_w^2)$, σ_b^2 and σ_w^2 are the noise variances at Bob and Willie, respectively. In addition, we assume that $x[i]$, $r_b[i]$, and $r_w[i]$ are mutually independent. We denote the transmit power of Alice as P (i.e., $\mathbb{E}\{|x[i]|^2\} = P$). Furthermore, we assume that Alice adopts Gaussian signaling, i.e., $x[i] \sim \mathcal{CN}(0,P)$.

B. Channel Coding Rate for Finite Blocklength

The received signal at Bob for each signal symbol is given by

$$y_b[i] = x[i] + r_b[i]. \tag{1}$$

As pointed out by [14], the decoding error probability at Bob is not negligible when n is finite. As such, for a given decoding error probability δ the channel coding rate of the channel from Alice to Bob can be approximated by [14, 16]

$$R \approx \log_2(1+\gamma_b) - \sqrt{\frac{\gamma_b(\gamma_b+2)}{n(\gamma_b+1)^2}}\frac{Q^{-1}(\delta)}{\ln(2)} + \frac{\log_2(n)}{2n}, \tag{2}$$

where $\gamma_b = P/\sigma_b^2$ is the signal-to-noise ratio (SNR) at Bob, and $Q^{-1}(\cdot)$ is the inverse Q-function. Equivalently, for a given channel coding rate R, the decoding error probability at Bob is given by

$$\delta = Q\left(\frac{\sqrt{n}(1+\gamma_b)\left(\ln(1+\gamma_b)+\frac{1}{2}\ln(n)-R\ln 2\right)}{\sqrt{\gamma_b(\gamma_b+2)}}\right). \tag{3}$$

C. Binary Hypothesis Testing at Willie

In order to detect Alice's presence, Willie is to distinguish the following two hypotheses

$$\begin{cases} \mathcal{H}_0: \ y_w[i] = r_w[i] \\ \mathcal{H}_1: \ y_w[i] = x[i] + r_w[i], \end{cases} \tag{4}$$

where $\mathcal{H}_0$ denotes the null hypothesis where Alice is not transmitting, $\mathcal{H}_1$ denotes the alternative hypothesis where Alice is transmitting, and $y_w[i]$ is the received signal at Willie. Following the assumptions detailed in Section II-A, we have the likelihood functions of $y_w[i]$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ as $f(y_w[i]|\mathcal{H}_0) = \mathcal{CN}(0,\sigma_w^2)$ and $f(y_w[i]|\mathcal{H}_1) = \mathcal{CN}(0,P+\sigma_w^2)$, respectively. In the cover communication, the ultimate goal of Willie is to minimize the total error rate, which is given by

$$\xi = P_F + P_M, \tag{5}$$

where $P_F \triangleq \Pr(\mathcal{D}_1|\mathcal{H}_0)$ is the false positive rate, $P_M \triangleq \Pr(\mathcal{D}_0|\mathcal{H}_1)$ is the miss detection rate, $\mathcal{D}_1$ and $\mathcal{D}_0$ are the binary decisions that infer whether Alice is present or not, respectively. We assume that Willie knows both P and σ_w^2 exactly, and thus the optimal test that minimizes ξ is the likelihood ratio test with $\lambda = 1$ as the threshold[1], which is given by

$$\frac{\mathbb{P}_1 \triangleq \prod_{i=1}^{n} f\left(y_w[i]|\mathcal{H}_1\right)}{\mathbb{P}_0 \triangleq \prod_{i=1}^{n} f\left(y_w[i]|\mathcal{H}_0\right)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} 1. \tag{6}$$

After performing some algebraic manipulations, (6) can be reformulated as

$$T \triangleq \frac{1}{n}\sum_{i=1}^{n}|y_w[i]|^2 \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \Gamma, \tag{7}$$

where T is the average power of each received symbol at Willie and Γ is the threshold for T, which is given by

$$\Gamma = \frac{(P+\sigma_w^2)\sigma_w^2}{P}\ln\left(\frac{P+\sigma_w^2}{\sigma_w^2}\right). \tag{8}$$

As per (6) and (7), we note that the radiometer is indeed the optimal detector when Willie knows the likelihood functions exactly (i.e., there are no nuisance parameters embedded in the likelihood functions). Following (7) and noting that T is a chi-squared random variable with $2n$ degrees of freedom, the false positive rate and miss detection rate are given by [9, 11]

$$P_F = \Pr(T > \Gamma|\mathcal{H}_0) = 1 - \frac{\gamma\left(n, \frac{n\Gamma}{\sigma_w^2}\right)}{\Gamma(n)}, \tag{9}$$

$$P_M = \Pr(T < \Gamma|\mathcal{H}_1) = \frac{\gamma\left(n, \frac{n\Gamma}{P+\sigma_w^2}\right)}{\Gamma(n)}, \tag{10}$$

where $\Gamma(n) = (n-1)!$ is the gamma function and $\gamma(\cdot,\cdot)$ is the incomplete gamma function given by

$$\gamma(n, x) = \int_0^x e^{-t}t^{n-1}dt. \tag{11}$$

With the radiometer as the optimal detector, following Pinsker's inequality, we have a lower bound on ξ, which is given by [5, 17, 18]

$$\xi \geq 1 - \sqrt{\frac{1}{2}\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)}, \tag{12}$$

where $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)$ is the Kullback-Leibler (KL) divergence from $\mathbb{P}_0$ to $\mathbb{P}_1$, which can be expressed as

$$\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) = n\left[\ln\left(\frac{P+\sigma_w^2}{\sigma_w^2}\right) - \frac{P}{P+\sigma_w^2}\right]. \tag{13}$$

D. Covert Requirement

Covert communication requires $\xi \geq 1-\epsilon$ for some arbitrarily small ϵ. As per (12), we can ensure $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ in order to guarantee $\xi \geq 1-\epsilon$. We note that $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ is a more strict constraint relative to $\xi \geq 1-\epsilon$ as per (12). From a conservative point of view and to avoid the complex expressions for P_F and P_M, in this work we adopt $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the requirement for covert communication. Also, the value of ϵ is especially very small in order to provide good covertness. Thus, in this work we only consider $\epsilon \in (0, 0.5]$ because $\epsilon > 0.5$ means that Willie is allowed to achieve more than 50% success detection rate.

III. Covert Communication with a Finite Number of Channel Uses

In this section, we first adopt the effective throughput to evaluate the performance of covert communication in AWGN channels with finite blocklength. Then, we determine the optimal n and P that maximize this effective throughput subject to the covert requirement.

A. Effective Throughput

The square root law states that Alice can transmit no more than $\mathcal{O}(\sqrt{n})$ bits in n channel uses covertly and reliably to Bob. Such scaling-law results are obtained when $n \to \infty$. As such, these square-law results cannot be applied in the covert communication with finite n. In this work, we focus on the amount of information that can be transmitted reliably from Alice to Bob for a given positive ϵ. Noting that the decoding error probability of a channel with finite blocklength is not negligible, we adopt the effective throughput from Alice to Bob as the main performance metric for the covert communication with finite blocklength, while utilizing the covert requirement as the constraint. The effective throughput from Alice to Bob is defined as [19, 20]

$$\eta = nR(1-\delta). \tag{14}$$

We note that η gives the average number of information bits that can be transmitted from Alice to Bob reliably (excluding information bits suffering from decoding errors) by utilizing a codeword with finite length n.

B. Optimal Number of Channel Uses and Transmit Power

The ultimate goal of our design in covert communication is to achieve the maximum effective throughput while guaranteeing the covert requirement. To this end, we first consider a fixed channel coding rate R and focus on the design of the number of channel uses and the transmit power P, since the design of n and P affects both the effective throughput from Alice to Bob and the detection performance at Willie. As such, for a given R the optimization problem in the covert communication of interest can be written as

$$\underset{n,P}{\operatorname{argmax}}\, \eta, \tag{15}$$

$$\text{s.t.} \quad \mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2, \tag{16}$$

$$n \leq N. \tag{17}$$

[1] We note that $\lambda = 1$ is due to the unknown or equal *a priori* probabilities, i.e., P_0 and P_1 are unknown or equal, where P_0 is the *a priori* probability that $\mathcal{H}_0$ is true, P_1 is the *a priori* probability that $\mathcal{H}_1$ is true, and $P_0 + P_1 = 1$. If both P_0 and P_1 are known, the total error rate is reformulated as $\xi = P_0P_F + P_1P_M$ and the optimal test that minimizes this reformulated ξ is the likelihood ratio test with $\lambda = P_1/P_0$. We also note that the assumption of equal *a priori* probabilities is commonly adopted in the literature of covert communication (e.g., [5, 10, 11]).

***Theorem** 1:* The optimal values of n and P that maximize the effective throughput η subject to $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ and $n \leq N$ are derived as

$$n^* = N, \tag{18}$$

$$P^* = (\sigma_w^2 + P^*)\left[\ln\left(\frac{P^*}{\sigma_w^2}+1\right) - 2\epsilon^2 N\right], \tag{19}$$

where P^* is the solution to the fixed-point equation (19).

Proof: The detailed proof is provided in Appendix. ■

Based on Theorem 1, we see that it is best for Alice to transmit over all available channel uses for covert communication, provided that the transmit power is optimized to maintain the same level of covertness despite that Willie has more observations when n is larger. The same level of covertness is achieved by reducing the transmit power when n becomes larger, which can be seen from (19) that P^* decreases with N. It is interesting to observe that both n^* and P^* are not functions of R. This demonstrates that the obtained n^* and P^* are globally optimal, regardless the value of the channel coding rate R. As such, the optimal value of R that maximizes the effective throughput subject to the covert requirement can be obtained through

$$R^* = \underset{0\leq R}{\operatorname{argmin}}\, NR\left[1-\delta(P^*,N,R)\right], \tag{20}$$

where $\delta(P^*,N,R)$ is obtained by substituting $P = P^*$ and $n^* = N$ into (3). We note that R^* can be also obtained through searching the optimal value of δ that maximizes η for $n^* = N$ and $P = P^*$. We define $\delta^* = \delta(P^*,N,R^*)$ and denote the maximum effective throughput as η^*, which is achieved by substituting P^*, n^*, R^*, and δ^* into (14).

IV. Numerical Results

In this section, we provide numerical results on the effective throughput subject to $\xi \geq 1-\epsilon$ to verify our analysis on the covert communication with $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the constraint.

In Fig. 2 we plot the maximum allowable total transmit power NP^* over the entire block versus ϵ. In this figure and the following figures, the curves for $\xi \geq 1-\epsilon$ are achieved by numerically evaluating the false positive and detection rates as per (9) and (10), respectively. In this figure, we observe that the NP^* with $\xi \geq 1-\epsilon$ as the constraint is higher than that with $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the constraint. This is due to the fact that the equality in (12) cannot be achieved in the considered system model, and hence $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ is a more strict constraint than $\xi \geq 1-\epsilon$. We also observe that NP^* increases (hence the effective throughput increases) as N increases, which can be explained by our Theorem 1. Finally, we observe that NP^* decreases (hence the effective throughput decreases) as ϵ decreases, which demonstrates the tradeoff between the covert requirement and the achievable effective throughput (e.g., a more strict covert requirement leads to a smaller effective throughput).

In Fig. 3, we plot NP^*, η, P^*, and η/N versus N in different sub-figures, respectively. As expected, we first observe that NP^* and η monotonically increase as N increases in Fig. 3 (a) and Fig. 3 (b), respectively. Although NP^* increases as shown in Fig. 3 (a), it is interesting to observe that the maximum allowable transmit power P^* monotonically decreases as N increases in Fig. 3 (c). This can be explained by (19) in our Theorem 1. Intuitively, this is due to the fact that as the number of observations at Willie increases, Alice has to reduce her transmit power in order to meet the same detection performance at Willie. In Fig. 3 (d), we observe that the effective throughput per channel use (i.e., η/N) monotonically increases as N increases. This is due to the fact that the decrease in δ (i.e., the decoding error probability given in (3)) caused by increasing N is more than the increase in δ caused by the reduction of P^* as shown in Fig. 3 (c). These aforementioned observations based on Fig. 3 demonstrate that increasing N not only helps Alice to allocate less transmit

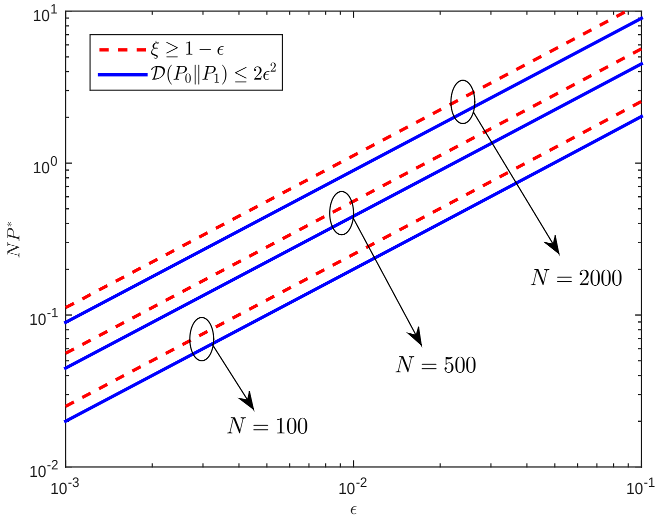

Fig. 2. Maximum allowable total transmit power NP^* versus ϵ for different values of N, where $\sigma_b^2 = \sigma_w^2 = 1$.

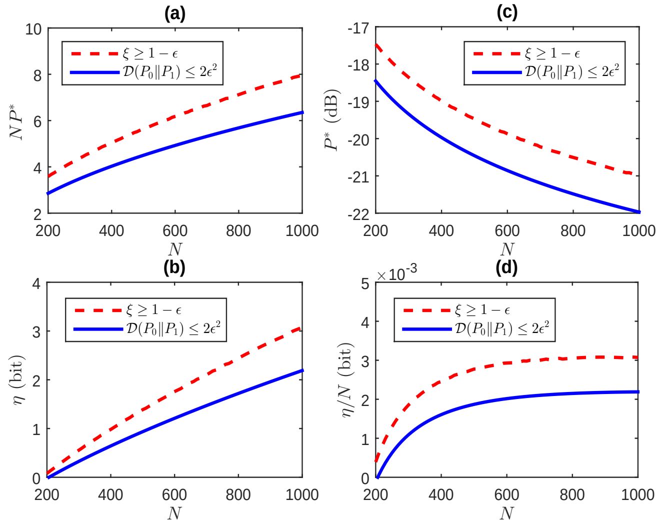

Fig. 3. NP^*, η, P^*, and η/N versus N, where $\sigma_b^2 = \sigma_w^2 = 1$, $\delta = 0.01$, and $\epsilon = 0.1$.

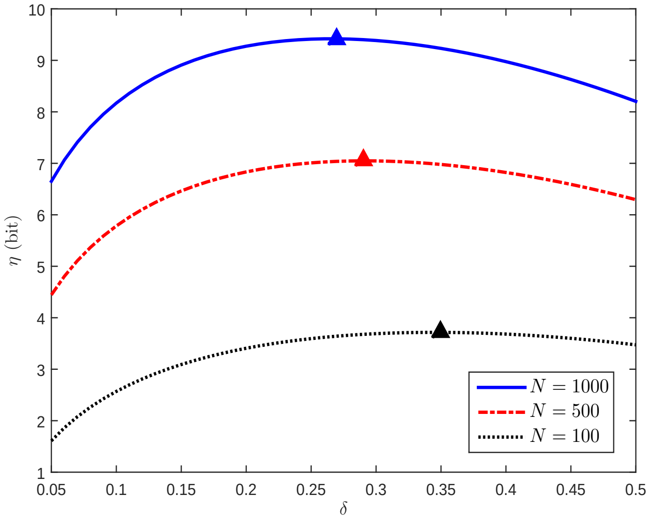

Fig. 4. Effective throughput η versus the decoding error probability δ for different values of N, where $\sigma_b^2 = \sigma_w^2 = 1$ and $\epsilon = 0.1$.

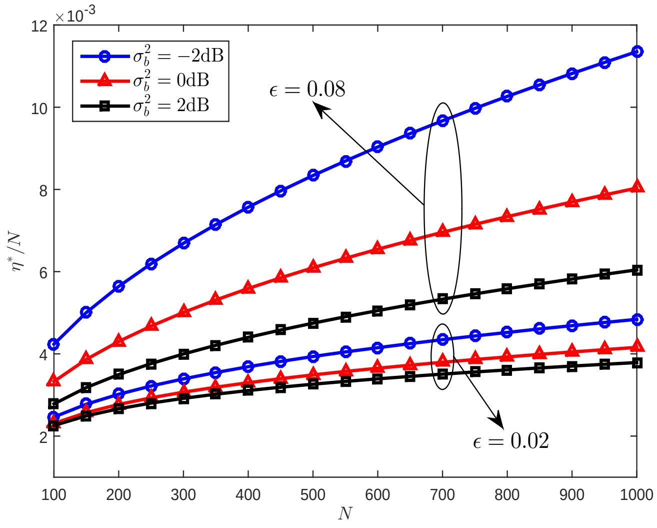

Fig. 5. Maximum effective throughput per channel use η^*/N versus N for different values of ϵ and σ_b^2, where $\sigma_w^2 = 1$.

power to each channel use in order to maintain the same level of covertness, but also reduces the decoding error probability in the communication from Alice to Bob, which turns out to improve the effective throughput of the covert communication.

In Fig. 4, we plot the effective throughput η subject to $\xi \geq 1 - \epsilon$ versus the decoding error probability δ. We first observe that the optimal value of δ that maximizes η indeed exists, based on which we can determine the optimal R. We also observe that the optimal value of δ decreases as N increases. As shown in Fig. 3 (c), the maximum allowable transmit power P^* decreases as N increases. As per (2), the observation, that both δ^* and P^* decreases as N increases, indicates that the optimal channel coding rate R^* decreases as N increases. We also plot the maximum effective throughput per channel use (i.e., η^*/N) versus N in Fig. 5. In this figure, we first observe that as N increases η^*/N increases, which is consistent with our observation found in Fig. 3 (d). We also observe that as ϵ increases slightly (e.g., from 0.02 to 0.08) η^*/N significantly increases. This demonstrates that the achievable effective throughput is very sensitive to the the covert requirement.

V. Conclusion

This work investigated the covert communication with finite blocklength (i.e., a finite number of channel uses $n \leq N$) over AWGN channels. We proved that the effective throughput of covert communication is maximized when all available channel uses are utilized, i.e., $n^* = N$. To guarantee the same level of covertness, the maximum allowable transmit power per channel use decreases as N increases, while the maximum allowable total transmit power over all channel uses increases as N increases. In contrast, we found that both the effective throughput and the effective throughput per channel use increase as N increases. This is due to the fact that increasing N not only reduces the transmit power allocated to each channel use, but also decreases the decoding error probability of the communication from Alice to Bob.

Acknowledgments

This work was supported by the Australian Research Council's Discovery Projects (DP150103905).

Appendix

We present our proof of Theorem 1 in the following 6 steps.

Step 1: We note that η and $\mathcal{D}(\mathbb{P}_0 \| \mathbb{P}_1)$ are both monotonically increasing functions of P and n. As such, we can conclude that the equality in the constraint (16) is always met in order to maximize η. Thus, we have $\mathcal{D}(\mathbb{P}_0 \| \mathbb{P}_1) = 2\epsilon^2$ and following (13) we have

$$n = \frac{2\epsilon^2}{f(\gamma_w)}, \tag{21}$$

where

$$f(\gamma_w) \triangleq \frac{\mathcal{D}(\mathbb{P}_0 \| \mathbb{P}_1)}{n} = \ln(\gamma_w + 1) - \frac{\gamma_w}{\gamma_w + 1}, \tag{22}$$

and $\gamma_w = P/\sigma_w^2$ is the SNR at Willie.

Step 2: We note $f(0) = 0$ and we derive the first derivative of $f(\gamma_w)$ with respect to γ_w as

$$\frac{\partial f(\gamma_w)}{\partial \gamma_w} = \frac{\gamma_w}{(\gamma_w + 1)^2} \geq 0, \tag{23}$$

which leads to the fact that $f(\gamma_w)$ is a monotonically increasing function of γ_w. With the constraint $\mathcal{D}(\mathbb{P}_0 \| \mathbb{P}_1) = 2\epsilon^2$, n is a monotonically decreasing function of $f(\gamma_w)$ as per (21), which results in that n is a monotonically decreasing function of γ_w (thus of P).

Step 3: Instead of directly proving $n^* = N$ for maximizing the effective throughput, we next prove that $n^* = N$ maximizes $n\gamma_w$ (i.e., maximizes nP) under the constraint (21) in the remaining steps. This is due to the fact that nP is the total transmit power for the n channel uses and the effective throughput increases as the total transmit power increases [14].

Step 4: We next prove that either $n = 1$ or $n = N$ maximizes $n\gamma_w$. To this end, in the following we first show that $n\gamma_w$ initially decreases and then increases with n. Following (21) and (22), we have

$$n\gamma_w = \frac{2\epsilon^2}{g(\gamma_w)}, \tag{24}$$

where $g(\gamma_w)$ is given by

$$g(\gamma_w) = \frac{\ln(1+\gamma_w)}{\gamma_w} - \frac{1}{1+\gamma_w}. \tag{25}$$

We then derive the first derivative of $g(\gamma_w)$ with respect to γ_w as

$$\frac{\partial g(\gamma_w)}{\partial \gamma_w} = \frac{h(\gamma_w)}{\gamma_w^2(1+\gamma_w)^2}, \tag{26}$$

where

$$h(\gamma_w) = 2\gamma_w^2 + \gamma_w - (1+\gamma_w)^2 \ln(1+\gamma_w). \tag{27}$$

We note that there are *only* two solutions to $h(\gamma_w) = 0$ for $\gamma_w \geq 0$, including $\gamma_w = 0$ and $\gamma_w = \gamma_w^{\dagger}$.[2] We also note that as $\gamma_w \to \infty$ we have $h(\gamma_w) \to -\infty$. Then, we can conclude that $h(\gamma_w) \geq 0$ for $\gamma_w < \gamma_w^{\dagger}$ and $h(\gamma_w) \leq 0$ for $\gamma_w \geq \gamma_w^{\dagger}$. As such, noting $\gamma_w^2(1+\gamma_w)^2 \geq 0$ and following (26), we have $\partial g(\gamma_w)/\partial \gamma_w \geq 0$ for $\gamma_w < \gamma_w^{\dagger}$ and $\partial g(\gamma_w)/\partial \gamma_w \leq 0$ for $\gamma_w \geq \gamma_w^{\dagger}$. This indicates that $g(\gamma_w)$ initially increases and then decreases with γ_w. As per (24), we know that $n\gamma_w$ monotonically decreases with $g(\gamma_w)$, which leads to the fact that $n\gamma_w$ first decreases and then increases as γ_w increases (i.e., $n\gamma_w$ has one minimum value but no maximum value). We recall that n is a monotonically decreasing function of γ_w under the constraint (21), which is proved following (23). Therefore, we conclude that $n\gamma_w$ first decreases and then increases as n increases, and thus the maximum value of $n\gamma_w$ is achieved either at $n = 1$ or $n = N$.

Step 5: We next prove that $n = N$ (not $n = 1$) maximizes $n\gamma_w$. Substituting $\gamma_w^{\dagger}$ into (21), we have $n^{\dagger} = 2\epsilon^2/f(\gamma_w^{\dagger})$. For $0 < \epsilon < 0.4835$, we have $n^{\dagger} < 1$ due to $f(\gamma_w^{\dagger}) > 0.4675$. When $n^{\dagger} < 1$, $n\gamma_w$ increases with n due to $n \geq 1$. As such, for $0 < \epsilon < 0.4835$ the optimal value of n that maximizes $n\gamma_w$ is N (i.e., $n^* = N$). For $0.4835 \leq \epsilon \leq 0.5$, we have $n^{\dagger} < 2$ again due to $f(\gamma_w^{\dagger}) > 0.4675$. We next confirm that even for $n^{\dagger} < 2$ we still have $n^* = N$. To this end, we only have to confirm $n\gamma_w$ for $n = 2$ is larger than that for $n = 1$. When $n = 1$, following (21) we have $f(\gamma_w) = 2\epsilon^2$. The maximum value of γ_w that guarantees $f(\gamma_w) = 2\epsilon^2$ (i.e., $n = 1$) is obtained when $\epsilon = 0.5$ since $f(\gamma_w)$ is a monotonically increasing function of γ_w as proved by (23). We obtain this maximum value by solving $f(\gamma_w) = 0.5$ as $\gamma_w^{n=1} < 2.3145$, which leads to $n\gamma_w < 2.3145$ when $n = 1$. When $n = 2$, following (21) we have $f(\gamma_w) = \epsilon^2$. The minimum value of γ_w that guarantees $f(\gamma_w) = \epsilon^2$ (i.e., $n = 2$) is obtained when $\epsilon = 0.4835$. We obtain this minimum value by solving $f(\gamma_w) = (0.4835)^2$ as $\gamma_w^{n=2} > 1.16$, which leads to $n\gamma_w > 2.32$ when $n = 2$. As such, we have $n\gamma_w < 2.3145$ when $n = 1$ and $n\gamma_w > 2.32$ when $n = 2$, which results in $n\gamma_w$ for $n = 2$ is larger than $n\gamma_w$ for $n = 1$. We recall that $n\gamma_w$ monotonically increases with n when $n \geq n^{\dagger}$. Therefore, for $0.4835 \leq \epsilon \leq 0.5$ the optimal value of n that maximizes $n\gamma_w$ is N.

Step 6: So far, we have proved $n^* = N$. Then, substituting $n^* = N$ into (21), we obtain the fixed-point equation in (19).

References

[1] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, "Reliable, deniable and hidable communication: A quick survey," in *Proc. IEEE Inf. Theory Workshop*, Nov. 2014, pp. 227–231.

[2] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: Fundamental limits of covert wireless communication," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 26–31, Dec. 2015.

[3] B. He, S. Yan, X. Zhou, and V. Lau, "On covert communication with noise uncertainty," *IEEE Commun. Lett.*, accepted to appear, DOI: 10.1109/LCOMM.2016.2647716, Dec. 2016.

[4] M. K. Simon, Jim K. Omura, Robert A. Scholtz, and Barry K. Levitt, *Spread Spectrum Communications Handbook*, McGraw-Hill, 1994.

[5] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[6] L. Wang, G. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[7] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise, in *Proc. IEEE Int'l. Symp. Info. Theory*, Jul. 2013, pp. 2945–2949.

[8] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[9] S. Lee and R. Baxley, "Achieving positive rate with undetectable communication over AWGN and Rayleigh channels," in *Proc. IEEE ICC, 2014*, Jun. 2014, pp. 780–785.

[10] B. A. Bash, D. Goeckel, and D. Towsley, LPD communication when the warden does not know when, in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 606–610.

[11] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1195–1205, Oct. 2015.

[12] T. Sobers, B. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert communication with the help of an uninformed jammer achieves positive rate," in *Proc. Asilomar Conf. on Signals, Systems, and Comput.*, Nov. 2015, pp. 625–629.

[13] D. Goeckel, B. Bash, S. Guha, and D. Towsley, Covert communications when the warden does not know the background noise power, *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 236–239, Feb. 2016.

[14] Y. Polyanskiy, H. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[15] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823–2835, Aug. 2015.

[16] G. Ozcan andM. C. Gursoy, "Throughput of cognitive radio systems with finite blocklength codes," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2541–2554, Nov. 2013.

[17] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Hoboken, NJ, 2002.

[19] X. Zhou, M. R. McKay, B. Maham, and A. Hjørungnes, "Rethinking the secrecy outage formulation: A secure transmission design perspective," *IEEE Commun. Lett.*, vol. 15, no. 3, pp. 302–304, Mar. 2011.

[20] S. Yan, N. Yang, G. Geraci, R. Malaney, and J. Yuan, "Optimization of code rates in SISOME wiretap channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6377–6388, Nov. 2015.

[2] We obtain $\gamma_w^{\dagger} \approx 2.1626$ by numerically solving $h(\gamma_w) = 0$.