SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2016; **9:**2359–2370
Published online 4 July 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.1503

RESEARCH ARTICLE

Design and analysis of the covert channel implemented by behaviors of network users

Yuwen Qian[1]*, Ting Sun[1], Jun Li[1], Chang Fan[1] and Huaju Song[2]

[1] School of Electronics and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[2] School of Biochemical and Environmental Engineering, Nanjing Xiaozhuang University, Nanjing 210011, China

ABSTRACT

In this paper, a novel covert channel, called covert behavior channel, is proposed according to behaviors of network users to solve the security and efficiency problem of the traditional covert channel. In the proposed channel, operation sequences of the network protocols are used as carriers of covert information. An encryption-based information embedding scheme is designed to improve security of the covert information. With the help of Markov model, the capacity of the proposed covert channel with time-varying noise is derived. The formulation for analyzing the covert behavior channel is presented against the channel noise aroused by discarding packets. By introducing corrected entropy-based algorithm to detect the covert behavior channel, the security of the channel is verified. Numerical results show that the proposed covert behavior channel is more secure than covert storage channels and achieves a better bit rate and robustness than that of covert timing channels. Copyright © 2016 John Wiley & Sons, Ltd.

KEYWORDS

network security; steganography; covert channel; capacity; robustness

***Correspondence**

Yuwen Qian, School of Electronics and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.
E-mail: admon1999@163.com

1. INTRODUCTION

The developments of computer networks and intrusion detection systems are forcing hackers to seek more subtle ways in stealing information. The network covert channel is an efficient tool, which allows two cooperating processes to transmit information in a manner that violates the system's security policy [1]. The network covert channel works not only as a hacking tool but also an important approach to transmit secret information such as private keys. The covert channel technology has become a novel method for network authentication, copyright protection and the evidence of cyber crime [2]. As such, the network covert channel attracts a wide attention.

Traditionally, network covert channels are characterized into two types: covert storage channel and covert timing channel [3]. A covert storage channel involves direct/indirect writing of a storage location by one process and the direct/indirect reading of the storage location by another process [4]. In a covert timing channel, covert information is encoded by varying packet rates, which is equivalent to modulating the transmitting time of packets [5]. Hence, covert timing channels can convey information through the arrival interval of packets, rather than through the contents of the packets themselves. Generally, covert timing channels are more secure than covert storage ones. However, because transmitting time of a packet is uncertain, varying the quality of network communication may make the timing channels fail [6]. This limits general applications for covert timing channels.

Recently, many works are contributed to improving the stability of the timing channel. An effective approach is to construct a timing channel by the pattern of arriving packets that can indicate behaviors of the transmitter [7,8]. Therefore, these covert timing channels can be viewed as covert behavior channel [9]. Inspired by the channel designed based on the arrival pattern of packets, Kundur *et al.* describe a covert channel implemented through packet sorting pattern [10]. In [10], a set of n packets can be arranged randomly, the entropy $\log_2(n!)$ can be achieved in the channel. Wolf constructs a behavior channel by modulating the pattern of acknowledge packets [11]. The receiving station can acknowledge each frame separately or acknowledge the first frame after several subsequent frames have arrived. Handel, however, proposed a behavior channel based on modulating the operators of clear-to-send/ready-to-send signals [12].

The pattern modulation of arriving packets typically results in the traffic with distinctive timing characteristics that deviates from legitimate traffic. Therefore, statistical tests that examine the shape and regularity of traffics can successfully detect these channels [13,14]. To improve security, the covert behavior channel is no longer limited to modulating the time feature of packets. Schulz *et al.* [15] use the sequence number of the internet protocol security (IPSEC) authentication header or encapsulating security payload to transmit covert information. URLs in HTTP are modulated by [16] to construct a distributed covert channel. Similarly, the average size of packets, the ratio of small and large packets, and the change of packet size patterns can act as carriers of the covert information [17,18]. In these channels, covert information is embedded into behaviors of network users. Because of not adopting the arriving or transmitting time of packets as carrier, these channels are more secure and effective. However, there is no general design for these covert behavior channels. The capacity of a covert channel without noise is intensively studied in [19,20]. These works usually do not take into account the noise in covert channel. However, there is a large amount of time-varying noise in actual covert channel systems that should not be neglected.

In this paper, to achieve a better trade-off between the security and effectiveness of the covert communication, we design a covert behavior channel by using operation sequences of network protocols. First, we systematically describe a typical model of covert behavior channel. Second, according to this model, we adopted the bi-directional communication protocol to reduce the error bit. By sharing a secret key between the transmitter and the receiver, covert information can be encrypted dynamically. Third, the closed form of capacity for the covert behavior channel with time-varying noise is derived based on Markov method. Finally, the robustness is analyzed for the covert behavior channel against the channel noise aroused by discarding packets.

Our contributions in this paper are listed as follows:

- We design the covert behavior channel by introducing the watermark bit embedding technology [21] used in multimedia steganography. Encryption method including replacement and scramble is used to keep security of the covert bit.
- We model the covert behavior channel with time-varying noise by using Markov chain. A dynamical channel capacity deriving method is adopted in terms of conditional probability of different system states, determined by channel noise.
- The definition of robustness is introduced to measure the stability of the covert information according to Wang and Reeves [22] when its carrier is attacked. The approach is introduced to derive the robustness of covert information by analyzing the successful probability of attacking on its carriers.
- Simulations show that the proposed covert behavior achieves a higher bit rate and better robustness than that of covert timing channel and is more undetectable than conventional covert storage ones.

We first present a design of a covert behavior channel and provide the details of its implementation in Section 3. In Section 4, the capacity, robustness, and security of the covert behavior channel are analyzed. The proposed schemes are evaluated in Section 5. Finally, the conclusion is given in Section 6.

2. BACKGROUND

Initially, the research on covert channels focuses on single systems [23,24]. However, with the advance of the network technology, covert channels become a pervasive security threat in the trusted distributed systems. Network covert channels have been used effectively by attackers to communicate with compromised hosts for stealing useful information. On the other hand, many practical tools are developed by employing covert channel to deliver the privacy information, such as session password, and authority information. Generally, these network covert channels are exploited with a variety of network protocols such as FTP, TCP, IP, and HTTP [17,18,25,26].

The covert storage channel, in general, adopts the unused header fields belonging to packets of these protocols that are designed for future protocol improvements to convey covert data[17,25]. For instance, The ID field in TCP and the option bits in IP have been used for storage channels [27]. Nevertheless, these covert storage channels can be detected easily by observing variations in unused packet header fields.

When it comes to covert timing channels, users can transmit covert information through the arrival patterns of packets, rather than through the contents of the packets themselves. The typical network timing channels is the packet sorting channels in which the order of packet arrival conveys information [10,28]. In addition, the user of timing channels can also make specific time intervals to carry covert information. However, these two kinds of covert timing channel can be effectively stopped by using a network pump. To countermeasure with the covert channel, Goldsmith and Varaiya [29] propose network pumps as intermediaries between networks terminals to filter network packets. Because the pumps can both modify packet inter-arrival times and reset the unused fields, we can use pump to defense against timing channels and storage ones. These defenses are aimed at stopping such channels rather than detecting them as shown in Figure 1.

To bypass the pump and detection device, instead of focusing on storage channels and timing channels, we propose the definition of the covert behavior channel, which conveys covert information by a sequence of operations. According to this definition, many network covert timing and storage channels exited can be considered as the covert behavior channels. For example, the network timing channels include packet sorting channels. The covert behavior

DOI: 10.1002/sec

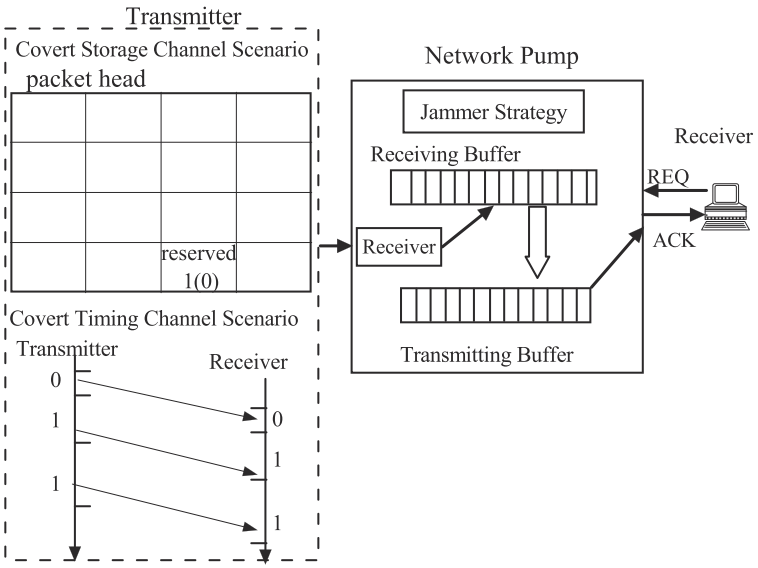

Figure 1. The network pump used to stop a covert storage channel and a covert timing channel. The jamming strategy in network pump is used to reset the reserved bits in the covert storage channel. The receiver buffer and transmitting buffer can work cooperatively to countermeasure with the covert channel timing.

channel can also be constructed based on arriving pattern of the packet, which includes packet losing, the sequence of arriving of network packets. Inspire by this idea, we design covert behavior channel with the sequence of network commands. From this perspective, covert behavior channels can be taken as a new covert channel, which has both features of a timing channel and a storage one.

3. DESIGN OF COVERT BEHAVIOR CHANNEL

3.1. System model

By bypassing network gateway, the covert channel can be used to transmit privacy information to the destination.

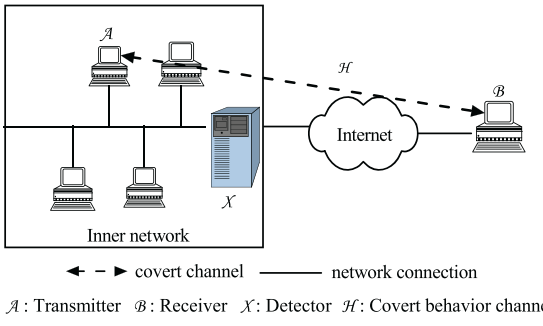

Figure 2. System model of network covert behavior channel. $\mathcal{A}$ is located in the inner network, and $\mathcal{B}$ is outside the inner network. The dashed line between $\mathcal{A}$ and $\mathcal{B}$ represents the covert channel that is realized via operation sequences of network protocol. $\mathcal{A}$ deliberately constructs the covert behavior to divulge information.

Figure 2 shows a typical network covert channel system. The transmitter $\mathcal{A}$ is supposed to be monitored by the detector $\mathcal{X}$, which filters each packet emitted from the network. The receiver $\mathcal{B}$ is located in the different network from $\mathcal{A}$. The covert behavior channel $\mathcal{H}$ is constructed by $\mathcal{A}$ to communicate with $\mathcal{B}$ for the purpose of bypassing the detector $\mathcal{X}$, which is shown as a dashed line in Figure 2.

In the covert behavior channel, behaviors of network users, such as downloading file, verifying accounts, or viewing web pages, are used as carriers to transmit secret information. These behaviors can be implemented by sequences of operations (or parameters) of network protocols, for example, Pop3. When the transmitter $\mathcal{A}$ exchanges information with the receiver $\mathcal{B}$ by this protocol, covert information carried by operation sequences is transmitted simultaneously. To do this, we install the client and server of the network protocol in $\mathcal{A}$ and $\mathcal{B}$, respectively. $\mathcal{A}$ communicates with $\mathcal{B}$ with selected sequences of operations (or parameters) of the network protocol to transmit secret information. Through decoding these operation sequences, $\mathcal{B}$ obtains the covert information.

3.2. Communication protocol design

In this section, we describe techniques used to maintain synchronization. Generally, the covert channel is unidirectional [22]. Because in these covert channels the receiver cannot echo any information to the transmitter, network errors, like losing packets and transmitting timeout, would cause an entire bit error of covert information. This is the primary reason for instability of the traditional covert channel.

To construct a stable communication protocol for the covert behavior channel, we design bi-directional

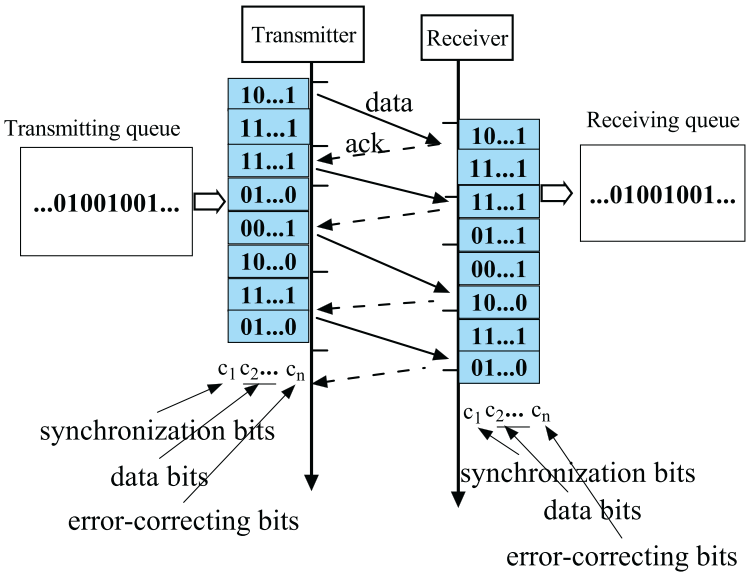

Figure 3. The bi-directional protocol of the covert behavior channel. Acknowledge packets are designed to inform the transmitter whether the receiver has received covert information correctly. A frame of covert information is composed of data bits, synchronization bits, and error-correcting bits.

communication protocol for covert behavior channel. As the unidirectional covert channel causes the loss of covert bit, we construct an acknowledge scheme. The receiver monitors each operation sequence received to determine the correctness of the covert information. Then, the receiver informs the covert transmitter whether the received information is correct by using acknowledging scheme. The acknowledging scheme is implemented by using the acknowledging operations of the network protocol. Taking transmission control protocol (TCP) for example, we select different acknowledge operations of TCP to inform the transmitter whether the covert information is received correctly, as shown in Figure 3.

In addition, the covert communication may not always accompany with the normal communication. There are only some particular periods used for transmitting covert information. Such a covert communication model is very effective for escaping from the monitor. However, this model also makes it more difficult for the receiver of the covert channel to be aware when the covert communication starts. To guarantee that the receiver can perceive the start and end of the covert communication, we specify two network behaviors to indicate the starting and ending of covert communications.

To reduce the spreading of the bit error, the covert message for transmission is subdivided into smaller blocks, called frames. An example frame consists of data bits, synchronization bits, and error-correcting bits.

3.3. Information embedding scheme

As it is known, sequences of system calls can indicate behaviors of users in the operation system. Similarly, the operation sequences of network protocols can also reflect the behaviors belonging to users. For instance, the behavior downloading a file in file transfer protocol (FTP) can be accomplished by the operation sequences: $\{CD, GET, BYE\}$. Our covert behavior channel exploits the operation sequences as the carrier.

To construct a covert behavior channel, we design the information embedding scheme for the covert transmitter as the following three steps. First, we specify a numeric table $W = \{w_1, w_2, w_3, \ldots, w_n\}$, where $w_i = i, 1 \leqslant i \leqslant n$. Each element in W is assigned with an operation sequence, which is used as the carrier of covert information. As shown in Figure 3, taking FTP as example, we construct the table $W = \{\{CD, GET, BYE\} = 1, \{CD, PUT, BYE\} = 2, \{CD, CD, BYE\} = 3, \ldots, n\}$. In this way, the operation sequence of the system is encoded with consecutive numbers. Second, we adopt $S = \{s_1, s_2, \ldots, s_{L-1}\}$ as the code set with L code words, where $s_i = i, 0 \leqslant i \leqslant L-1$. At last, we map covert information coding with the code words in S to the set W.

As shown in Figure 4, to map the symbol i ($i \in S$) to the element in W. We partition W into M subsets, whose length is equal to that of S. As a sequence, the length of the W should be deliberately designed as times the length of the S so that M can be guaranteed as an integer. Then, the mapping function $E(i, K)$ is adopted as

$$E\left(i, K_j\right) = \left(K_j - 1\right) \times L + O(i) \tag{1}$$

where K_j is the password used to ensure the security of the covert information, generated by

$$K_j = C_0 + R_j \tag{2}$$

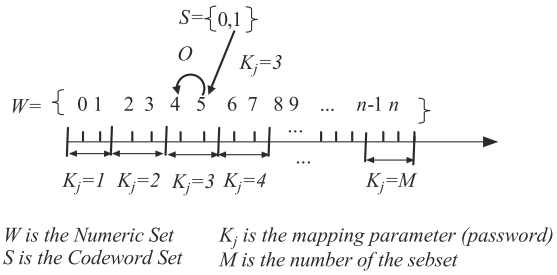

Figure 4. An example of information embedding scheme that maps 1 in S to 4 in W. K_j is used to select the subset set to be 3, and O is used to inverse the position of elements.

where C_0 is a constant, R_j is a pseudorandom sequence, and j is an integer. The rule for generating R_j is negotiated by the transmitter and receiver. $O(i)$ is given as

$$O(i) = (L-1) - i \bmod L \tag{3}$$

Therefore, $(K_j - 1) \times L$ in Equation (1) is used to find the subset in W for the symbol i.

Before covert communication, the transmitter of the covert channel sends the information of W, S, and the rule of generating K_j to the receiver, and the receiver is to decode the covert bit from received operation sequence. After receiving the signal, the receiver recovers it by a decoding function, given as

$$D\left(s, K_j\right) = O\left(s - \left(K_j - 1\right) \times L\right) \tag{4}$$

where s is the received signal.

For validating the correctness of the information embedding scheme, we give Theorem 1.

Theorem 1. *Given any $L > 0$, $K_j > 0$, and a covert signal* i, $D\left(E\left(i, K_j\right), K_j\right) = i$.

Proof. For any $i > 0$, we can find a unique integer $a \geqslant 0$ and a unique integer $0 \leqslant b \leqslant L$ satisfied

$$E\left(i, K_j\right) = a \times L + b \tag{5}$$

Let the received signal be s, then $s = E(i, K_j)$. According to (2) and (1), s can be given as

$$s = \left(K_j - 1\right) \times L + L - 1 - i \bmod L \tag{6}$$

Substituting (6) into (4) yields

$$D\left(s, K_j\right) = g\left(\left(K_j - 1\right) \times L + L - 1 - i \bmod L, K_j\right) \tag{7}$$

Let $a = K_j - 1$, $b = L - 1 - i \bmod L$. Combining (4) and (7) yields

$$D\left((a \times L + L - 1 - i \bmod L, K_j\right) = O\left(b\right) \tag{8}$$

According to (1), $O(b)$ can be obtained by

$$O(b) = L - 1 - (L - 1 - i \bmod L) \bmod L = i \tag{9}$$

Hence, $D(s, K_j) = i$. This completes the proof of Theorem 1. □

4. PERFORMANCE ANALYSIS

4.1. Capacity analysis

As mentioned in Section 1, most of the works on capacity analysis of the covert channel are performed under the assumption that the channel is noiseless. However, the network is a very complex system, in which network traffics and structures vary with time frequently. This leads to packet loss, packet retransmission, and disorder of arrival packets, which bring noises to covert channels. Even the covert channel designed with a connected protocol like FTP may also be affected by these noises. For example, the operator sequence $\{CD, GET, POST\}$ may be received as $\{CD, GET\}$ after the packet with the operation $POST$ lost.

We design the covert behavior based on the network protocol-based connected transport mode, which transmits packets according to the packet order. Then, we can extract the operation in terms of the sequence number of the TCP, and the operator sequence received would be ordered. Therefore, covert behavior channel cannot be interfered by the noise aroused by disorder of the packets. In addition, we adopt the coding method with the equal operation length (Section 3.3). As shown in Figure 5, if a packet is lost, w_i would be received as w_j. Thanks to the sequence number of the TCP packets, we can perceive the occurrence of the losing packet, thus not to take the first operation in w_{i+1} as the last one in w_i. As a result, the

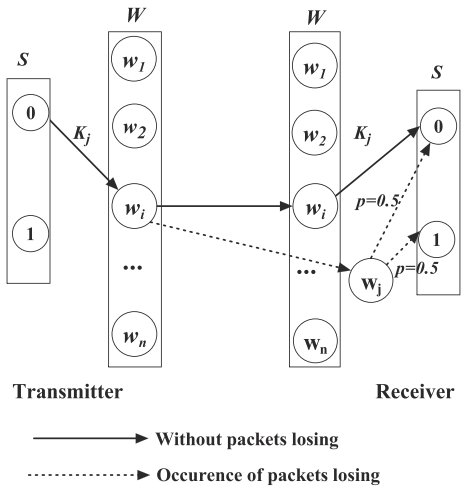

Figure 5. A noisy covert behavior channel model with the binary coding scheme. After the packet is lost, the received operation sequence w_j becomes shorter than that of the original one w_i.

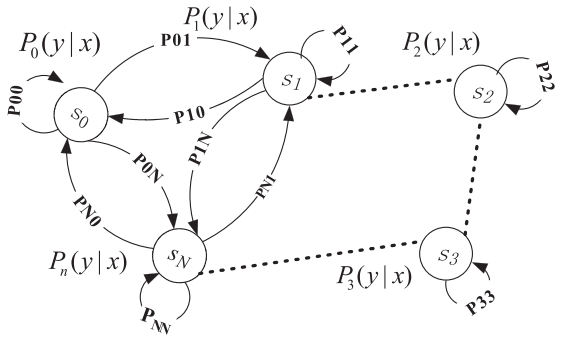

Figure 6. The model for the covert behavior using Markov chain with finite states.

length of the operation sequence w_j is less than that of the elements in W. In this case, The w_j can be interpreted as 0 or 1 with the same probability, which does not rely on the parameter K. Hence, the covert behavior channel with noise adopting binary coding scheme can be taken as a Z-channel.

However, the noise in the covert behavior is time varying, which leads to the changing of the bit error rate. Being different with the noise, the bit error rate is not time varying. It changes only if the noise shifts considerably. As a result, the noise can be graded to several levels. In terms of these levels, a covert channel system can be modeled in several states. In fact, the next state of the system relies on the previous and initial state. Furthermore, the transition probabilities of these states are independent of the input and output, and the covert behavior is a Z-channel. Thus, we model the covert behavior channel with time-varying noise as a Markov chain with finite states in the space $U = \{s_1, \ldots, s_N\}$. These states correspond to N different discrete memoryless channels, forming an irreducible, periodic, stationary Markov chain [29,30], as shown in Figure 6.

We define the state at time n of the Markov chain as Ψ_n, which is positive recurrent and ergodic. Let P be the matrix of transition probabilities for Ψ_n, so

$$P_{km} = p\left(\Psi_{n+1} = s_m | \Psi_n = s_k\right) \tag{10}$$

where s_k, s_m are states in U.

The input and output of the covert behavior at time n are denoted as x_n and y_n, respectively. Then, according to the conditional input/output probability at time n, we have

$$p\left(y_n | x_n, \Psi_n\right) = \sum_{k \in N} p_k\left(y_n | x_n\right) I\left(\Psi_n, s_k\right) \tag{11}$$

where $p_k(y|x) = p(y|x, \Psi = s_k)$, $I(\Psi_n, s_k)$ is an indicator function given as

$$I(\Psi_n, s_k) = \begin{cases} 1, \Psi_n = s_k \\ 0, \Psi_n \neq s_k \end{cases} \tag{12}$$

We use the notation $r^n \triangleq (r_1, ..., r_n)$ for $r = x, y$ or Ψ. Because the state at time n+1 is independent of previous input/output pairs when conditioned on Ψ_n, we have

$$p\left(\Psi_{n+1} | \Psi_n, x^n, y^n\right) = p\left(\Psi_{n+1} | \Psi_n\right) \tag{13}$$

Because these channels are memoryless, then

$$p\left(y_{n+1} | \Psi_{n+1}, x_{n+1}, \Psi^n, x^n, y^n\right) = p\left(y_{n+1} | \Psi_{n+1}, x_{n+1}\right) \tag{14}$$

If we assume that the x_n's are independent of the past input/output, it follows that

$$p\left(y_{n+1}, x_{n+1} \middle| \Psi_{n+1}, \Psi^n, x^n, y^n\right) = p\left(y_{n+1}, x_{n+1} \middle| \Psi_{n+1}\right) \tag{15}$$

From (13) and (15), we obtain

$$p\left(y^N, x^N \middle| \Psi^N\right) = \prod_{n=1}^{N} p\left(y_n, x_n \middle| \Psi_n\right) \tag{16}$$

Similarly, from (14) and (15) we can obtain

$$p\left(y_{n+1} \middle| \Psi_{n+1}, \Psi^n, y^n\right) = p\left(y_{n+1} \middle| \Psi_{n+1}\right) \tag{17}$$

We denote conditional state distributions by the N-dimensional random vectors $\pi_n = (\pi_n(1), ..., \pi_n(N))$ and $\boldsymbol{\rho}_n = (\rho_n(1), ..., \rho_n(N))$, respectively, where

$$\rho_n(k) = p\left(\Psi_n = s_k \middle| y^{n-1}\right) \tag{18}$$

$$\pi_n(k) = p\left(\Psi_n = s_k \middle| x^{n-1}, y^{n-1}\right) \tag{19}$$

Then, the capacity of covert behavior channel with time-varying noise is given by Theorem 2.

Theorem 2. *A time-varying noise covert behavior channel, with common finite input and output alphabets denoted by* X *and* Y*, which capacity is given as*

$$C = \frac{1}{n} \sum_{i=1}^{n} \left(-\log \sum_{l=1}^{N} p\left(y_i | \Psi_i = s_l\right) \rho_i(l) + \log \sum_{l=1}^{N} p_l\left(y_i | x_i\right) \pi_i(l) \right) \tag{20}$$

where N *is the number of state and* x_i *and* y_i *are the elements in* X *and* Y.

Proof. Let X^n and Y^n be the input and output sequence of the covert behavior channel. Hence, X^n can be defined as the set $\{x_1, x_2, \ldots, x_n\}$, in which all $x_i, 1 \leqslant i \leqslant n$ has N possibility. Given any X^n and Y^n, there is a mutual information $I(X^n; Y^n)$, which is the amount of information transmitted by the channel. Thus, if a X^n is chosen from

DOI: 10.1002/sec

all the input sequence that leads to the mutual information achieving its maximum value, then a maximum of the mutual information can be considered as the capacity of the covert channel, given by

$$C = \lim_{n\to\infty} \max_{\Gamma(X^n)} \frac{1}{n} I\left(X^n; Y^n\right) \tag{21}$$

where $\Gamma(X^n)$ is the set of all input distributions on X^n, $I(X^n; Y^n)$ is mutual information, which can be expressed as

$$I\left(X^n; Y^n\right) = H\left(Y^n\right) - H\left(Y^n|X^n\right) \tag{22}$$

where $H(Y) = E(-\log p(y))$ and $H(Y|X) = E(-\log p(y|x))$. By using information theory, we have

$$H\left(Y^n\right) = \sum_{i=1}^{n} H\left(Y_i|Y^{i-1}\right) \tag{23}$$

$$H\left(Y^n|X^n\right) = \sum_{i=1}^{n} H\left(Y_i|X_i, Y^{i-1}, X^{i-1}\right) \tag{24}$$

Substituting (23), (24) in (22) yields

$$C = \lim_{n\to\infty} \max_{\Gamma(X^n)} \frac{1}{n} \left(\sum_{i=1}^{n} H\left(Y_i|Y^{i-1}\right) - \sum_{i=1}^{n} H\left(Y_i|X_i, Y^{i-1}, X^{i-1}\right)\right) \tag{25}$$

Because the definition of the entropy is

$$H\left(Y_i|X_i, X^{i-1}, Y^{i-1}\right) = E\left(-\log p\left(y_i|x_i, x^{i-1}, y^{i-1}\right)\right) \tag{26}$$

In terms of (11) and (24), we have

$$E\left(-\log p\left(y_i|x_i, x^{i-1}, y^{i-1}\right)\right) = E\left(-\log \sum_{l=1}^{N} p\left(y_i|x_i, \Psi_i = s_l\right) p\left(\Psi_i = s_l|x^{i-1}, y^{i-1}\right)\right) \tag{27}$$

Also, by (16), (17), and (23), $H(Y_i|Y^{i-1})$ can be expressed as

$$H\left(Y_i|Y^{i-1}\right) = E\left(-\log \sum_{l=1}^{N} p\left(y_i|\Psi_i = s_l\right) p\left(\Psi_i = s_l|y^{i-1}\right)\right) \tag{28}$$

Substituting (27), (28) in (25) yields

$$C = \lim_{n\to\infty} \max_{p(X^n)} \frac{1}{n} \sum_{i=1}^{n} \left(E\left(-\log \sum_{l=1}^{N} p(y_i|\Psi_i = s_l)\rho_i(l)\right) - E\left(-\log \sum_{l=1}^{N} p(y_i|x_i, \Psi_i = s_l)\pi_i(l)\right)\right) \tag{29}$$

because $\rho_i(l)$ and $\sum_{l=1}^{N} p(y_i|\Psi_i = s_l)$ are constants, and

$$E\left(-\log \sum_{l=1}^{N} p\left(y_i|\Psi_i = s_l\right)\rho_i\left(l\right)\right) = \left(-\log \sum_{l=1}^{N} p\left(y_i|\Psi_i = s_l\right)\rho_i\left(l\right)\right) \tag{30}$$

Also,

$$E\left(-\log \sum_{l=1}^{N} p(y_i|x_i, \Psi_i = s_l)\pi_i(l)\right) = -\log \sum_{l=1}^{N} p(y_i|x_i, \Psi_i = s_l)\pi_i(l) \tag{31}$$

Thereby, Equation (20) holds. This completes Theorem 2.

Theorem 2 gives a method to estimate the capacity of the covert behavior channel. In Equation (20), the computations of π_l and ρ_l are key steps. In what follows, we give a recursive method to calculate them. According to the Bayes ruler, we have

$$p\left(\Psi_i|x^i, y^i\right) = \frac{p\left(y_i|\Psi_i, x_i\right) p(x_i|x^{i-1}) p\left(\Psi_i|x^{i-1}, y^{i-1}\right) p\left(x^{i-1}, y^{i-1}\right)}{p\left(x^i, y^i\right)} \tag{32}$$

In (32), $p\left(x^i, y^i\right)$ can be given as

$$p\left(x^i, y^i\right) = \sum_{l=1}^{N} p\left(x^i, y^i, \Psi_i = s_l\right) \tag{33}$$

By (11) and (14), it follows that

$$p\left(x^i, y^i\right) = \sum_{l=1}^{N} p\left(y_i|x_i\right) p\left(x_i|x^{i-1}\right) p\left(\Psi_i = s_l|x^{i-1}, y^{i-1}\right) \cdot p\left(x^{i-1}, y^{i-1}\right) \tag{34}$$

Substituting (34) in (32) yields

$$p\left(\Psi_i|x^i, y^i\right) = \frac{p\left(y_i|\Psi_i, x_i\right) p\left(\Psi_i = s_j|x^{i-1}, y^{i-1}\right)}{\sum_{j=1}^{N} p\left(y_i|x_i, \Psi_i = s_j\right) p\left(\Psi_i = s_j|x^{i-1}, y^{i-1}\right)} \tag{35}$$

For a particular state l, we have

$$\pi_{i+1}(l) = \frac{\sum_{j=1}^{N} p_j\left(y_i|x_i\right) \pi_i(j) P_{jl}}{\sum_{l=1}^{N} p_l\left(y_i|x_i\right) \pi_i(l)} \tag{36}$$

Similarly, we can obtain the recursive formula to compute ρ

$$\rho_{i+1}(l) = \frac{\sum_{j=1}^{N} p_j\left(y_i|\Psi_i = s_j\right) \rho_n(j) P_{jl}}{\sum_{l=1}^{N} p_l\left(y|\Psi_i = s_j\right) \rho_n(l)} \tag{37}$$

The initial value for π and ρ can be given as

$$\pi_0 = \left[p(\Psi_0 = s_1), p(\Psi_0 = s_2), ..., p(\Psi_0 = s_N)\right] \tag{38}$$

□

4.2. Robustness analysis

The robustness of the embedded information bit is the ability to protect the covert information from interferences with the carrier objects. The bit error of covert information mainly comes from the transmission errors, such as packet losing or timeout. Let X be the covert bit sent and X^i be the signal received. Then, we define robustness of X as the probability $Pr\left(X^i - X = 0\right)$. Then, we obtain Theorem 3.

Theorem 3. *Given a covert behavior channel, after attacked by losing packets, the lower bound and upper bound of the robustness for a covert information are* $\left(1 - \frac{2}{N}\right) + \frac{2}{N} \times \frac{1}{L}$ *and 1, respectively.*

Proof. According to Sections 3.3 and 4.1, the robustness of covert information is determined by whether the carrier can be attacked successfully by discarding packets. As shown in Figure 4, by taking w_i as the carrier of the covert bit, the probabilities that w_i is attacked successfully are shown in Table I.

In Table I, the second column is the permutation number of remaining operations in w_i after discarding packets, and the third column is the probability that w_i is attacked successfully. N is the total number of the operation sequences, and L is the length of the operation sequence. Thus, the average probability that w_i is attacked successfully can be written as

Table I. Robustness of the covert behavior channel.

Losing packet numbers	Permutations	P_{r_suc}
1	$(L-1)*N$	L/N^L
2	$2*(L-2)*N^2$	$L*(L-2)/N^L$
3	$3*(L-3)*N^3$	$L*(L-3)/N^L$
...	...	...
L-2	$(L-2)*1*N^{L-2}$	$(L-2)*2/N^L$

$$\begin{aligned} I &= \frac{\frac{N}{N^L} C_L^1 + \frac{N^2}{N^L} C_L^2 + \cdots + \frac{N^{L-1}}{N^L} C_L^{L-1}}{L-1} \\ &= \frac{N C_L^1 + N^2 C_L^2 + \cdots + N^{L-1} C_L^{L-1}}{N^L \times (L-1)} \end{aligned} \tag{39}$$

According to binomial theorem, we have

$$I = \frac{(N+1)^L - N^L - 1}{N^L (L-1)} \tag{40}$$

where $L \geqslant 2$. let $f(x) = \frac{(N+1)^x - N^x - 1}{N^x(x-1)}$, then

$$f'(x) = \frac{N^x\left(-(1+N)^x + N^x - 2N + 1\right)}{\left(N^x(x-2)\right)^2} + \frac{N^x \log N\,(xN + x - 2)}{\left(N^x(x-2)\right)^2} < 0 \tag{41}$$

Assuming $x = 2$, we can obtain the maximum of the robustness $I = \frac{2}{N}$. When x approaches infinity, the limitation of I can be expressed as

$$\lim_{x \to \infty} \frac{(N+1)^x - N^x - 1}{N^x (x-1)} = 0 \tag{42}$$

After w_i is successfully attacked, it will be received as w_j as shown in Figure 4. According to the definition of the robustness, we can conclude that the upper bound of the robustness is 1 and the lower bound is

$$Pr\left(X^i - X = 0\right) = \left(1 - \frac{2}{N}\right) + \frac{2}{N} \times \frac{1}{L} \tag{43}$$

This completes Theorem 3. □

4.3. Security analysis

To detect the covert channel, a widely adopted approach is introduced by Steven in [13] to recognize covert network flows with corrected entropy. In this approach, the change of the corrected entropy provides a critical clue for covert channel detection. The detection idea can be described by

$$CCE\left(X_m|X_{m-1}\right) = CE\left(X_m|X_{m-1}\right) + perc\left(X_m\right) EN\left(X_1\right) \tag{44}$$

where CCE is the corrected conditional entropy, EN and CE represent the estimates of the entropy and conditional

DOI: 10.1002/sec

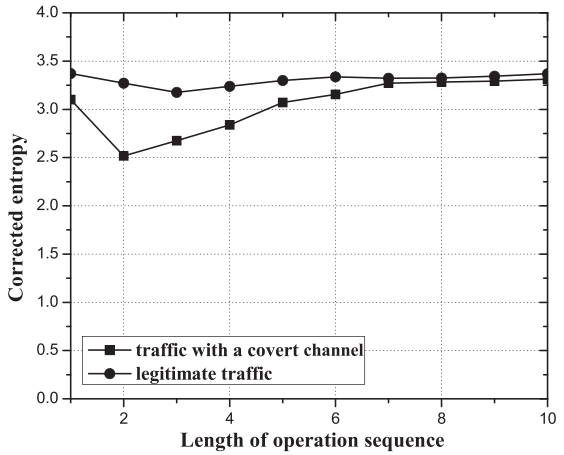

Figure 7. The corrected entropy of the behaviors based on file transfer protocol with $m = 5$ and $w = 2000$.

entropy, and m is the operation sequence length, which means the number of the operators belonging in the sequence, and X_m is the number of unique patterns with length m, and $perc(X_m)$ is the percentage of pattern X_m in all pattern. The minimum of the corrected conditional entropy is considered to be the best estimate of the entropy rate with the available data.

For analyzing the security of the behavior channel towards this detection approach, we conduct our detection experiment. Figure 7 shows the behavior of covert traffic generated by our proposed method along with the legitimate traffic observed. The corrected entropy curve of covert traffic is nearly identical to that of legitimate traffic except that the length of operator sequence is 2. If the sequence length is set as 5 or more than 5, the detector fails to discriminate a covert traffic from the legitimate traffic. As a result, it is very difficult for the algorithm to differentiate the traffic containing covert channel from legitimate traffic.

5. SIMULATION RESULTS

The simulation environment is set up according to Figure 2. In the covert behavior channel, four behaviors are chosen as carriers, which are downloading files, uploading files, transmitting files, and changing direction. To produce a significant amount of operation sequences, we select about one million FTP packets as samples from the data set NZIX-II [31]. Then, we select the operation from the data set to comprise the four different behaviors. A 0.5-M file is specified as the covert information. To compare the performance of covert behavior with the representative covert channels, we also implement the covert channel covert TCP according to [32] and IP covert timing channel (IPCTC) according to [7], shown as Figure 8.

In order to construct a noisy covert behavior channel, we inject noise data chosen from the mentioned data set

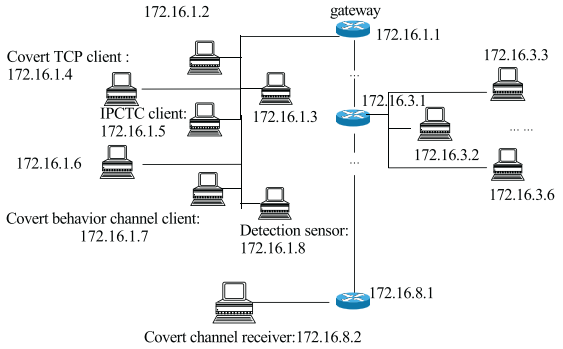

Figure 8. Simulation environment of the different covert channels.

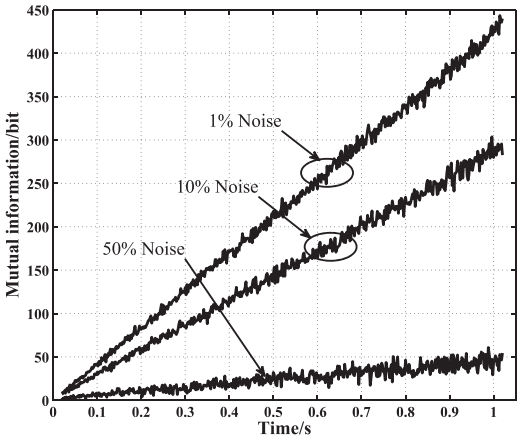

Figure 9. Capacity of the covert behavior channel with different noises.

into the channel. We break the covert behavior channel into blocks of 300 packets and randomly replace blocks of the covert traffic with the non-covert traffic until we achieve three of our desired noise levels, which are no noise, 20% noise, and 40% noise. Therefore, there are three noise states in our system. We mix packets of these three states randomly to construct three testing environments, in which proportions of the noise are 1%, 10%, and 50% to indicate the communication environments with low noise, middle noise, and heavy noise. We, first, probe the approximative probability of occurrence of the three noise states (0, 20%, and 40%) in our laboratory, respectively. Then, the transmitting probabilities of states are recorded as the matrix of transition probabilities, as mentioned in (10). we arrange the proportion of packets, which satisfy the matrix of transition probabilities constructed, to used as the input of our covert behavior channel. Figure 9 shows the capacity curves of the proposed covert behavior channel. These curves are obtained by the approximate expression given by (21). We can notice that with the increase of the noise, the performance of capacity becomes worse.

Figure 10 shows bit rate curves of the covert behavior channel, IPCTC, and covert TCP in the network environ-

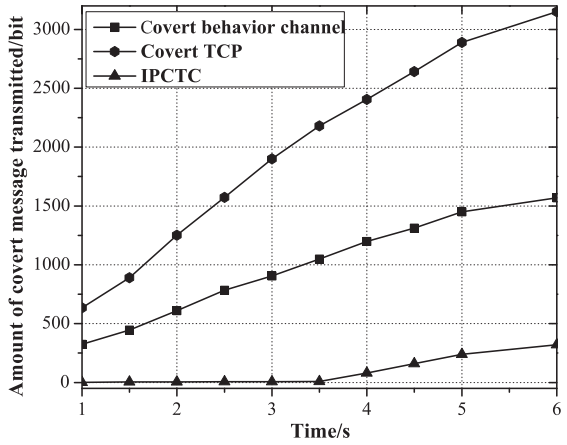

Figure 10. Achieved bit rate of three covert channels with 10% noise. TCP, transmission control protocol; IPCTC, IP covert timing channel.

ment with 10% noise. From Figures 9 and 10, it follows that the analytical results and the simulation results are consistent, which validate the correctness of our derivation. According to Figure 10, the bit rate of covert TCP outperforms that of our proposed covert behavior channel. The reason is that we need several packets to transmit a covert bit in covert behavior channel. However, in the covert TCP, a packet can carry one covert bit. We can also notice that bit rates of the covert behavior channel and covert TCP is much larger than that of the covert timing channel, which is because the bit rate of timing channel is restricted from the time interval of arriving packets.

To evaluate the complexity of algorithms of constructing the given three covert channels, we adopt the total time of computation as the indicator. A thread is installed in the each node (transmitting node and receiving node) to count the time required for embedding and extracting the covert information. We collect the time from each node, and the results are listed as Table II.

From Table II, we notice that IPCTC and covert behavior channel need less time for a transmission. The reason is that the implement of a covert TCP generally involves the link level programming, which requires packets to be moved from one buffer to another thus bringing time delay to the transmission.

We construct the scenario to evaluate the robustness of the covert behavior channel towards discarding packet . All network packets from client $\mathcal{A}$ to the server $\mathcal{B}$ are stored and forwarded by $\mathcal{X}$. After receiving packets from $\mathcal{A}$, $\mathcal{X}$ delays and discards them randomly. We run covert TCP, IPCTC, and the covert behavior channel for 20 times to transmit files respectively. Figure 11 shows the average bit error rate acquired in $\mathcal{B}$ of the three channels. The impact of the random delays and packet losing on the IPCTC is much larger than that of the covert behavior channel and Covert TCP. By analyzing log files, we notice that one bit is error in IPCTC will cause error of the successive bits. Furthermore, robustness performance of the covert behavior channel is better than that of Covert TCP, which is attributed to the synchronization and the error-rate correcting scheme.

Table III shows detection rate of two detection algorithms, the entropy-based approach and Kolmogorov–Smirnov test, on different covert channels with a 2000-packet detection window. In terms of the performance on detection rate and false report rate, we select the threshold $KSTest > 0.35$ for the Kolmogorov–Smirnov tester according to [13] and $CCE = 1.5, 2.2$ and 1.4 to detect the three covert channels for the entropy-based tester. The recognition rate of IPCTC and covert TCP is more than 90%. The traffic distribution of an IPCTC is different with that of the legitimate traffic, which makes it easy to be identified by entropy-based detector and Kolmogorov–Smirnov tester. In a covert TCP channel, positions for storing covert information in a packet are set to be default values in legitimate traffic. Hence, these two detectors work very well in recognizing the channel. However, these two algorithms fail in finding a covert behavior channel because the distribution

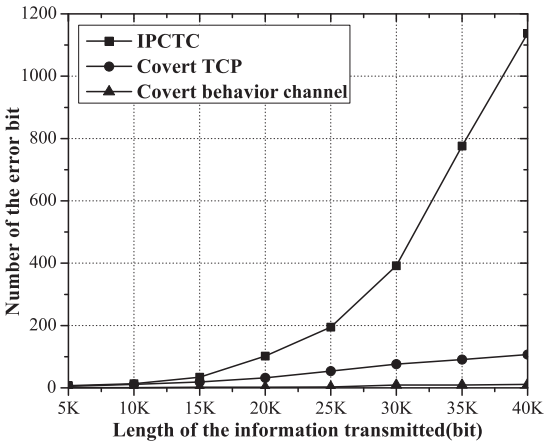

Figure 11. Robustness of covert channels against the attacking of discarding packet. TCP, transmission control protocol; IPCTC, IP covert timing channel.

Table II. Complexity comparison of constructing covert channels.

Covert channel	Covert bit/transmission	Time/transmission (ms)
Covert behavior channel	0.33	1.83
IPCTC	1	1.88
Covert TCP	1	2.89

IPCTC, IP covert timing channel; TCP, transmission control protocol.

Table III. Detection result of the covert behavior.

	Entropy-based test		Kolmogorov–Smirnov test	
	Accuracy (%)	False report (%)	Accuracy (%)	False report (%)
Proposed covert behavior channel	73.5	16	66	21
Covert TCP	97.4	9	94	14
IPCTC	92.1	14	90.1	16

IPCTC, IP covert timing channel; TCP, transmission control protocol.

of covert traffic is close to that of the legitimate one, which is caused by the encryption method.

6. CONCLUSION

In this paper, we present a network covert behavior channel. The operation sequence of the network protocols is used to transmit secret information. To ensure the security of covert behavior channel, modulation algorithm with encryption is designed. A Markov model is used to derive the capacity of covert behavior channel with time-varying noise. Compared with the covert timing channel, the capacity of covert behavior is improved significantly. The quantity analysis on robustness is conduct on the covert behavior channel under the attacking of discarding packets. Moreover, the experimental method proves that the covert channel is undetectable to the effective covert channel detector. Evaluation results show that the covert behavior channel is secure and effective.

ACKNOWLEDGEMENTS

This work was supported in part by open research fund of National Mobile Communications Research Laboratory, Southeast University (grant no. 2013D02), the Fundamental Research Funds for the Central Universities (grant no. 30920130122004), Natural Science Foundation of Jiangsu Province(grants no. BK20131353), and the National Natural Science Foundation of China (grant nos. 61271230, 61472190, and 61301107).

REFERENCES

1. Snoeren A, Partridge C, Sanchez L. Single packet IP trace back. *ACM/IEEE Transaction on networking* 2002; **10**(6): 721–734.
2. Peng P, Ning P, Reeves DS. On the secrecy of timing-based active watermarking trace-back techniques, 2006.
3. Lampson BW. A note on the confinement problem. *Communications of the ACM* 1973; **16**(10): 613–615.
4. Zander S, Branch P, Armitage G. A survey of covert channels and countermeasures in computer network protocols. *IEEE Communications Surveys and Tutorials* 2007; **9**(3): 44–57.
5. Kiyavash, Koushanfar N, Coleman F, Rodrigues TP. A timing channel spyware for the CSMA/CA protocol. *IEEE Transactions on Information Forensics and Security* 2013; **8**(3): 477–487.
6. Zi X, Yao L, Jiang X, Pan L, Li J. Evaluating the transmission rate of covert timing channels in a network. *Computer Networks: The International Journal of Computer and Telecommunications Networking* 2011; **55**(12): 2760–2771.
7. Cabuk S, Brodley CE, Shields C. IP covert channel detection. *Transactions on Information and System Security of ACM* 2009; **12**(4): 1–22.
8. Sarah S, Wang H, Chun C, Ness S, Bagchi, Saurabh. Capacity bounds on timing channels with bounded service times. *IEEE International Symposium on Information Theory*, Nice, France, 2007; 981–985.
9. Shen Y, Huang L, Lu X, Yang W. A novel comprehensive steganalysis of transmission control protocol/internet protocol covert channels based on protocol behaviors and support vector machine. *Security and Communication Networks* 2015; **8** (7): 1279–1290.
10. Ahsan K, Kundur D. Practical data hiding in TCP/IP. *IET-CIRED Seminar on Smart Grids for Distribution*, France, 2002; 7–14.
11. Wolf M. Covert channels in LAN protocols. In *Local Area Network Security*. Springer: Karlsruhe, FRG, 1989; 91–101.
12. Handel T, Sandford M. Hiding data in the OSI network model. *IEEE Proceeding of 1st Int'l Wksp of Information Hiding*, Cambridge,UK, 1996; 23–38.
13. Gianvecchio S, Wang H. An entropy based approach to detect covert timing channels. *IEEE on Dependable and Secure Computing* 2011; **8**(6): 785–797.
14. Zhao H, Shi Y. Detecting covert channels in computer networks based on chaos theory. *IEEE on Information Forensics and Security* 2013; **8**(2): 273–283.
15. Schulz S, Bochum R, Varadharajan V, Sadeghi AR. The silence of the LANs: efficient leakage resilience for IPsec VPNs. *IEEE Transactions on Information Forensics and Security* 2014; **9**(2): 221–232.
16. Wang F, Huang L, Miao H, Tian M. A novel distributed covert channel in HTTP. *Security and Communication Networks* 2014; **6**(7): 1031–1041.

17. Crespi V, Cybenko G, Giani A. Engineering statistical behaviors for attacking and defending covert channels. *IEEE Journal of Selected Topics in Signal Processing* 2013; **7**(1): 124–136.
18. Zou X. *The Research on Information Hiding Based on Command Sequence of FTP Protocol*. Springer-Verlag: Berlin, Germany, 2005.
19. Zander S, Armitage G, Branch P. Capacity of temperature-based covert channels. *IEEE on Communications Letters* 2011; **15**(1): 82–84.
20. Moskowitz IS, Kang MH. Covert channels-here to stay. *Proceedings of the Ninth Annual Conference on Computer Assurance*, Gaithersburg, MD, 1994; 235–244.
21. Cox I, Miller M, Bloom J. *Digital Watermarking*. Morgan Kaufmann Publishers: San Francisco, California, 2002.
22. Wang X, Reeves D. Robust correlation of encrypted attack traffic through stepping stones by flow watermarking. *IEEE Transactions on Dependable and Secure Computing* 2011; **8**(3): 434–449.
23. McHugh J. Covert channel analysis, 1995.
24. Porras PA, Kemmerer RA. Covert flow trees: a technique for identifying and analyzing covert storage channels. *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, Oakland, CA ,California, 1991; 36–51.
25. Dyatlov A, Castro S. Exploitation of data streams authorized by a network access control system for arbitrary data transfers: tunneling and covert channels over the HTTP protocol, 2003.
26. Smith JC. Covert shells. *Sans Institute Information Security Reading Room*, 2000.
27. Giffin J, Greenstadt R, Litwack P, Tibbetts R. Covert messaging through TCP timestamps. *Workshop on Privacy Enhancing Technologies* 2002; **2482**: 194–208.
28. Ahsan K. Covert channel analysis and data hiding in TCP/IP. *Master's thesis*, University of Toronto, 2000.
29. Goldsmith AJ, Varaiya PP. Capacity mutual information and coding for finite-state Markov channels. *IEEE Transaction on Information Theory* 1996; **42**(3): 868–886.
30. Lin S, Li Y, Li Y, Ai B, Zhong Z. Finite-state Markov channel modeling for vehicle-to-infrastructure communications. *IEEE Communications 6th International Symposium Wireless Vehicular (WIVEC)*, Vancouver, BC, Canada, 2014; 1–5.
31. Lippmann R, Haines J, Fried DJ, Korba J, Das K. The 1999 DARPA off-line intrusion detection evaluation. *Computer Networks* 2000; **34**(4): 579–595.
32. Rowland C. Covert channels in the TCP/IP protocol. *First Monday:Peer-Reviewed Journal on the Internet* 1997; **2**(5): 1–5.