

Received 20 October 2022, accepted 16 November 2022, date of publication 18 November 2022,
date of current version 28 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223357

RESEARCH ARTICLE

Covert Communications Over Non-Orthogonal Multiple Overt Channels

HIEN QUANG TA^{1,3}, (Member, IEEE), KHUONG HO-VAN^{2,3},
DANIEL BENEVIDES DA COSTA⁴, (Senior Member, IEEE), SANG WU KIM⁵,
AND HOON OH⁶, (Member, IEEE)

¹School of Electrical Engineering, International University, Ho Chi Minh City 700000, Vietnam

²Department of Telecommunications Engineering, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City 700000, Vietnam

³Vietnam National University (Ho Chi Minh City), Thu Duc, Ho Chi Minh City 700000, Vietnam

⁴Technology Innovation Institute, Masdar City, Abu Dhabi, United Arab Emirates

⁵Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011, USA

⁶Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan 44610, South Korea

Corresponding author: Hoon Oh (hoonoh@ulsan.ac.kr)

This work was supported by the Institute of Information and Communication Technology Planning and Evaluation (IITP) Grant by the Korean Government (MSIT) (Development of 5G-based Shipbuilding and Marine Smart Communication Platform and Convergence Service) under Grant 2020-0-00869.

ABSTRACT The paper exploits overt information of non-orthogonal multiple access (NOMA) systems for camouflage to hide covert information. The optimum strategy of hiding information at the transmitter, accounting for the probability of the warden's detection error and the resulting spectral efficiency of the covert information under the covertness requirement, is determined. The results show that the optimum strategy of hiding the covert information is to superimpose it onto the overt NOMA information which is hardest-decoded, increasing to 1 the probability of detection error at the warden when the number of NOMA users accretes. This indicates that the multiplicity of NOMA users can be leveraged to hide the covert information in NOMA systems. In addition, various results prove the increase in the covert spectral efficiency with increasing the number of NOMA users while guaranteeing the probability of the warden's detection error close to 1.

INDEX TERMS Covert communication, non-orthogonal multiple access, probability of detection.

I. INTRODUCTION

As a promising candidate to utilize and share spectrum efficiently among users in future networks, non-orthogonal multiple access (NOMA) has demonstrated prospective gains over the single-antenna [1], [2] and multi-antenna [3] orthogonal multiple access (OMA). Under security requirements and relying on physical layer security (PLS) perspective, several papers have revealed that NOMA-based systems outperform its OMA counterpart in terms of secrecy sum rate [4], [5], [6], [7].

Recently, covert communication has received much attention due to its crucial role in improving user privacy by hiding transmission towards a warden. It has found impor-

tant applications ranging from military to commercial wireless networks. The authors in [8] pioneeringly studied the theoretical limit of covert communications by considering square root law (SRL) tools: the transmitter can send covertly and reliably $\mathcal{O}(\sqrt{U})$ bits over U channels. This means that $\mathcal{O}(\sqrt{U})/U$ tends to 0 as U goes to infinity. Then, [9], [10], and [11] expanded the SRL for discrete memoryless channels (DMCs). In [9], [10], and [11], the Big-O notation completely characterized the constant hidden. Additionally, [12] studied covert communications in DMCs from the perspective of the second order asymptote subject to distinct covertness constraints. Owing to the asymmetry property of Kullback-Leibler divergence, [13] proved the Gaussian signalling to be optimal for one covertness constraint yet for another. Several efforts have been strived to surpass the SRL in the manner that a positive covert rate is achievable for infinite number

The associate editor coordinating the review of this manuscript and approving it for publication was Yan Huo.

of channel uses or $\mathcal{O}(U)$ instead of $\mathcal{O}(\sqrt{U})$ bits can be sent reliably and covertly in U channel uses. To guarantee a positive covert rate, the noise uncertainty at the warden was exploited in [14], [15], and [16]. More specifically, the signal transmission was proved not to be detected if the detector's signal-to-noise ratio (SNR) does not exceed a threshold, a.k.a. the SNR wall in [16]. This is because the estimated noise power does not agree with the real one, so called *noise uncertainty*. The positive covert rate probability was analyzed theoretically in [15] as the adversary suffers noise uncertainty. Then, [14] analyzed the covert throughput under two practical noise uncertainty paradigms in additive white Gaussian noise (AWGN) channels. The core idea is to exploit the warden's noise uncertainty to supply covertness and achieve a non-zero covert rate. Jamming [17], [18], interference [19], [20] and artificial noise [21], [22] are applicable sources of noise uncertainty. Other uncertainties at the warden were also exploited, such as the uncertainty of transmission time [23], that of transmit power [24], and that of channel [25], [26] as an additional factor, along with the noise uncertainty, to improve covertness and increase the covert rate.

Most recently, superimposing (embedding) the covert information into an overt transmission has been developed from an information theoretic viewpoint in the broadcast channel [27] and in the multiple access channels [28], which was inspired by [29] wherein information transmission is hidden by exploiting hardware imperfection (dirty constellation). Then, [30] investigated superimposing the covert information into the overt information at the relay in cooperative networks with amplify-and-forward relay. The covertness has been also studied in the NOMA networks with two-users and exploiting the warden's uncertainties of random transmit powers [31] or employing intelligent reflecting surfaces [32]. The idea of hiding information into overt transmissions was to exploit existing (overt) transmissions combined with warden's uncertainties to enlarge the dynamic range for covert transmissions.

In literature, covert communication can be categorized into two mainstreams: hiding information in noise [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [21], [22], [23], [24], [25], [26] and hiding information in existing transmissions [27], [28], [29], [30], [31], [32]. The former focuses on sending only the covert information and mainly exploiting the warden's uncertainties, such as noise and channel, which depends on the warden's hardware performance and cannot be predicted or unknown to the transmitter. Thus, it is impossible for the transmitter to design and control the practical covert transmissions since the warden's uncertainties are unknown to the transmitter. On the other hand, hiding information in existing transmissions can significantly increase the covert throughput and can control the covertness requirement since the uncertainty of the existing transmissions at the warden, i.e. whether the covert information is hidden into them, can be controlled by the transmitter. Although covertness has been considered in NOMA systems [30], [31], [32], its analysis relies on the energy detection assumption at the warden,

which is not optimal especially when the warden succeeds in decoding the overt message and knows its codeword [33]. As a result, the covertness in NOMA systems has not been quantified accurately.

Therefore, this paper considers hiding information in multiple overt transmissions of NOMA systems with multiple users and assuming that the CSI of existing transmissions is publicly known. It is determined the optimum strategy of the warden to optimize its detection probability and that of the transmitter to induce maximum detection error at the warden and the spectral efficiency of the covert information under the covertness requirement. The results prove that the covert information can be decoded only if all overt information wherein the covert information is superposed onto, can be restored. This can be referred to conceal the covert information by exploiting the *decoding uncertainty* of overt information at the warden and suggests to superimpose it onto the overt information which is hardest to be decoded. The results also show that it can guarantee the maximum probability of the warden's detection error close 1 for sufficiently large number of NOMA information. Therefore, the multiplicity of NOMA overt information can be leveraged to hide information. Then, the covert spectral efficiency increases, in trade-off of the high loss of the overt spectral efficiency owing to covert communications, as the number of transmit antennas or NOMA users accretes.

This paper contributes the following:

- A paradigm to conceal the covert information by utilizing the multiplicity of NOMA overt users is proposed.
- The optimum strategy of transmission at the transmitter to minimize the detection probability of the warden is determined. By leveraging the multiplicity of NOMA overt users without requiring any warden's uncertainty, such as noise or channel gain, this strategy can achieve an increasing-to-1 detection error probability at the warden. This finding is under practical consideration where the transmitter can base on the NOMA network parameters to control and design its covert transmission.
- The warden's detection error probability, the covert spectral efficiency and the loss of the overt spectral efficiency for the covert transmission are analyzed and characterized with general number of NOMA users.
- A non-zero covert rate can be obtained whilst guaranteeing the decoding error probability at the warden close to 1. This can efficiently achieved by exploiting the existing NOMA network with the multiplicity of overt users.

The next part of this paper presents the system model. Then, the decoding strategy of the warden, along with the optimum decoding threshold to optimize its detection probability, are described in Part III. Next, Part IV determines the best cover set chosen by the transmitter to minimize the warden's detection probability, while Part V presents the derivation of the decoding error probability at the warden. Subsequently, Part VI derives the detection outage probability

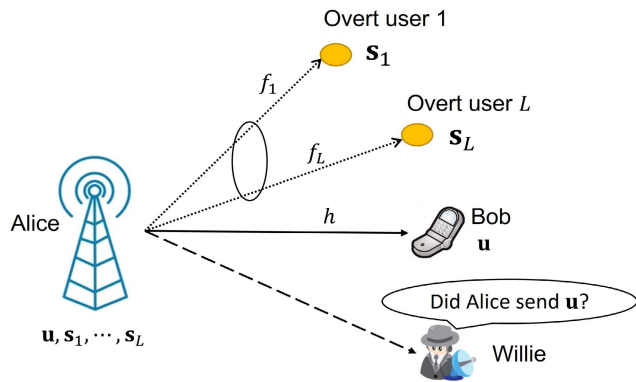


FIGURE 1. System setup - Alice hides the covert information u toward Willie by using overt NOMA information s_1, \dots, s_L as a camouflage.

of the covert information. The spectral efficiencies of the overt and covert information and the spectral efficiency loss of the overt information for the covert information are derived in Part VII. Part VIII extends the analysis for the multi-antenna transmitter case, whilst Part IX presents numerous results along with insightful discussions. Eventually, Part X makes the paper conclusion and mentions the future work.

II. COVERT COMMUNICATIONS IN NOMA SYSTEMS

Fig. 1 illustrates a downlink NOMA system composed of L NOMA information s_1, \dots, s_L , where $s_l = (s_{1,l}, \dots, s_{n,l})$, $l \in [1, L]$, of L overt users and the transmitter (Alice) opportunistically transmits covert information $u = (u_1, \dots, u_n)$ to a covert user (Bob) toward a warden (Willie), who is not to decode the covert information, but to detect if any information other than NOMA information or the transmission of the covert information. The random coding argument is employed to generate codewords [27]. That is, L overt codewords s_l for $1 \leq l \leq L$ and one covert codeword u are produced by taking symbols independently from a complex Gaussian distribution with zero mean and variance of P_l and $\sum_{l \in \mathcal{S}} P_l$, respectively, where $\sum_{l=1}^L P_l = P$. The designed specifications of the system, i.e. $P_l, \sum_{l \in \mathcal{S}} P_l$ and ϕ are supposed to be available at all receivers including Willie. It is also supposed that the codebooks for overt information are publicly known while that for the covert information is secretly shared between Bob and Alice, and unknown to Willie.

Considering superimposing u onto a set, denoted as \mathcal{S} , of some overt NOMA information where $|\mathcal{S}| \leq L$ and its complement is denoted as \mathcal{S}^c . For example, $\mathcal{S} = \{1, 5, 9\}$ means that u is superimposed onto s_1, s_5 , and s_9 . The optimum cover set \mathcal{S} , or the set of overt messages where the covert message is superimposed onto, to maximize the warden’s detection error will be discussed in Section IV. Since Willie does not know whether Alice transmits the covert message or not, the transmitted signal is viewed under two hypotheses of no transmission of u , denoted as H_0 , and the alternative,

TABLE 1. Frequently-used notations.

The transmit rate of the covert information	R_u
The channel gain at the j -th overt user	f_j
Number of NOMA overt users or information	L
Number of Alice’s transmit antennas	M
Number of Willie’s antennas	N
The j -th overt information	s_j
Covert information	u
Cover set	\mathcal{S}
The transmit rate of the j -th overt information	R_j
Decoding set at Willie	\mathcal{D}
Power of the l -th overt information	P_l
Power of the covert information	$\sum_{l \in \mathcal{S}} P_l$
Power ratio allocated to overt information	ϕ
The channel gain variance at Willie	σ_g^2
The channel gain variance at the covert user (Bob)	σ_b^2
The channel gain variance at the overt user	σ_f^2
The channel gain at Willie	g
The channel gain at the covert user (Bob)	h
The maximum detection error probability	ξ_{max}
Covert spectral efficiency	η_u
Overt spectral efficiency	η_s
The loss of overt spectral efficiency	$\eta_{s,loss}$
Covertness requirement	ϵ
Received signal at the j -th overt user	y_j
Received signal at Willie	y_w
Received signal at the covert user (Bob)	y_b
The background noise variance	σ_n^2

denoted as H_1 ,

$$x = \begin{cases} \sum_{l=1}^L s_l, & H_0, \\ \sqrt{\phi} \sum_{l \in \mathcal{S}} s_l + \sqrt{1-\phi}u + \sum_{l \in \mathcal{S}^c} s_l, & H_1, \end{cases} \quad (1)$$

where $\phi \in (0, 1)$ is the power ratio allotted to the overt information. It should be noted that, under both hypotheses H_0 and H_1 , all NOMA messages are still transmitted to avoid being detected by the warden and that the total transmit power is identical.

All nodes are assumed to have single antenna and denote the fading channel from Alice to the j -th overt user, Bob and Willie, as f_j, h and g , respectively, which are independent and identically distributed (i.i.d) complex Gaussian random variables with zero mean and variance of σ_f^2, σ_h^2 , and σ_g^2 , respectively. The assumption of i.i.d. channel gains have practical appeal in scattered networks [30], [31], [33]. The multi-antenna scenario at the warden and transmitter will be extended in Sections V and VIII, respectively. Assuming the forward channel training (Alice sends a pilot based on which the overt users estimates the channel gains and feedback to Alice), then Alice perfectly knows $f_j, j = 1, \dots, L$. However, Bob does not feedback h to Alice to prevent Willie from detecting his presence and thence, h is unknown to Alice. The channel gain at the overt users, without loss of generality, are assumed in descending order, i.e. $|f_1|^2 \geq |f_2|^2 \geq \dots \geq |f_L|^2$. We denote $n_x = \{n_{1,x}, \dots, n_{n,x}\} \sim CN(0, \sigma_n^2 I)^1$ as a noise

¹ $CN(0, \sigma^2 I)$ denotes the complex Gaussian noise vector that each variable is Gaussian-distributed with zero mean and variance σ^2 .

vector at the receiver x for $x = \{j, w, b\}$, where j , w and b indicates the j -th overt user, Willie and Bob, respectively. For compactness, Table 1 lists frequently-used notations.

III. WARDEN'S DETECTION

This part describes the detection strategy at the warden and derives the optimum detection threshold to maximize the warden's detection probability as well as the resulting minimum detection error probability. It follows from (1) that Willie receives the signal to be

$$y_w = \begin{cases} g \sum_{l=1}^L s_l + n_w, & H_0, \\ g(\sqrt{\phi} \sum_{l \in \mathcal{S}} s_l + \sqrt{1-\phi}u + \sum_{l \in \mathcal{S}^c} s_l) + n_w, & H_1. \end{cases} \quad (2)$$

Due to the forward channel training at Alice, Willie is assumed to perfectly know its channel gain g . Since Willie has to decide H_0 or H_1 , regarding the transmission of u from his observation of y_w with the length of n , it applies the optimum method of the likelihood ratio test (LRT) for its detection to maximize the detection probability.

If assuming that Willie knows the overt information $\{s_l\}$, the LRT test or the optimum detection at Willie is given by [17]

$$\frac{\Pr(y_w|g, \{s_l\}, H_1)}{\Pr(y_w|g, \{s_l\}, H_0)} \underset{H_1}{\overset{H_0}{>}} \delta, \quad (3)$$

where $\Pr(y_w|g, \{s_l\}, H_0)$ and $\Pr(y_w|g, \{s_l\}, H_1)$, when u in (2) is treated as noise because the codebook of u is unknown to Willie, are given by [33]

$$\Pr(y_w|g, \{s_l\}, H_0) = \frac{e^{-\|y_w - g \sum_{l=1}^L s_l\|^2 / \sigma_n^2}}{(\pi \sigma_n^2)^n}, \quad (4)$$

and

$$\Pr(y_w|g, u, \{s_l\}, H_1) = \frac{e^{-\frac{\|y_w - g(\sqrt{\phi} \sum_{l \in \mathcal{S}} s_l + \sum_{l \in \mathcal{S}^c} s_l)\|^2}{\sigma_n^2 + (1-\phi)|g|^2 \sum_{l \in \mathcal{S}} P_l}}}{(\pi(\sigma_n^2 + (1-\phi)|g|^2 \sum_{l \in \mathcal{S}} P_l))^n}, \quad (5)$$

respectively, and δ is a detection threshold.

It should be noted that the optimum detection in (3) depends on Willie's knowing all overt information. Then, it is necessary to consider another statistic test, rather than (3). In the next subsection, we will describe the detail of warden's detection strategy and its optimum detection. Then, we will prove that, to detect the covert information, the warden must decode all overt information where the covert information is superimposed onto.

A. DETECTION STRATEGY

For convenience, we notate the set of indices of the overt NOMA information which Willie decodes after the successive interference cancellation (SIC) is finished, as \mathcal{D} and its complement as \mathcal{D}^c . Since Willie does not know $s_{l \in \mathcal{D}^c}$

but their codebook (their distribution) is known, he can produce a series of likely codewords for $s_{l \in \mathcal{D}^c}$ and perform the marginalized LRT as the optimal detection [33], [34],

$$\Delta := \frac{E_{s_{l \in \mathcal{D}^c}}[\Pr(y_w|g, \{s_l\}, H_1)]}{E_{s_{l \in \mathcal{D}^c}}[\Pr(y_w|g, \{s_l\}, H_0)]} \underset{H_1}{\overset{H_0}{>}} \delta, \quad (6)$$

where $E_H[\cdot]$ notates the expectation operator with respect to the random variable H ,

$$E_{s_{l \in \mathcal{D}^c}}[\Pr(y_w|g, \{s_l\}, H_0)] = \frac{e^{-\|y_w - g \sum_{l \in \mathcal{D}} s_l\|^2 / \sigma_0^2}}{(\pi \sigma_0^2)^n}, \quad (7)$$

and

$$E_{s_{l \in \mathcal{D}^c}}[\Pr(y_w|g, \{s_l\}, H_1)] = \left(\frac{1}{\pi \sigma_1^2}\right)^n e^{-\|y_w - g(\sqrt{\phi} \sum_{l \in \mathcal{S} \cap \mathcal{D}} s_l + \sum_{l \in \mathcal{S}^c \cap \mathcal{D}} s_l)\|^2 / \sigma_1^2} \quad (8)$$

with $\sigma_0^2 = \sigma_n^2 + |g|^2 \sum_{l \in \mathcal{D}^c} P_l$ and $\sigma_1^2 = \sigma_n^2 + |g|^2(\phi \sum_{l \in \mathcal{S} \cap \mathcal{D}^c} P_l + (1-\phi) \sum_{l \in \mathcal{S}} P_l + \sum_{l \in \mathcal{S}^c \cap \mathcal{D}^c} P_l)$. The proof of (7) and (8) are provided in Appendix A.

Different from previous studies where the overt information is supposed to be available at the detector [30], [31], [32] and the detection error mainly relies on the LRT in (3), our work considers unknown overt information and thus, Willie should perform the optimum marginalized LRT as in (6), i.e. averaging over unknown overt information $s_{l \in \mathcal{D}^c}$. To determine the probability of detection error at the warden, the following considers two scenarios that Willie decodes unsuccessfully all $s_{l \in \mathcal{S}}$ ($\mathcal{S} \cap \mathcal{D} = \emptyset$) and Willie succeeds in decoding at least one $s_{l \in \mathcal{S}}$ ($\mathcal{S} \cap \mathcal{D} \neq \emptyset$).

1) IF $\mathcal{S} \cap \mathcal{D} = \emptyset$, i.e. WILLIE FAILS TO DECODE ALL $s_{l \in \mathcal{S}}$
 Since $\mathcal{S} \cap \mathcal{D} = \emptyset$, we obtain $\sum_{l \in \mathcal{S} \cap \mathcal{D}^c} P_l = \sum_{l \in \mathcal{S}} P_l$ and $\sum_{l \in \mathcal{S}^c \cap \mathcal{D}^c} P_l = \sum_{l=1}^L P_l - \sum_{l \in \mathcal{S}} P_l - \sum_{l \in \mathcal{D}} P_l$, which yields $\sigma_1^2 = \sigma_0^2$. It follows from (7) and (8) that

$$\begin{aligned} E_{s_{l \in \mathcal{D}^c}}[\Pr(y_w|g, \{s_l\}, H_1)] &= \frac{e^{-\|y_w - g \sum_{l \in \mathcal{S}^c \cap \mathcal{D}} s_l\|^2 / \sigma_0^2}}{(\pi \sigma_0^2)^n} \\ &= \frac{e^{-\|y_w - g \sum_{l \in \mathcal{D}} s_l\|^2 / \sigma_0^2}}{(\pi \sigma_0^2)^n} \\ &= E_{s_{l \in \mathcal{D}^c}}[\Pr(y_w|g, \{s_l\}, H_0)], \end{aligned} \quad (9)$$

which, from (6), yields $\Delta = 1$. Then, the missed detection probability, $P_M := \Pr(\Delta \leq \delta | H_1)$, and the false alarm probability, $P_F := \Pr(\Delta > \delta | H_0)$, are $(P_F, P_M) = (0, 1)$ or $(P_F, P_M) = (1, 0)$. Hence, the sum of the missed detection and false alarm probabilities reduces to

$$P_M + P_F = 1 \quad (10)$$

for all δ . This means that Willie is unable to detect the transmission of u if he fails to decode all $s_{l \in \mathcal{S}}$.

2) IF $\mathcal{S} \cap \mathcal{D} \neq \emptyset$, I.e. WILLIE SUCCEEDS IN DECODING AT LEAST ONE $s_{l \in \mathcal{S}}$

Taking the natural logarithm on Δ in (6) and dividing it by n , we obtain

$$\Delta' := \frac{1}{n} \ln \Delta = \frac{1}{n} \ln \left(\frac{E_{s_{l \in \mathcal{D}^c}} [\Pr(y_w | g, \{s_l\}, H_1)]}{E_{s_{l \in \mathcal{D}^c}} [\Pr(y_w | g, \{s_l\}, H_0)]} \right), \quad (11)$$

which, from (7) and (8), yields

$$\Delta' = \ln \left(\frac{\sigma_0^2}{\sigma_1^2} \right) + \frac{1}{n} \frac{\|y_w - g \sum_{l \in \mathcal{D}} s_l\|^2}{\sigma_0^2} - \frac{1}{n} \frac{\|y_w - g(\sqrt{\phi} \sum_{l \in \mathcal{S} \cap \mathcal{D}} s_l + \sum_{l \in \mathcal{S}^c \cap \mathcal{D}} s_l)\|^2}{\sigma_1^2}. \quad (12)$$

As $n \rightarrow \infty$, Δ' converges to

$$\begin{aligned} \Delta' &\rightarrow \ln \left(\frac{\sigma_0^2}{\sigma_1^2} \right) + \frac{E[\|y_{w,i} - g \sum_{l \in \mathcal{D}} s_{l,i}\|^2]}{\sigma_0^2} \\ &\quad - \frac{E[\|y_{w,i} - g(\sqrt{\phi} \sum_{l \in \mathcal{S} \cap \mathcal{D}} s_{l,i} + \sum_{l \in \mathcal{S}^c \cap \mathcal{D}} s_{l,i})\|^2]}{\sigma_1^2} \\ &= \begin{cases} \ln(\sigma_0^2/\sigma_1^2) + \tau_0, & H_0, \\ \ln(\sigma_0^2/\sigma_1^2) + \tau_1, & H_1, \end{cases} \end{aligned} \quad (13)$$

where $\tau_0 = 1 - (\sigma_0^2 + |g|^2(1 - \sqrt{\phi})^2 \sum_{l \in \mathcal{S} \cap \mathcal{D}} P_l) / \sigma_1^2$ and $\tau_1 = (\sigma_1^2 + |g|^2(1 - \sqrt{\phi})^2 \sum_{l \in \mathcal{S} \cap \mathcal{D}} P_l) / \sigma_0^2 - 1$. Then, the entire detection error probability reduces to

$$\begin{aligned} P_M + P_F &= \Pr(\Delta' \leq \delta' | H_1) + \Pr(\Delta' > \delta' | H_0) \\ &\rightarrow 0 \end{aligned} \quad (14)$$

by adopting the detection threshold $\delta' \in (\ln(\sigma_1^2/\sigma_0^2) + \tau_0, \ln(\sigma_1^2/\sigma_0^2) + \tau_1)$.

Remark 1: If any $s_{l \in \mathcal{S}}$ is decoded successfully by Willie, the covert information u can be detected perfectly for sufficiently large n , i.e. Willie is permitted to inspect the received signal in a long duration. This is equivalent to that if Willie fails to decode all $s_{l \in \mathcal{S}}$, then it is unable to detect the transmission of u .

B. OPTIMUM DETECTION THRESHOLD FOR WILLIE

Willie aims at minimizing the entire detection error probability by adopting properly the detection thresholds δ' and δ . The entire detection error probability, averaging over two scenarios that Willie decodes successfully at least one $s_{l \in \mathcal{S}}$ and Willie decodes unsuccessfully all $s_{l \in \mathcal{S}}$, can be expressed as

$$\begin{aligned} P_M + P_F &= \Pr(\Delta' \leq \delta', \mathcal{S} \cap \mathcal{D} \neq \emptyset | H_1) \\ &\quad + \Pr(\Delta \leq \delta, \mathcal{S} \cap \mathcal{D} = \emptyset | H_1) \\ &\quad + \Pr(\Delta' > \delta', \mathcal{S} \cap \mathcal{D} \neq \emptyset | H_0) \\ &\quad + \Pr(\Delta > \delta, \mathcal{S} \cap \mathcal{D} = \emptyset | H_0). \end{aligned} \quad (15)$$

For $\mathcal{S} \cap \mathcal{D} \neq \emptyset$, Willie can choose $\delta' \in (\ln(\sigma_1^2/\sigma_0^2) + \tau_0, \ln(\sigma_1^2/\sigma_0^2) + \tau_1)$ to induce the second and fourth terms in (15) to be zero, which reduces (15) to

$$P_M + P_F = \Pr(\Delta \leq \delta, \mathcal{S} \cap \mathcal{D} = \emptyset | H_1)$$

$$\begin{aligned} &+ \Pr(\Delta > \delta, \mathcal{S} \cap \mathcal{D} = \emptyset | H_0) \\ &= \Pr(1 \leq \delta, \mathcal{S} \cap \mathcal{D} = \emptyset | H_1) \\ &\quad + \Pr(1 > \delta, \mathcal{S} \cap \mathcal{D} = \emptyset | H_0), \end{aligned} \quad (16)$$

or

$$P_F + P_M = \min_{\delta} \{ \Pr(\mathcal{S} \cap \mathcal{D} = \emptyset | H_0), \Pr(\mathcal{S} \cap \mathcal{D} = \emptyset | H_1) \} \quad (17)$$

which (16) is derived due to $\Delta = 1$ for $\mathcal{S} \cap \mathcal{D} = \emptyset$ as shown in Section III-A1.

It follows from (2) where the signal to interference plus noise ratio (SINR) for $s_{l \in \mathcal{S}}$ under H_1 is reduced by a factor of ϕ compared to that under H_0 , we have $\Pr(\mathcal{S} \cap \mathcal{D} = \emptyset | H_0) < \Pr(\mathcal{S} \cap \mathcal{D} = \emptyset | H_1)$. Hence, Willie should choose $\delta < 1$ to minimize $P_M + P_F$ and then, the resulting minimum entire detection error probability is obtained as

$$\begin{aligned} \xi_{min} &:= \Pr(\mathcal{S} \cap \mathcal{D} = \emptyset | H_0) \\ &= \Pr \left(\bigcap_{j \in \mathcal{S}} I(s_j; y_{w,j}) < R_j \middle| H_0 \right), \end{aligned} \quad (18)$$

where R_j is the transmission rate of the j -th overt information,

$$y_{w,j} = y_w - g \sum_{l \in \mathcal{D}_j} s_l = g \sum_{l \in \mathcal{D}_j^c} s_l + n_w \quad (19)$$

is the received signal after successively cancelling s_l , $l \in \mathcal{D}_j$, and \mathcal{D}_j is the decoding set before decoding s_j . One should note that the minimum detection error probability in (18) represents the best detection capability at the warden and depends on the cover set \mathcal{S} or the transmitter's choice of which set of overt information wherein the covert information is superposed onto. Relied on this, the next following section then considers the optimum transmission strategy at the transmitter to minimize the warden's detection probability by properly choosing the optimum cover set \mathcal{S} .

IV. OPTIMUM COVER SET \mathcal{S} FOR ALICE

In this section, we determine the optimum cover set chosen by Alice for maximizing the minimum detection error possibility ξ_{min} and then, derive the detection error probability, denoted as ξ_{max} , when this optimum cover set is chosen. Alice aims to maximize ξ_{min} by finding the optimum cover set \mathcal{S} . One can obtain from (18) that the optimum $|\mathcal{S}|$ to maximize ξ_{min} is 1, meaning that the covert information should be superposed onto one overt information. Then, the resulting maximum entire detection error probability is expressed as

$$\begin{aligned} \xi_{max} &:= \max_{\mathcal{S}} \min_{\delta} P_F + P_M \\ &= \max_{j \in \mathcal{S}} \Pr(I(s_j; y_{w,j}) < R_j | H_0), \end{aligned} \quad (20)$$

where $|\mathcal{S}| = 1$ and

$$I(s_j; y_{w,j}) = \log_2 \left(1 + \frac{|g|^2 \gamma_j}{1 + |g|^2 (\sum_{l \in \mathcal{D}_j^c} \gamma_l - \gamma_j)} \right) \quad (21)$$

under H_0 , and $\gamma_j = P_j / \sigma_n^2$.

Since $|\mathcal{S}| = 1$ and $j \in \mathcal{S}$, $\mathcal{S} = \{j\}$. This indicates that the covert information should be superimposed onto only one overt NOMA information. Next, we determine the overt transmission rate R_j in NOMA systems and then, which overt NOMA information should be chosen for the covert information to superimpose onto in order for the entire detection error probability to be maximized. Before deriving the expression of ξ_{max} of (20), we describe the design of L NOMA overt information to guarantee that all overt users can decode their own information regardless of the transmission of the covert information.

Overt Transmission: Assuming the SIC is applied by each user, i.e., the k -th user will firstly recover s_m , $m > k$, and then remove them from y_k in a successive manner [2]. The remaining signals s_m , $m < k$, are considered as noise. It follows from (1) that, when $\mathcal{S} = \{j\}$, the j -th overt user receives the signal under two hypotheses H_0 and H_1 to be

$$y_j = \begin{cases} f_j \sum_{l=1}^L s_l + n_w, & H_0, \\ f_j(\sqrt{\phi}s_j + \sqrt{1-\phi}u + \sum_{l \neq j} s_l) + n_j, & H_1. \end{cases} \quad (22)$$

One can see that R_j under H_0 is higher than that under H_1 due to additional interference of the covert information u under H_1 . Then, R_j should be determined under H_1 to guarantee that all overt users can decode their own information regardless of the transmission of the covert information. Hence, the j -th user after suppressing s_l , $j < l$, from y_j obtains

$$\begin{aligned} \hat{y}_j &= y_j - f_j \sum_{l>j} s_l \\ &= f_j \left(\sqrt{\phi}s_j + \sqrt{1-\phi}u + \sum_{l<j} s_l \right) + n_j. \end{aligned} \quad (23)$$

As a result, the maximum rate of s_j under H_1 is expressed as

$$R_j = \log_2 \left(1 + \frac{|f_j|^2 \phi \gamma_j}{1 + |f_j|^2 (\sum_{l<j} \gamma_l + (1-\phi)\gamma_j)} \right). \quad (24)$$

Then, ξ_{max} in (20) is obtained from (21) and (24) by

$$\begin{aligned} \xi_{max} &= \max_j \Pr \left(|f_j|^2 > \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l \in \mathcal{D}_j^c} \gamma_l - \sum_{l=1}^j \gamma_l))^+} \right), \end{aligned} \quad (25)$$

where $(x)^+ = \max\{0, x\}$. When g and \mathcal{D}_j^c are unavailable at Alice, the maximum of (25) is obtained by minimizing $\sum_{l=1}^j \gamma_l$ and maximizing $|f_j|^2$. Because $\sum_{l=1}^j \gamma_l$ reduces and $|f_j|^2$ increases with declining j , the maximum of (25) is attained as $j = 1$, viz. superposing u onto s_1 , yielding

$$\xi_{max} = \Pr \left(|f_1|^2 > \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+} \right). \quad (26)$$

Remark 2: The best hiding strategy for Alice to maximize the entire detection error probability is to superpose the covert

information u onto the overt information s_1 that experiences the highest channel gain hence can be transmitted with the highest data speed.

V. MAXIMUM ENTIRE DETECTION ERROR PROBABILITY

This part evaluates the maximum entire detection error probability, ξ_{max} and derives it in closed-form expression. The challenge is that \mathcal{D}_1^c depends on $|g|^2$. Based on the law of total probability, ξ_{max} is derived as

$$\begin{aligned} \xi_{max} &= 1 - \sum_{k=0}^L \Pr \left(|f_{k+1}|^2 \leq |g|^2 < |f_k|^2, \right. \\ &\quad \left. |f_1|^2 \leq \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+} \right), \end{aligned} \quad (27)$$

where $|f_0|^2 = \infty$ and $|f_{L+1}|^2 = 0$. To compute ξ_{max} , the probability in (27) is computed for the case of $k = 0$ and $1 \leq k \leq L$ as follows.

i) $k = 0$: For $|f_1|^2 \leq |g|^2$, Willie can apply SIC to decode s_2, \dots, s_L since $R_{j,1} = \log_2(1 + |f_j|^2 \gamma_j / (1 + |f_j|^2 \sum_{l<j} \gamma_l))$ and $I(s_j, y_{w,j}) = \log_2(1 + |g|^2 \gamma_j / (1 + |g|^2 \sum_{l<j} \gamma_l))$, for $j > 1$, after cancelling s_l , $l > j$, from y_w . Hence $\mathcal{D}_1^c = \{1\}$. Since $|g|^2 / (\phi - |g|^2(1-\phi)\gamma_1)^+ > |g|^2$, we obtain

$$\begin{aligned} \Pr \left(|f_1|^2 \leq \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1)\gamma_1)^+}, |f_1|^2 \leq |g|^2 \right) &= \Pr \left(|f_1|^2 \leq \frac{|g|^2}{(\phi - |g|^2(1-\phi)\gamma_1)^+}, |f_1|^2 \leq |g|^2 \right) \\ &= \Pr(|f_1|^2 \leq |g|^2) \\ &= \Pr \left(\bigcap_{l=1}^L \{|f_l|^2 \leq |g|^2\} \right) \\ &= \int_0^\infty \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \\ &= \frac{\sigma_f^2}{\sigma_g^2} B \left(\frac{\sigma_f^2}{\sigma_g^2}, L + 1 \right), \end{aligned} \quad (28)$$

where the Beta function [35] is notated as $B(p, q) = \int_0^1 z^{p-1}(1-z)^{q-1} dz = \int_0^\infty (1-e^{-y})^{q-1} (e^{-y})^p dy$.

ii) $1 \leq k \leq L$: For $|f_k|^2 > |g|^2 \geq |f_{k+1}|^2$, Willie can also apply SIC to decode s_{k+1}, \dots, s_L , hence $\mathcal{D}_1^c = \{1, \dots, k\}$. Let

$$\nu_k(|g|^2) = \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l=1}^k \gamma_l - \gamma_1))^+}. \quad (29)$$

Since $\Pr(\nu_k(|g|^2) \geq |f_1|^2, |f_k|^2 > |g|^2 \geq |f_{k+1}|^2) = 0$ if $\nu_k(|g|^2) \leq |g|^2$ or equivalently $\phi_k \leq |g|^2$, wherein

$$\phi_k = \frac{1-\phi}{(\phi \sum_{l=1}^k \gamma_l - \gamma_1)^+}, \quad (30)$$

we obtain

$$\Pr \left(|f_k|^2 > |g|^2 \geq |f_{k+1}|^2, \right)$$

$$\begin{aligned}
 & \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+} \geq |f_1|^2) \\
 &= \Pr(|f_k|^2 > |g|^2 \geq |f_{k+1}|^2, v_k(|g|^2) \geq |f_1|^2) \\
 &= \int_0^{\phi k} \Pr(|f_k|^2 > x \geq |f_{k+1}|^2, v_k(x) \geq |f_1|^2) \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \\
 &= \int_0^{\phi k} \binom{L}{k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{v_k(x)}{\sigma_f^2}}\right)^k \frac{e^{-\frac{x}{\sigma_g^2}}}{\sigma_g^2} dx,
 \end{aligned} \tag{31}$$

where (31) can be derived from [36, Eq. (2.2.13)].

Based on (27), (28) and (31), the maximum entire detection error probability is expressed as

$$\begin{aligned}
 & \xi_{max} \\
 &= 1 - \frac{\sigma_f^2}{\sigma_g^2} B\left(\frac{\sigma_f^2}{\sigma_g^2}, L+1\right) - \sum_{k=1}^L \binom{L}{k} \\
 & \times \int_0^{\phi k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{v_k(x)}{\sigma_f^2}}\right)^k f_{|g|^2}(x) dx
 \end{aligned} \tag{32}$$

where

$$f_{|g|^2}(x) = \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} \tag{33}$$

is the probability density function (PDF) of the random variable $|g|^2$. Note that, the PDF in (33) will be changed to $\frac{x^{N-1} e^{-x/\sigma_g^2}}{(N-1)! \sigma_g^{2N}}$ if Willie has N antennas rather than single one.

Remark 3: $\xi_{max} \rightarrow 1$ as $L \rightarrow \infty$. That is, if L is sufficiently high, then the transmission of u is undetectable. Appendix B provides the proof. This result indicates that the multiplicity of overt users in the scattered NOMA network can be leveraged to hide the covert information. This is under practical consideration in which no uncertainty, such as noise or channel gain, at the warden is required and can be controlled by the transmitter for the case when the warden is a node in the network.

Remark 4: If $\mathcal{D}_1^c = \{1\}$, namely Willie can decode all information prior to decoding s_1 , which is adopted in the secrecy analysis of NOMA [6], [7], we obtain from (27) that

$$\begin{aligned}
 \xi_{max} &= 1 - \Pr\left(\frac{|g|^2}{(\phi + |g|^2(\phi - 1)\gamma_1)^+} \geq |f_1|^2\right) \\
 &= 1 - \int_0^\infty \left(1 - e^{-\frac{x}{\sigma_f^2(\phi + x(\phi - 1)\gamma_1)^+}}\right)^L f_{|g|^2}(x) dx.
 \end{aligned} \tag{34}$$

Fig. 2 compares ξ_{max} evaluated using (32) and (34). One can see that the difference is small at low SNR (P/σ_n^2) but it becomes larger for larger SNR, indicating that the simplifying assumption of $\mathcal{D}_1^c = \{1\}$ is valid only at low SNR.

Covertness requirement: if $\xi_{max} \geq 1 - \epsilon$ for any $\epsilon > 0$, where ϵ represents the covertness requirement [14], [17],

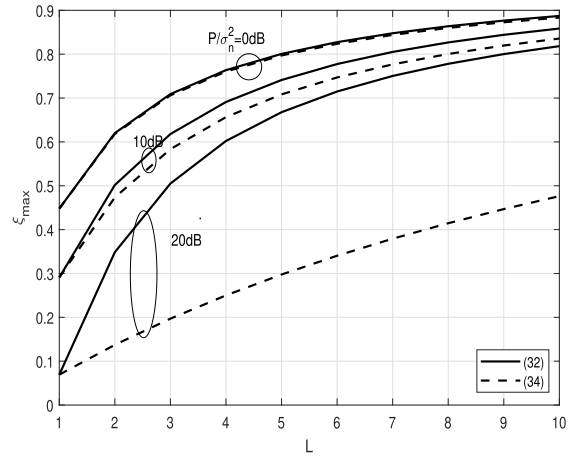


FIGURE 2. The maximum entire detection error probability, ξ_{max} , against L for distinct values of P/σ_n^2 ; $M = 1, N = 1, P_l = P/L$ for $1 \leq l \leq L, \phi = 0.9$ and $\sigma_f^2 = \sigma_g^2 = 1$.

[33], then a covert transmission is achievable. This means $\xi_{max} \rightarrow 1$ at sufficiently small $\epsilon \rightarrow 0$ or the inefficient detection.

VI. DECODING OUTAGE PROBABILITY OF THE COVERT INFORMATION

This part determines the decoding outage probability of the covert information at Bob (covert user) assuming that the covert information is transmitted at a constant rate R_u . When u is superimposed onto s_1 , Bob receives the signal as

$$y_b = h \left(\sqrt{1 - \phi} u + \sqrt{\phi} s_1 + \sum_{l=2}^L s_l \right) + n_b, \tag{35}$$

where $n_b \sim CN(0, \sigma_n^2)$. Thence, the decoding outage probability of the covert information is computed as

$$P_{o,u} = \Pr(I(u; y_b) < R_u). \tag{36}$$

Relied on the law of total probability, $P_{o,u}$ reduces to

$$P_{o,u} = 1 - \sum_{k=0}^L \Pr(I(u; y_b) \geq R_u, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2). \tag{37}$$

If $|f_k|^2 > |h|^2 \geq |f_{k+1}|^2$, Bob can restore s_{k+1}, \dots, s_L and cancel them from y_b to get

$$y'_b = \begin{cases} h \left(\sqrt{1 - \phi} u + \sqrt{\phi} s_1 + \sum_{l=2}^k s_l \right) + n_b, & 1 \leq k \leq L, \\ h \sqrt{1 - \phi} u + n_b, & k = 0. \end{cases} \tag{38}$$

Then, Bob can achieve the rate to be

$$I(u; y'_b) = \begin{cases} \log_2 \left(1 + \frac{|h|^2(1-\phi)\gamma_1}{1 + |h|^2(\sum_{i=2}^k \gamma_i + \phi\gamma_1)} \right), & 1 \leq k \leq L, \\ \log_2 (1 + |h|^2(1 - \phi)\gamma_1), & k = 0. \end{cases} \tag{39}$$

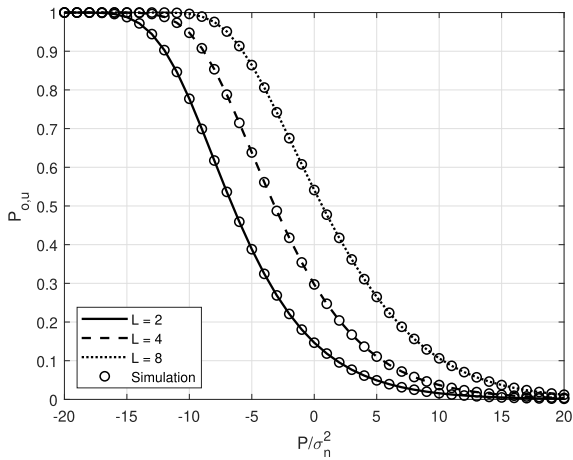


FIGURE 3. Decoding outage probability, $P_{o,u}$, against P/σ_n^2 for distinct values of L ; $R_u = 0.01$, $\phi = 0.9$, $M = 8$, $N = 1$, $P_l = P/L$ for $l \in [1, L]$, and $\sigma_f^2 = \sigma_h^2 = 1$.

Therefore, we obtain

$$\begin{aligned}
 P_{o,u} &= 1 - \sum_{k=0}^L \Pr(I(u; y'_b) \geq R_u, |f_k|^2 > |h|^2 \geq |f_{k+1}|^2) \\
 &= 1 - \sum_{k=0}^L \Pr(|h|^2 \geq \mu_k, |f_k|^2 > |h|^2 \geq |f_{k+1}|^2),
 \end{aligned} \tag{40}$$

where

$$\mu_k = \begin{cases} \frac{2^{R_u} - 1}{\left((1 - \phi)\gamma_1 - (2^{R_u} - 1)(\sum_{l=2}^k \gamma_l + \phi\gamma_1) \right)^+}, & 1 \leq k \leq L, \\ (2^{R_u} - 1)/((1 - \phi)\gamma_1), & k = 0. \end{cases} \tag{41}$$

Then, one obtains

$$\begin{aligned}
 P_{o,u} &= 1 - \sum_{k=0}^L \int_{\mu_k}^{\infty} \binom{L}{k} \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} \right)^k \frac{e^{-x/\sigma_h^2}}{\sigma_h^2} dx \\
 &= 1 - \frac{\sigma_f^2}{\sigma_h^2} \sum_{k=0}^L \binom{L}{k} B \left(e^{-\frac{\mu_k}{\sigma_f^2}}; k + \frac{\sigma_f^2}{\sigma_h^2}, L - k + 1 \right),
 \end{aligned} \tag{42}$$

where the incomplete Beta function [35] is notated as $B(e^{-x}; p, q) = \int_0^{e^{-x}} z^{p-1}(1-z)^{q-1} dz = \int_x^{\infty} (1 - e^{-y})^{q-1} (e^{-y})^p dy$ wherein $B(0; p, q) = 0$ and $B(1; p, q) = B(p, q)$ [35].

Fig. 3 shows the decoding outage probability, $P_{o,u}$, against the transmit SNR P/σ_n^2 for distinct values of L . This figure shows that the decoding outage probability $P_{o,u}$ declines when the SNR, P/σ_n^2 , accretes and the decline is more significant for smaller L . This is because of the increasing interference of the overt information. Thus, although multiplicity of overt information in NOMA systems can help increasing ξ_{max}

(shown in Fig. 4), it will decline the decoding performance at the covert user. Moreover, this figure illustrates the agreement between the simulation and analytical results.

VII. SPECTRAL EFFICIENCY

This part analyzes the covert spectral efficiency, interpreted as the maximum reliable transmission rate (bits/s) over a given bandwidth between Bob and Alice constrained by the covert condition of $\xi_{max} \geq 1 - \epsilon$ for some ϵ that stands for the covertness requirement. We also analyze the overt spectral efficiency and its loss due to sharing the transmission power for the covert information.

A. COVERT INFORMATION

Because the covert information is successfully decoded on $1 - P_{o,u}$ transmissions, the average rate received over several transmission bursts is $R_u(1 - P_{o,u})$ [37]. Then, the covert spectral efficiency (bits/s) is defined as [33]

$$\begin{aligned}
 \max_{R_u, P} \eta(R_u, P) &= L \times B \times R_u(1 - P_{o,u}) \\
 \text{s.t. } \xi_{max} &\geq 1 - \epsilon
 \end{aligned} \tag{43}$$

for a positive value ϵ wherein B is the bandwidth of a single information when the orthogonal multiple access (OMA) is applied. Note that, when NOMA is applied, the bandwidth of L channels of L overt NOMA information is $L \times B$ [2], which is used to compute the spectral efficiency in the objective function of (43).

Since ξ_{max} is a declining function of P , the maximum allowed power, denoted as P^* , can be found by having $\xi_{max} = 1 - \epsilon$. Also since $P_{o,u}$ is a decreasing function of P (shown in Fig. 3), the covert throughput is obtained as

$$\eta_u := \max_{R_u} \eta(R_u, P^*) \tag{44}$$

by numerical search.

B. OVERT INFORMATION

The average spectral efficiency (bits/s) of the overt information s_1 where the covert information is superimposed onto can be obtained from (24) as

$$\eta_s = L \times B \times \int_0^{\infty} \log_2 \left(1 + \frac{x\phi\gamma_1}{1 + x(1 - \phi)\gamma_1} \right) f_{|f_1|^2}(x) dx, \tag{45}$$

where

$$f_{|f_1|^2}(x) = L(1 - e^{-x/\sigma_f^2})^{L-1} e^{-x/\sigma_f^2} / \sigma_f^2 \tag{46}$$

is the probability density function (PDF) of $|f_1|^2$ [38].

The loss of overt spectral efficiency, which is computed by the difference of the overt spectral efficiency when the covert information is not sent ($\phi = 1$) and is sent ($\phi \neq 1$), is given by

$$\begin{aligned}
 \eta_{s,loss} &= L \times B \times \int_0^{\infty} \left[\log_2(1 + x\gamma_1) \right. \\
 &\quad \left. - \log_2 \left(1 + \frac{x\phi\gamma_1}{1 + x(1 - \phi)\gamma_1} \right) \right] f_{|f_1|^2}(x) dx
 \end{aligned}$$

$$\begin{aligned}
 & -\log_2 \left(1 + \frac{x\phi\gamma_1}{1+x(1-\phi)\gamma_1} \right) \Big] f_{|f_1|^2}(x) dx \\
 & = L \times B \times \int_0^\infty \log_2(1+x(1-\phi)\gamma_1) f_{|f_1|^2}(x) dx.
 \end{aligned} \tag{47}$$

Note that the loss of overt spectral efficiency indicates the performance degradation of NOMA systems when the covert transmission is embedded.

At low SNR ($\gamma_1 \ll 1$), the loss of overt spectral efficiency can be approximated from $\ln(1+x) \simeq x$ for $x \ll 1$ as

$$\begin{aligned}
 \eta_{s,loss} & \simeq \frac{L \times B \times (1-\phi)\gamma_1}{\ln(2)} \times \int_0^\infty x f_{|f_1|^2}(x) dx \\
 & = \frac{L \times B \times (1-\phi)\gamma_1 \sigma_f^2}{\ln(2)} \times \sum_{i=1}^L \frac{1}{i},
 \end{aligned} \tag{48}$$

which is derived from [38]. One can see that the loss of overt spectral efficiency in (48) increases with a coefficient of $L \times \ln(L)$ as L increases.

VIII. MULTIPLE ANTENNA AT ALICE

This part considers the case that Alice has M antennas and the NOMA signal is transmitted through one of them. Since merely one RF chain is used, this case reduces energy consumption and hardware complexity without losing the diversity gain [2], [39]. Given that the m -th antenna is selected, let $|f_{m,1}| \geq |f_{m,2}| \geq \dots \geq |f_{m,L}|$ notate the ordered channel gain between Alice and L overt receivers. Then based on (26), the maximum entire detection error probability given that the m -th antenna is selected is given by

$$\begin{aligned}
 & \xi_{max}(m) \\
 & = \Pr \left(|f_{m,1}|^2 > \frac{|g_m|^2}{(\phi + |g_m|^2(\phi \sum_{l \in \mathcal{D}_1^c(m)} \gamma_l - \gamma_1))^+} \right),
 \end{aligned} \tag{49}$$

where g_m is the channel gain between the m -th antenna of the transmitter and Willie, $\mathcal{D}_1(m)$ is the decoding set of Willie before restoring s_1 and $\mathcal{D}_1^c(m)$ is the complement of $\mathcal{D}_1(m)$. Next, we will derive the optimum antenna selection at Alice to minimize the warden's detection probability. Finally, the spectral efficiency and the maximum entire detection error probability will be derived when the optimum antenna selection is applied.

A. OPTIMUM ANTENNA SELECTION FOR ALICE

Selecting the antenna to maximize $\xi_{max}(m)$ yields

$$\begin{aligned}
 & \xi_{max} \\
 & = \max_{1 \leq m \leq M} \xi_{max}(m) \\
 & = \max_{1 \leq m \leq M} \Pr \left(|f_{m,1}|^2 > \frac{|g_m|^2}{(\phi + |g_m|^2(\phi \sum_{l \in \mathcal{D}_1^c(m)} \gamma_l - \gamma_1))^+} \right).
 \end{aligned} \tag{50}$$

When $|g_m|^2$ and $\mathcal{D}_1^c(m)$ are unavailable at the transmitter, the maximum of (50) is obtained by choosing the

antenna that maximizes $|f_{m,1}|^2$. If we let $|f_{m^*,1}|^2 = \max_{l \in [1,L], m \in [1,M]} |f_{m,l}|^2$, where $m^* = \arg \max_{m,l} |f_{m,l}|^2$, then the resulting maximum entire detection error probability is expressed as²

$$\begin{aligned}
 & \xi_{max} \\
 & = \Pr \left(|f_{m^*,1}|^2 > \frac{|g_{m^*}|^2}{(\phi + |g_{m^*}|^2(\phi \sum_{l \in \mathcal{D}_1^c(m^*)} \gamma_l - \gamma_1))^+} \right) \\
 & = 1 - \sum_{k=0}^L \Pr \left(|f_{m^*,1}|^2 \leq \nu_k(|g_{m^*}|^2), \right. \\
 & \quad \left. |f_{m^*,k}|^2 > |g_{m^*}|^2 \geq |f_{m^*,k+1}|^2 \right),
 \end{aligned} \tag{51}$$

where $|f_{m^*,0}|^2 = \infty$, $|f_{m^*,L+1}|^2 = 0$ and it is assumed that $|f_{m^*,k}|^2 > |f_{m^*,k+1}|^2$ for $k \in [0, L]$.

Next, the maximum entire detection error probability at Willie as well as decoding outage probability at Bob when the optimum antenna selection is applied at Alice will be derived.

B. MAXIMUM ENTIRE DETECTION ERROR PROBABILITY

Conditioned on $|f_{m^*,1}|^2 = t$, where its PDF is given by [36]

$$f_{|f_{m^*,1}|^2}(t) = LM(1 - e^{-t/\sigma_f^2})^{LM-1} e^{-t/\sigma_f^2} / \sigma_f^2, \tag{52}$$

the common un-ordered CDF of $(M-1)$ ordered variables $|f_{m^*,2}|^2 \geq \dots \geq |f_{m^*,L}|^2$ is given by

$$F(x|t) = (1 - e^{-x/\sigma_f^2}) / (1 - e^{-t/\sigma_f^2}), x \in (0, t]. \tag{53}$$

Then, we have for given $|f_{m^*,1}|^2 = t$,

$$\begin{aligned}
 & \Pr(|f_{m^*,k}|^2 > |g_{m^*}|^2 \geq |f_{m^*,k+1}|^2 | t) \\
 & = \binom{L-1}{k-1} \frac{(1 - e^{-x/\sigma_f^2})^{L-k} (e^{-x/\sigma_f^2} - e^{-t/\sigma_f^2})^{k-1}}{(1 - e^{-t/\sigma_f^2})^{L(M-1)}}.
 \end{aligned} \tag{54}$$

It follows from (52), (53), and (54), and the binomial expansion of

$$\left(1 - e^{-\frac{t}{\sigma_f^2}} \right)^a = \sum_{l=0}^a \binom{a}{l} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{t}{\sigma_f^2}} \right)^l \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^{a-l} \tag{55}$$

that, since $|g_{m^*}|^2$ is distributed exponentially with mean σ_g^2 , we obtain (56), as shown at the bottom of the next page, for $1 \leq k \leq L$, and for $k = 0$,

$$\begin{aligned}
 & \Pr(|f_{m^*,1}|^2 \leq \nu_0(|g_{m^*}|^2), |f_{m^*,1}|^2 \leq |g_{m^*}|^2) \\
 & = \frac{\sigma_f^2}{\sigma_g^2} B \left(\frac{\sigma_f^2}{\sigma_g^2}, LM + 1 \right).
 \end{aligned} \tag{57}$$

Thence, based on (51), (56), and (57), the maximum entire detection error probability when the optimum antenna selection is applied at Alice is obtained by (58), as shown at the bottom of the next page. When $M = 1$, (58) reduces to (32).

² [39] also investigated the antenna selection, which achieves the maximum channel gain of the resulting strongest user (information), to attain a nearly-optimal sum rate in NOMA systems.

C. SPECTRAL EFFICIENCY

Substituting $|h|^2$ with $|h_{m^*}|^2$ and $|f_k|^2$ with $|f_{m^*,k}|^2$ for $0 \leq k \leq L$, one obtains from (40) that

$$\begin{aligned}
 P_{o,u} &= 1 - \sum_{k=0}^L \Pr(|h_{m^*}|^2 \geq \mu_k, |f_{1,k}|^2 > |h_{m^*}|^2 \geq |f_{1,k+1}|^2) \\
 &= 1 - \frac{\sigma_f^2}{\sigma_h^2} \sum_{k=1}^L \binom{L}{k} \sum_{l=0}^{L(M-1)} \frac{Mk}{k+l} \binom{L(M-1)}{l} \\
 &\quad \times B\left(e^{-\frac{\mu_k}{\sigma_f^2}}; k+l + \frac{\sigma_f^2}{\sigma_h^2}, LM - k - l + 1\right), \quad (59)
 \end{aligned}$$

where the second term of (59) is derived similar to (56). When $M = 1$, (59) reduces to (42).

Therefore, the spectral efficiency of the covert information can be derived from (50), (59) and the optimization problem (43). The spectral efficiency of the overt information with its loss due to the covert transmissions can also be obtained from (45) and (47), respectively, where the PDF of $|f_1|^2$ is changed to

$$f_{|f_1|^2}(x) = LM(1 - e^{-x/\sigma_f^2})^{LM-1} e^{-x/\sigma_f^2} / \sigma_f^2. \quad (60)$$

IX. NUMERICAL RESULTS

This part presents the numerical results. For comparing the spectral efficiency of covert and overt information for NOMA with L users, we consider the standard bandwidth of $B = 200$ KHz in GSM.

A. MAXIMUM ENTIRE DETECTION ERROR PROBABILITY (MEDEP)

Fig. 4 shows the MEDEP, ξ_{max} , when the total SNR, P/σ_n^2 , increases with L , versus L for different values of M . One can see that, as L increases, ξ_{max} increases and converges to 1 even when the total SNR, P/σ_n^2 , increases with L . This demonstrates that the multiplicity of overt information (users) in NOMA can be leveraged to conceal the transmission of covert information. One can also see that the larger M , the faster the convergence speed. This is because the degree of freedom of the highest channel gain of the overt message where the covert message is superimposed onto, increases with increasing M , making faster the convergence speed of the MEDEP (as proved in Appendix B). Moreover, this figure illustrates the agreement between the simulation and analytical results.

Fig. 5 plots the MEDEP, ξ_{max} , versus L for different number of the warden’s antenna. One can see that, as N increases, ξ_{max} decreases significantly. This is because the multi-antenna warden has higher chance to decode the overt messages and, consequently, to detect the covert information.

Fig. 6 shows the MEDEP, ξ_{max} , against the total transmit SNR, P/σ_n^2 , for distinct values of L . This figure exposes that ξ_{max} decreases as the SNR increases and that the decrease is negligible for large L . One can also see that for $L > 1$, ξ_{max} converges to a non-zero constant even if the transmission power is very large. This exposes that the interference of many overt information is the helpful source to hide the presence of the covert information.

Fig. 7 shows the MEDEP, ξ_{max} , against α , for distinct values of L . One can see that ξ_{max} increases as α increases.

$$\begin{aligned}
 &\Pr(|f_{m^*,1}|^2 \leq v_k(|g_{m^*}|^2), |f_{m^*,k+1}|^2 \leq |g_{m^*}|^2 < |f_{m^*,k}|^2) \\
 &= \int_0^{\phi_k} \left(\int_x^{v_k(x)} \Pr(|f_{m^*,k+1}|^2 \leq x < |f_{m^*,k}|^2 | t) f_{|f_{m^*,1}|^2}(t) \right) \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \\
 &= \binom{L}{k} Mk \int_0^{\phi_k} \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^{L-k} \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \int_x^{v_k(x)} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{t}{\sigma_f^2}} \right)^{k-1} \left(1 - e^{-\frac{t}{\sigma_f^2}} \right)^{L(M-1)} \frac{e^{-t/\sigma_f^2}}{\sigma_f^2} dt \\
 &= \binom{L}{k} Mk \sum_{l=0}^{L(M-1)} \binom{L(M-1)}{l} \int_0^{\phi_k} \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^{LM-k-l} \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \int_x^{v_k(x)} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{t}{\sigma_f^2}} \right)^{k+l-1} \frac{e^{-t/\sigma_f^2}}{\sigma_f^2} dt \\
 &= \binom{L}{k} Mk \sum_{l=0}^{L(M-1)} \binom{L(M-1)}{l} \int_0^{\phi_k} \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^{LM-k-l} \frac{\left(e^{-x/\sigma_f^2} - e^{-v_k(x)/\sigma_f^2} \right)^{k+l} e^{-x/\sigma_g^2}}{k+l} dx. \quad (56)
 \end{aligned}$$

$$\begin{aligned}
 \xi_{max} &= 1 - \frac{\sigma_f^2}{\sigma_g^2} B\left(\frac{\sigma_f^2}{\sigma_g^2}, LM + 1\right) - \sum_{k=1}^L \binom{L}{k} \sum_{l=0}^{L(M-1)} \frac{Mk}{k+l} \binom{L(M-1)}{l} \\
 &\quad \times \int_0^{\phi_k} \left(1 - e^{-\frac{x}{\sigma_f^2}} \right)^{LM-k-l} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{v_k(x)}{\sigma_f^2}} \right)^{k+l} \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx. \quad (58)
 \end{aligned}$$

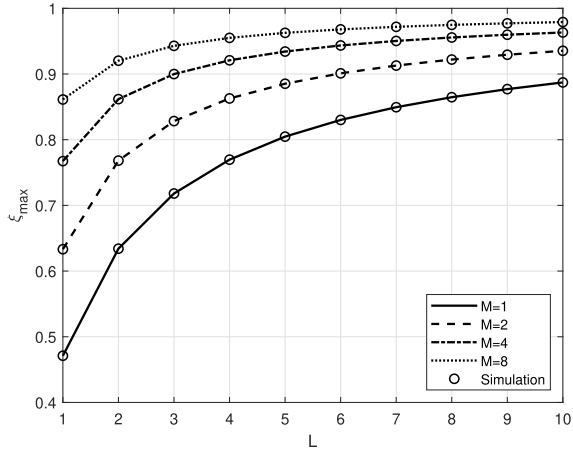


FIGURE 4. The MEDEP, ξ_{max} , against L for distinct values of M ; $N = 1$, $P_I/\sigma_n^2 = -10$ dB for $l \in [1, L]$, $\phi = 0.9$, and $\sigma_f^2 = \sigma_g^2 = 1$.

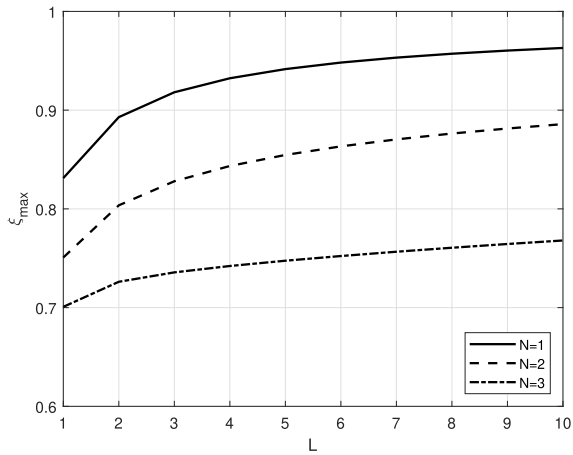


FIGURE 5. The MEDEP, ξ_{max} , against L for distinct number of the warden's antennas N ; $M = 8$, $P_I/\sigma_n^2 = -10$ dB for $l \in [1, L]$, $\phi = 0.9$, and $\sigma_f^2 = \sigma_g^2 = 1$.

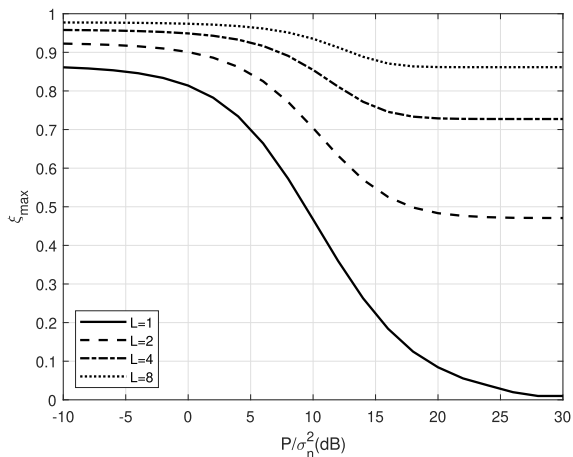


FIGURE 6. The MEDEP, ξ_{max} , against the transmit SNR, P/σ_n^2 , for distinct values of L ; $P_I = P/L$ for $l \in [1, L]$, $\phi = 0.9$, $M = 8$, $N = 1$ and $\sigma_f^2 = \sigma_g^2 = 1$.

This is because of less power located to the covert information, harder detection at Willie. One can also see that ξ_{max} increases significantly in all ranges of α as L increases.

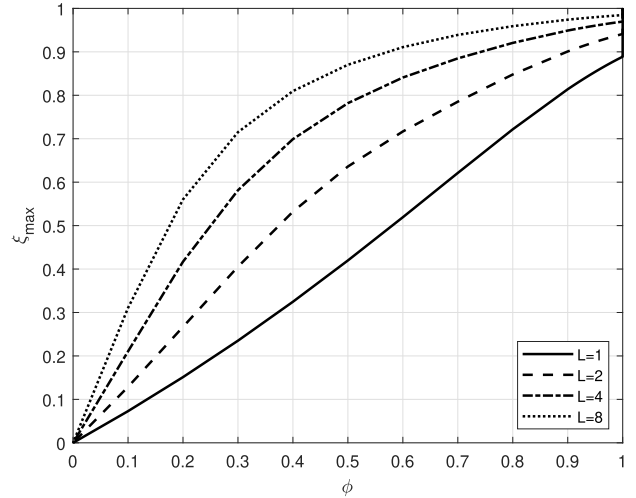


FIGURE 7. The MEDEP, ξ_{max} , against ϕ for distinct values of L ; $P_I = P/L$ for $l \in [1, L]$, $M = 8$, $N = 1$, $P/\sigma_n^2 = 0$ dB and $\sigma_f^2 = \sigma_g^2 = 1$.

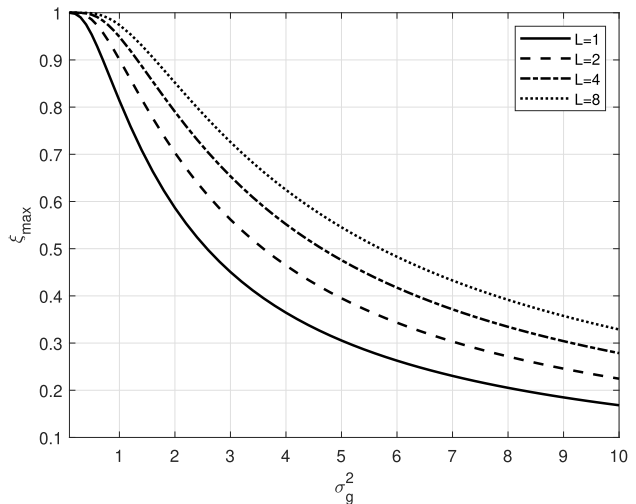


FIGURE 8. The MEDEP, ξ_{max} , against σ_g^2 , for distinct values of L ; $P_I = P/L$ for $l \in [1, L]$, $\phi = 0.9$, $M = 8$, $N = 1$, $P/\sigma_n^2 = 0$ dB and $\sigma_f^2 = 1$.

Fig. 8 shows the MEDEP, ξ_{max} , against σ_g^2 , for distinct values of L . One can see that ξ_{max} decreases dramatically as σ_g^2 increases (Willie is located closer to Alice). This is because Willie with better channel has more chance to successfully decode the overt information, hence detect the covert information.

B. SPECTRAL EFFICIENCY

Fig. 9 shows the covert spectral efficiency, η_u (Kbits/s), against the MEDEP, ξ_{max} , when the SNR is changed. This figure demonstrates that the covert spectral efficiency can be traded with ξ_{max} , i.e. η_u decreases as ξ_{max} increases. One can also see that when ξ_{max} is close to 1, which is the region of interest, the covert spectral efficiency can be increased if L is increased. This indicates that the multiplicity of overt information in NOMA can help hiding the covert information. However, for less strict requirements of covertness (small

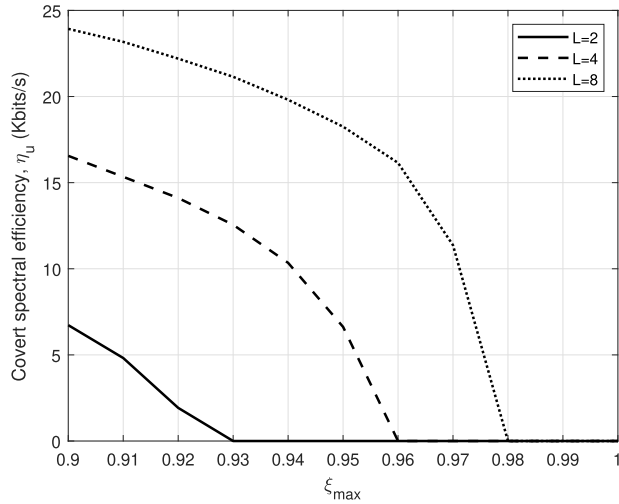


FIGURE 9. Covert spectral efficiency (Kbits/s), η_u , versus the MEDEP, ξ_{max} ; $\phi = 0.9$, $M = 8$, $N = 1$, $P_l = P/L$ for $l \in [1, L]$, and $\sigma_f^2 = \sigma_h^2 = \sigma_g^2 = 1$.

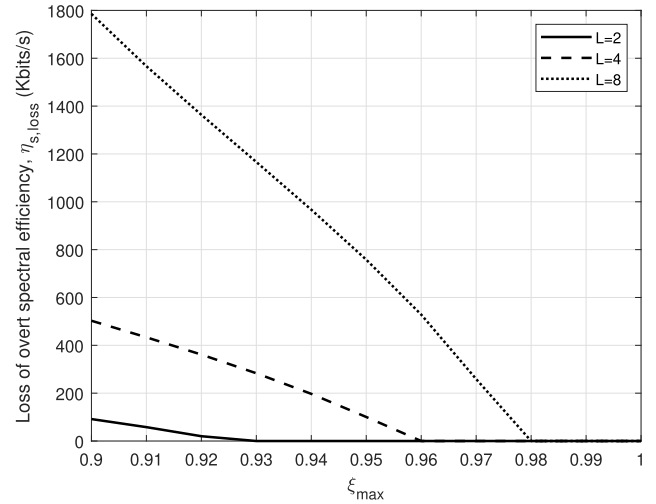


FIGURE 11. Loss of overt spectral efficiency (Kbits/s) versus the MEDEP, ξ_{max} ; $\phi = 0.9$, $M = 8$, $N = 1$, $P_l = P/L$ for $l \in [1, L]$, and $\sigma_f^2 = \sigma_h^2 = \sigma_g^2 = 1$.

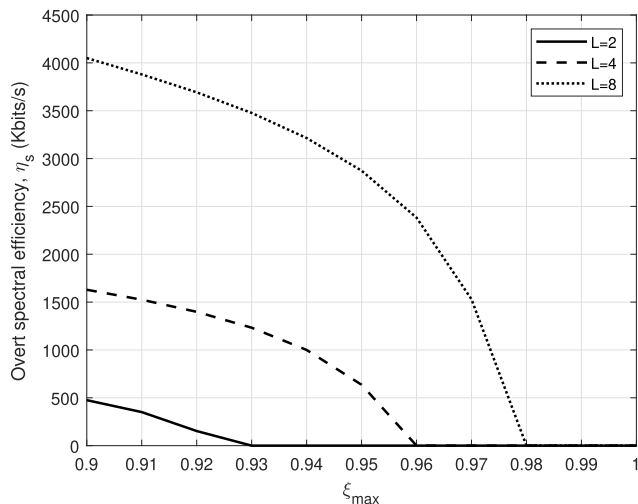


FIGURE 10. Overt spectral efficiency (Kbits/s) versus the MEDEP, ξ_{max} ; $\phi = 0.9$, $M = 8$, $N = 1$, $P_l = P/L$ for $l \in [1, L]$, and $\sigma_f^2 = \sigma_h^2 = \sigma_g^2 = 1$.

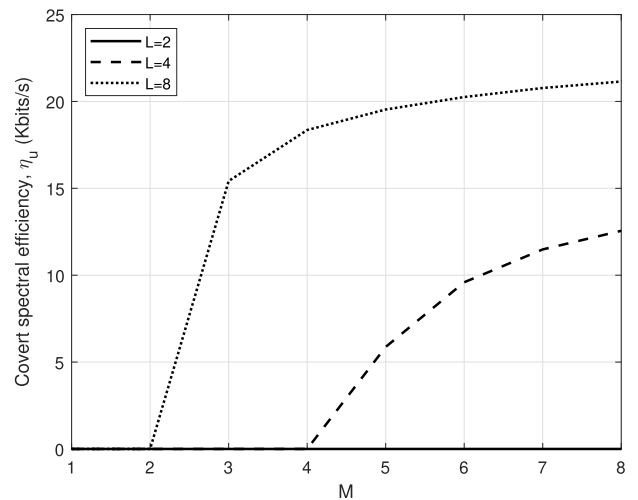


FIGURE 12. Covert spectral efficiency (Kbits/s) versus M for different values of L ; $\phi = 0.9$, $N = 1$, $P_l = P/L$ for $l \in [1, L]$, $\epsilon = 0.07$ and $\sigma_f^2 = \sigma_h^2 = \sigma_g^2 = 1$.

ξ_{max}), an increase of L will significantly increase decoding outage of u (as shown in Fig. 3), resulting in a decrease of the covert spectral efficiency, as comparing η_u when $L = 4$ and $L = 8$ in the Fig. 9.

Fig. 10 shows the overt spectral efficiency, η_s (Kbits/s), versus the MEDEP, ξ_{max} , as the SNR is varied. One can see that the overt spectral efficiency decreases as ξ_{max} increases, which is because the transmit SNR decreases with increasing ξ_{max} . One can also see that the overt spectral efficiency accretes with accreting L .

Fig. 11 shows the loss of overt spectral efficiency, $\eta_{s,loss}$ (Kbits/s), against the MEDEP, ξ_{max} , when the SNR is changed. This figure illustrates that, to provide a certain covert spectral efficiency, the loss of overt spectral efficiency is significantly high. For example, to provide about 7 (Kbits/s) of the covert spectral efficiency (in Fig. 9 at $\xi_{max} = 0.9$ and $L = 2$), the 100 (Kbits/s) loss of overt spectral efficiency is required (in Fig. 11 at $\xi_{max} = 0.9$ and $L = 2$).

One can also see that the loss is more significant for larger L or less strict ξ_{max} .

Fig. 12 shows the covert spectral efficiency (Kbits/s) versus M , as the SNR is varied. One can see that the covert spectral efficiency accretes as the number of transmit antennas M accretes and that the larger L , the more considerable the increase. One can also see that, when M is less than a threshold (at 2 and 4 for $L = 4$ and $L = 8$, respectively) or the number of transmit antenna is not sufficiently high, the covert spectral efficiency goes to zero.

Fig. 13 shows the covert spectral efficiency (Kbits/s) versus σ_g^2 , as the SNR is varied. This figure demonstrates that the covert spectral efficiency declines drastically to zero, which the covert communication may not be achieved, when σ_g^2 accretes (Willie is located closer to Alice) regardless of number of overt information L . This is because Willie easily decodes all overt information, hence detect the covert information, when located close to Alice (as seen in Fig. 8).

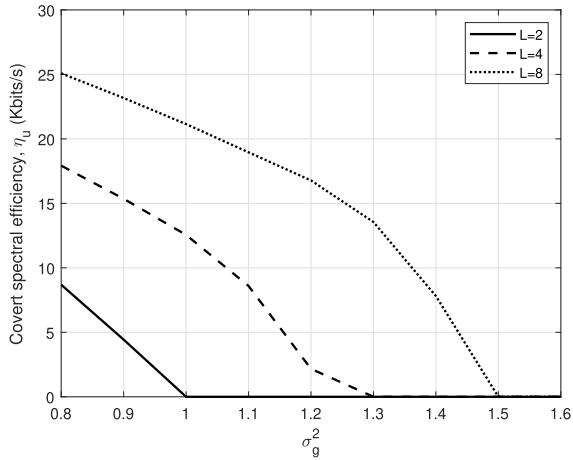


FIGURE 13. Covert spectral efficiency (Kbits/s) versus σ_g^2 for different values of L ; $P_l = P/L$ for $l \in [1, L]$, $M = 8$, $N = 1$, $\phi = 0.9$, $\epsilon = 0.07$ and $\sigma_f^2 = \sigma_h^2 = 1$.

X. CONCLUSION AND FUTURE WORK

This paper investigated hiding a covert information under other information in NOMA systems. The optimum detection method at the warden, which minimizes the entire detection error probability, and the optimum superimposition method at the transmitter, which maximizes the minimum total detection error probability, were proposed. Numerous results proved the increase and the convergence to 1 of the entire detection error probability with accreting the number of users (information). We also found that the covert spectral efficiency can be increased if the number of overt users is increased while guaranteeing the entire detection error probability close to 1. The covert performance is much better if the transmitter has multi-antennas. These results showed the practical application of the proposed scheme, in which the transmitter can control and design the covert transmission by adjusting the NOMA network parameters for the case when the warden is a node within the network.

For the future work, a complicated NOMA network, such as clustering or beamforming with multi-antenna at the transmitter and the overt users should be considered to characterize the improvement of the current work. While the proposed scheme showed its promising to apply in the real networks, it has not also considered the multi-antenna or arbitrary location of the warden, which should be considered in the future work. After all, designing a hidden network or covert transmissions under existing networks is a difficult task and many other existing networks, such as cognitive radio or massive MIMO, are also interesting work for the study of hiding covert information.

APPENDIX A

This Appendix provides the proof of (7) and (8). Based on (4) and (6), one obtains

$$\begin{aligned} E_{s_l \in \mathcal{D}^c} [\Pr(y_w | g, \{s_l\}, H_0)] \\ = \left(\frac{1}{\pi \sigma_n^2} \right)^n E_{s_l \in \mathcal{D}^c} \left[e^{-\|y_w - g \sum_{l=1}^L s_l\|^2 / \sigma_n^2} \right]. \end{aligned} \quad (61)$$

Let $x = e^{j\theta} \sum_{l \in \mathcal{D}^c} s_l$, $g = |g|e^{j\theta}$ and $y = y_w - g \sum_{l \in \mathcal{D}^c} s_l$. Because s_l , $l \in \mathcal{D}^c$, is an i.i.d circular symmetric complex Gaussian vector (cscGv), the PDF of x is a cscGv with mean 0 and common variance $\sum_{l \in \mathcal{D}^c} P_l$ for any θ . Then, one obtains

$$\begin{aligned} E_{s_l \in \mathcal{D}^c} [\Pr(y_w | g, \{s_l\}, H_0)] \\ = \left(\frac{1}{\pi \sigma_n^2} \right)^n E_x \left[e^{-\|y - |g|x\|^2 / \sigma_n^2} \right] \\ = \left(\frac{1}{\pi \sigma_n^2} \right)^n \prod_{i=1}^n E_{x_i} \left[e^{-|y_i - |g|x_i|^2 / \sigma_n^2} \right] \\ = \left(\frac{1}{\pi \sigma_n^2} \right)^n \prod_{i=1}^n \int_{x_i} e^{-|y_i - |g|x_i|^2 / \sigma_n^2} \frac{e^{-|x_i|^2 / \sum_{l \in \mathcal{D}^c} P_l}}{\pi \sum_{l \in \mathcal{D}^c} P_l} dx. \end{aligned} \quad (62)$$

To complete the derivation of (62), one needs to solve the last integral. Towards this end, one needs to solve the following integral as

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-(y - |g|x)^2 / \sigma_n^2} \frac{e^{-x^2 / \sum_{l \in \mathcal{D}^c} P_l}}{\sqrt{\pi \sum_{l \in \mathcal{D}^c} P_l}} dx \\ = \sqrt{\frac{\sigma_n^2}{\sigma_0^2}} e^{-y^2 / \sigma_0^2} \int_{-\infty}^{\infty} \frac{e^{-\frac{(x - y|g| \sum_{l \in \mathcal{D}^c} P_l / \sigma_0^2)^2}{\sigma_n^2 \sum_{l \in \mathcal{D}^c} P_l / \sigma_0^2}}}{\sqrt{\pi \sigma_n^2 \sum_{l \in \mathcal{D}^c} P_l / \sigma_0^2}} dx \\ = \sqrt{\frac{\sigma_n^2}{\sigma_0^2}} e^{-y^2 / \sigma_0^2}, \end{aligned} \quad (63)$$

where y and x are real, $\sigma_0^2 = \sigma_n^2 + |g|^2 \sum_{l \in \mathcal{D}^c} P_l$. Hence, we obtain

$$\begin{aligned} \int_{x_i} e^{-|y_i - |g|x_i|^2 / \sigma_n^2} \frac{e^{-|x_i|^2 / \sum_{l \in \mathcal{D}^c} P_l}}{\pi \sum_{l \in \mathcal{D}^c} P_l} dx \\ = \int_{-\infty}^{\infty} e^{-\frac{(Re\{y_i\} - |g|Re\{x_i\})^2}{\sigma_n^2}} \frac{e^{-Re\{x_i\}^2 / \sum_{l \in \mathcal{D}^c} P_l}}{\sqrt{\pi \sum_{l \in \mathcal{D}^c} P_l}} d(Re\{x_i\}) \\ \times \int_{-\infty}^{\infty} e^{-\frac{(Im\{y_i\} - |g|Im\{x_i\})^2}{\sigma_n^2}} \\ \times \frac{e^{-Im\{x_i\}^2 / \sum_{l \in \mathcal{D}^c} P_l}}{\sqrt{\pi \sum_{l \in \mathcal{D}^c} P_l}} d(Im\{x_i\}) \\ = \frac{\sigma_n^2}{\sigma_0^2} e^{-|y_i|^2 / \sigma_0^2}, \end{aligned} \quad (64)$$

where $Im\{x\}$ and $Re\{x\}$ denote the imaginary and real parts of x , respectively. Then, we obtain from (62) and (64) that

$$\begin{aligned} E_{s_l \in \mathcal{D}^c} [\Pr(y_w | g, \{s_l\}, H_0)] \\ = \frac{e^{-\|y\|^2 / \sigma_0^2}}{(\pi \sigma_0^2)^n} \\ = \frac{e^{-\|y_w - g \sum_{l \in \mathcal{D}^c} s_l\|^2 / \sigma_0^2}}{(\pi \sigma_0^2)^n}. \end{aligned} \quad (65)$$

Similarly, since $\mathcal{S} = (\mathcal{S} \cap \mathcal{D}) \cup (\mathcal{S} \cap \mathcal{D}^c)$ and $\mathcal{S}^c = (\mathcal{S}^c \cap \mathcal{D}) \cup (\mathcal{S}^c \cap \mathcal{D}^c)$, we obtain from (5) and (6) that

$$\begin{aligned} & E_{s_l \in \mathcal{D}^c} [\Pr(y_w | g, u, \{s_l\}, H_1)] \\ &= \left(\frac{1}{\pi(\sigma_n^2 + (1-\phi)|g|^2 \sum_{l \in \mathcal{S}} P_l)} \right)^n \\ & \quad \times E_{s_l \in \mathcal{D}^c} \left[e^{-\frac{\|y_w - g(\sqrt{\phi} \sum_{l \in \mathcal{S}} s_l + \sum_{l \in \mathcal{S}^c} s_l)\|^2}{\sigma_n^2 + (1-\phi)|g|^2 \sum_{l \in \mathcal{S}} P_l}} \right] \\ &= \left(\frac{1}{\pi\sigma_1^2} \right)^n e^{-\|y_w - g(\sqrt{\phi} \sum_{l \in \mathcal{S} \cap \mathcal{D}} s_l + \sum_{l \in \mathcal{S}^c \cap \mathcal{D}} s_l)\|^2 / \sigma_1^2}, \quad (66) \end{aligned}$$

where $\sigma_1^2 = \sigma_n^2 + |g|^2(\phi \sum_{l \in \mathcal{S} \cap \mathcal{D}^c} P_l + (1-\phi) \sum_{l \in \mathcal{S}} P_l + \sum_{l \in \mathcal{S}^c \cap \mathcal{D}^c} P_l)$.

APPENDIX B

Let $\nu(g) = \frac{|g|^2}{(\phi + |g|^2(\phi \sum_{l \in \mathcal{D}^c} \gamma_l - \gamma_1))^+}$. Then, it follows (26) that

$$\begin{aligned} \xi_{max} &= 1 - \Pr(|f_1|^2 \leq \nu(g)) \\ &\stackrel{(a)}{=} 1 - \Pr(\bigcap_{l=1}^L |f_l|^2 \leq \nu(g)) \\ &\rightarrow 1 \quad (67) \end{aligned}$$

as $L \rightarrow \infty$, where (a) follows from $|f_1|^2 = \max_{1 \leq l \leq L} |f_l|^2$.

Note that when Alice has M antennas, replacing $|f_1|^2$ by $|f_{m^*,1}|^2$ into (67) yields

$$\xi_{max} = 1 - \Pr(|f_{m^*,1}|^2 \leq \nu(g)), \quad (68)$$

which converges to 1 with higher degree of freedom since $|f_{m^*,1}|^2$ is selected from $M \times L$ random variables. Hence, the convergence speed increases with increasing M .

REFERENCES

[1] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[2] Z. Ding, X. Lei, G. K. Karagiannis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[3] Q. Sun, S. Han, C.-L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[4] L. Lv, Z. Ding, Q. Ni, and J. Chen, "Secure MISO-NOMA transmission with artificial noise," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6700–6705, Jul. 2018.

[5] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Secure beamforming in downlink MISO nonorthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7563–7567, Aug. 2017.

[6] Y. Zhang, H.-M. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 930–933, May 2016.

[7] B. He, A. Liu, N. Yang, and V. K. N. Lau, "On the design of secure non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2196–2206, Oct. 2017.

[8] B. A. 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.

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

[10] L. Wang, G. W. 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.

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

[12] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[13] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, "Gaussian signalling for covert communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3542–3553, Jul. 2019.

[14] B. He, S. Yan, X. Zhou, and V. K. N. Lau, "On covert communication with noise uncertainty," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 941–944, Apr. 2017.

[15] 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.

[16] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, Feb. 2008.

[17] T. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6193–6206, Sep. 2017.

[18] T.-X. Zheng, H.-M. Wang, D. W. K. Ng, and J. Yuan, "Multi-antenna covert communications in random wireless networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1974–1987, Mar. 2019.

[19] B. He, S. Yan, X. Zhou, and H. Jafarkhani, "Covert wireless communication with a Poisson field of interferers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6005–6017, Sep. 2018.

[20] W. Ma, Z. Niu, W. Wang, S. He, and T. Jiang, "Covert communication with uninformed backscatters in hybrid active/passive wireless networks: Modeling and performance analysis," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2622–2634, Apr. 2022.

[21] R. Soltani, D. Goeckel, D. Towsley, B. A. Bash, and S. Guha, "Covert wireless communication with artificial noise generation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7252–7267, Nov. 2018.

[22] F. Shu, T. Xu, J. Hu, and S. Yan, "Delay-constrained covert communications with a full-duplex receiver," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 813–816, Jun. 2019.

[23] B. Bash, D. Goeckel, and D. Towsley, "Covert communication gains from adversary's ignorance of transmission time," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8394–8405, Dec. 2016.

[24] S. Yan, B. He, X. Zhou, Y. Cong, and A. L. Swindlehurst, "Delay-intolerant covert communications with either fixed or random transmit power," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 129–140, Jan. 2019.

[25] K. Shahzad and X. Zhou, "Covert wireless communications under quasi-static fading with channel uncertainty," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1104–1116, 2021.

[26] H. Q. Ta, Q.-V. Pham, K. Ho-Van, and S. W. Kim, "Covert communication with noise and channel uncertainties," *Wireless Netw.*, vol. 28, no. 1, pp. 161–172, Jan. 2022.

[27] K. S. Kumar Arumugam and M. R. Bloch, "Embedding covert information in broadcast communications," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2787–2801, Oct. 2019.

[28] K. S. K. Arumugam and M. R. Bloch, "Covert communication over a K -user multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7020–7044, Nov. 2019.

[29] A. Dutta, D. Saha, D. Grunwald, and D. Sicker, "Secret agent radio: Covert communication through dirty constellations," in *Proc. Int. Workshop Inf. Hiding*, Cham, Switzerland: Springer, 2012, pp. 160–175.

[30] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, "Covert communication achieved by a greedy relay in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4766–4779, Jul. 2018.

[31] L. Tao, W. Yang, S. Yan, D. Wu, X. Guan, and D. Chen, "Covert communication in downlink NOMA systems with random transmit power," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 2000–2004, Nov. 2020.

[32] L. Lv, Q. Wu, Z. Li, Z. Ding, N. Al-Dhahir, and J. Chen, "Covert communication in intelligent reflecting surface-assisted NOMA systems: Design, analysis, and optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1735–1750, Mar. 2022.

[33] S. W. Kim and H. Q. Ta, "Covert communications over multiple overt channels," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1112–1124, Feb. 2022.

[34] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

- [35] G. E. Andrews, R. Askey, R. Roy, R. Roy, and R. Askey, *Special Functions*, vol. 71. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [36] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*. Philadelphia, PA, USA: SIAM, 2008.
- [37] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [38] H. A. David and N. Nagaraja, *Order Statistics*, 3rd ed. Wiley, 2003.
- [39] Y. Yu, H. Chen, Y. Li, Z. Ding, L. Song, and B. Vucetic, "Antenna selection for MIMO nonorthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3158–3171, Apr. 2018.



HIEN QUANG TA (Member, IEEE) received the B.S. degree in electrical and electronic engineering from the Ho Chi Minh University of Technology, Vietnam, and the Ph.D. degree in electrical engineering from Iowa State University, Ames, IA, USA. He is currently a Lecturer with the Department of Electrical Engineering, International University—Vietnam National University, Ho Chi Minh City, Vietnam. His current research interests include wireless communications and networking.



Technology. His major research interests include modulation and coding techniques, diversity techniques, digital signal processing, energy harvesting, physical layer security, and cognitive radio.

KHUONG HO-VAN received the B.E. (Hons.) and M.S. degrees in electronics and telecommunications engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2001 and 2003, respectively, and the Ph.D. degree in electrical engineering from the University of Ulsan, South Korea, in 2007. From 2007 to 2011, he joined McGill University, Canada, as a Postdoctoral Fellow. He is currently an Associate Professor at the Ho Chi Minh City University of



of Quebec, Montreal, QC, Canada. From 2010 to 2022, he was with the Federal University of Ceará, Brazil. From January 2019 to April 2019, he was a Visiting Professor at the Lappeenranta University of Technology (LUT), Finland, with financial support from Nokia Foundation. He was awarded with the prestigious Nokia Visiting Professor Grant. From May 2019 to August 2019, he was with the King Abdullah University of Science and Technology (KAUST), Saudi Arabia, as a Visiting Faculty. From September 2019 to November 2019, he was a Visiting Researcher at Istanbul Medipol University, Turkey. From 2021 to 2022, he was a Full Professor at the National Yunlin University of Science and Technology (YunTech), Taiwan. Since 2022, he has been with the AI and Digital Science Research Center, Technology Innovation Institute (TII), a global research center and the applied pillar of Abu Dhabi's Advanced Technology Research Council. His Ph.D. Thesis was awarded the Best Ph.D. Thesis in electrical engineering by the Brazilian Ministry of Education (CAPES) at the 2009 CAPES Thesis Contest. He has acted as a symposium co-chair/a track co-chair in numerous IEEE flagship conferences. He is the editor of several IEEE journals.

DANIEL BENEVIDES DA COSTA (Senior Member, IEEE) was born in Fortaleza, Ceará, Brazil, in 1981. He received the B.Sc. degree in telecommunications from the Military Institute of Engineering (IME), Rio de Janeiro, Brazil, in 2003, and the M.Sc. and Ph.D. degrees in electrical engineering, area: telecommunications from the University of Campinas, São Paulo, Brazil, in 2006 and 2008, respectively. From 2008 to 2009, he was a Postdoctoral Research Fellow with INRS-EMT, University



ing, Iowa State University. Previously, he was a Faculty Member of KAIST; a Visiting Associate Professor with the California Institute of Technology, Pasadena, CA, USA; and a Technical Consultant with the Wireless Systems Research Department, AT&T Labs, Middletown, NJ, USA. His current research interests include wireless communications and security. He was a recipient of the Best Paper Award from the IEEE International Symposium on Spread Spectrum Techniques and Applications, Piscataway, NJ, in September 2000, and the IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS), Washington, DC, USA, in November 2005. He also received the Best Poster Award from the IEEE Conference on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON), San Diego, CA, in June 2007. He served as the Co-Chair for the Technical Program Committee of IEEE GLOBECOM Communication Theory Symposium 2013 and the Publications Chair for IEEE WCNC 2017. He served as an Associate Editor for IEEE COMMUNICATIONS LETTERS from 2000 to 2006. He is currently serving as the Editor for IEEE COMMUNICATIONS SURVEYS AND TUTORIALS and an Associate Editor for IEEE ACCESS.

SANG WU KIM received the B.S. degree in electronic engineering from Yonsei University, Seoul, South Korea, in 1981, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1983, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1987. He is currently an Associate Professor with the Department of Electrical and Computer Engineer-



for the data services of the CDMA and IMT2000 handset products. He has served as the Department Head of the School of IT Convergence, from 2016 to 2017. He has been leading the Digital Manufacturing IT Research Center, University of Ulsan, South Korea, as a Director, since 2013. He has published over 60 refereed journals and issued over 30 patents for the last decade and made many technology transfers to the IT companies through Foundation for Industry Cooperation, University of Ulsan. His research interests include embedded systems, mobile ad hoc and sensor networks, real-time computing, and context-aware computing. He is a member of IEICE and ICASE and a Lifetime Member of KICS and KISA. He received the Best Paper Award from the National Academy of Science, USA, in 1995, and the Best Paper Award from the 13th International Conference on Future Networks and Communications, Elsevier, in 2018.

HOON OH (Member, IEEE) received the B.S.E.E. degree from the Sung Kyun Kwan University, Seoul, and the M.Sc. and Ph.D. degrees in computer science from Texas A&M University, College Station, TX, USA, in 1993 and 1995, respectively. From 1983 to 1989 and 1996 to 2000, he worked as a Software Engineer and a Software Architect at the Corporate Research Center of Samsung Electronics, respectively. He was involved in developing communication protocols

...