COVERT TIMING CHANNELS CODES FOR COMMUNICATION OVER INTERACTIVE TRAFFIC

Negar Kiyavash

Dept. of Computer Science
Information Trust Institute
University of Illinois
kiyavash@illinois.edu

Todd Coleman *

Dept. of Electrical and Computer Eng.
Coordinated Science Laboratory
University of Illinois
colemant@illinois.edu

ABSTRACT

This paper presents the first practical perfectly-secure steganography codes for covert communication via packet timings across interactive traffic relayed over network queuing systems. It has recently been shown that sparse-graph linear codes followed by shaping techniques, combined with message-passing decoding, can enable practical timing channel codes with low symbol error rates near the information capacity of the famous "Bits Through Queues" channel. Inspired by this new class of codes, we use an alternative shaping technique that employs random dithers and construct provably secure steganographic codes for communication using packet timings in interactive traffic. To validate the perfect secrecy of our steganographic codes, we model interactive traffic as a two-state Markov Modulated Poisson Process (MMPP) and show its goodness-of-fit.

Index Terms— Covert Communication, Steganography, Timing Channels, Interactive Traffic

1. INTRODUCTION

Historically, timing channels are synonymous with covert channels [14, 10]. Covert channels are mechanisms for communicating information in ways that are difficult to detect. Packet networks are designed with the goal of communicating through packet contents and their headers; hence, the timing channel induced by the inter packet timings provides a side channel that can be utilized for covert communication. Figure 1 illustrates a timing covert channels between two parties. The eavesdropper (a.k.a wire tapper) sees the exchange of packets but fails to realize the covert communication in packet timings. Besides covert communication, recently a host of new security applications have arisen where it is desired to communicate - not by means of packets contents - but by utilizing the inter packet timings [12, 15].

*Both authors would like to acknowledge the support of NSF through grants CCF 07-29061 and CNS 08-31488.

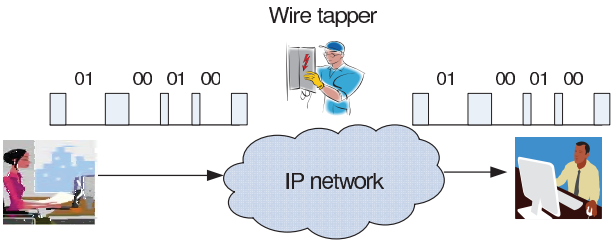

Fig. 1. Covert communication: An eavesdropper inspects the packet contents, but is unable to decode messages modulated by the packet timings.

This paper discusses a practical implementation of perfectly-secure steganographic codes for queuing channels, where the covertext is interactive traffic (e.g. an SSH flow). Here, we consider a communication channel where the encoder communicates covert information based upon timings between successive packets over an interactive traffic session. A receiver observes packet timings after they have traveled through a communication network with queues at intermediate router nodes. Based upon the encoding mechanism, the statistical structure of the network queues, and the packet timings it observes, the receiver decodes the covert message. In the famous 'Bits Through Queues" result, Anantharam & Verdu characterized - in closed form - the capacity a channel where a single server queue with exponential service times is placed between the transmitter and receiver [1]. Recently, a class of low complexity codes that approach capacity of such channels were introduced [6]. Building upon the work in [6], we propose the first perfectly-secure class of steganographic codes with low decoding complexity for communication over queuing channels.

1.1. Summary of Results and Organization

The main contribution of this paper is the introduction and practical implementation of perfectly-secure steganographic codes [4, Definition 1] for queuing channels, where the covert

text is interactive traffic (e.g. SSH flow). More precisely, we consider a communication channel where the encoder (a user typing over an interactive session such as SSH) communicates with a receiver. The encoder conveys covert information by modifying timings between successive packets that act as covertext. A receiver observes packet timings after they have traveled through a communication network with queues at intermediate router nodes.

In Section 2, we present a model for interactive traffic, as it is essential to the analysis of the perfect secrecy of our steganographic codes. Specifically, we model the interactive connection as a two-state Markov-modulated Poisson process (MMPP) [7], where one state corresponds to a user typing characters and the other state corresponds to periods of silence. As the receiver is involved in the interactive session, it is aware of the states of the MMPP. In Subsection 2.1, we show statistically that our two-state MMPP indeed fits the interactive traffic well. Once the underlying distribution of the interactive traffic is parameterized, we use a novel dither-based technique to shape our codes according to the two-state MMPP which in turn guarantees that our code will have perfect secrecy. In Section 3 we present our code construction which is based on the low-complexity codes of [6]. The security of our steganographic codes is discussed in Section 3.1. Finally, in Section 4, we demonstrate the performance of our scheme with a linear complexity decoder on simulated queuing channels.

To summarize: 1) we model interactive traffic as a two-state Markov Modulated Poisson Process (MMPP) and validate its goodness-of-fit statistically; 2) using an alternative shaping technique that employs random dithers, we construct provably good steganographic codes for communication using packet timings in interactive traffic; 3) we demonstrate the performance of our codes in terms of symbol error rates over simulated queuing channels.

2. MODEL OF INTERACTIVE TRAFFIC

We first present a model for interactive traffic, as it is essential to the analysis of perfect-secrecy of our steganographic codes. Given that the traffic might be encrypted (e.g. SSH traffic), in modeling interactive traffic, we do not consider the content of the packets; likewise, the sizes of packets representing keystrokes are likely to be uniform. We thus consider only the arrival time of the packets in the flow, allowing us to model the flow as a point process.

Suppose we observed packet arrivals at times $t_1 < t_2 < \cdots < t_n$ in a fixed interval $(0, \tau]$ such that t_i is the time the i-th packet arrived. The collection of arrival times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ specifies a flow f. Furthermore, we model the interactive connection as a Markov-modulated Poisson process (MMPP) [7]. The set of possible states are $\{0, 1\}$, where state 0 corresponds to user typing characters and state 1 corresponds to periods of silence, when the user is expecting a response from the receiver. Figure 2 depicts this two-state MMPP.

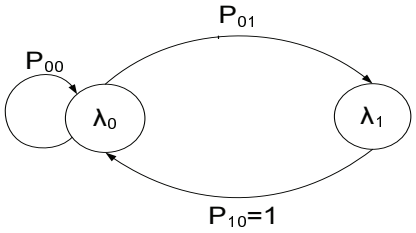

Fig. 2. The embedded two-state Markov chain.

When the process is in state 0, packet arrivals are modeled as a Poisson process of rate λ_0; i.e. the inter-arrival times $\{Z_i\}$ are independent and identically distributed (i.i.d.) according to an exponential distribution with rate λ_0:

$$f_Z(z) = \lambda_0 e^{-\lambda_0 z}.$$

When the process is in state 1, the arrivals are again modeled as Poisson but with rate $\lambda_1 < \lambda_0$. Given that state 1 corresponds to a period of silence (no packet arrivals), as soon as a packet arrives, the embedded Markov chain transitions to state 0. Therefore, the transition probabilities $\{P_{ij}, i, j = 0, 1\}$ of the embedded Markov chain $\{\phi_n, n \geq 0\}$ are as follows:

$$\begin{aligned} P_{00} + P_{01} &= 1, \\ P_{01} = 1, P_{11} &= 0 \end{aligned} \tag{1}$$

and the embedded Markov chain is defined by the matrix:

$$\begin{bmatrix} P_{00} & 1 \\ 1 - P_{00} & 0 \end{bmatrix}$$

The steady state probabilities π_0, π_1 of the embedded chain ϕ_n are given by:

$$\pi_0 = \frac{1}{2 - P_{00}}, \qquad \pi_1 = \frac{1 - P_{00}}{2 - P_{00}} \tag{2}$$

2.1. Parameter Selection and Goodness of Fit

We estimated the parameters P_{00}, λ_0, and λ_1 of our MMPP model by using network traces of SSH connections taken at a wireless access point in our institution. For a trace, we estimate the state transition probabilities, the underlying state sequence of the embedded Markov chain, and the corresponding rates λ_0 and λ_1 of each state using the EM algorithm [13]. Our estimated values for the transition probability P_{00} and the rates λ_0 and λ_1 were as follows:

$$P_{00} = .96 \qquad \lambda_0 = 5.6 \qquad \lambda_1 = 0.57 \tag{3}$$

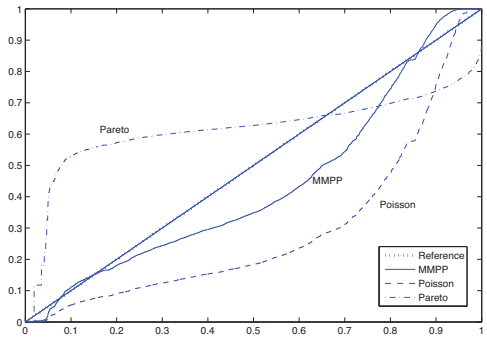

Fig. 3. Q–Q plot of Poisson and MMPP models with our sample data.

To assess the goodness of fit of our MMPP model with parameters of (3), we used a quantile–quantile (Q–Q) plot [3]. Using the theoretical CDF of the model, the observations are mapped into values in interval $[0, 1]$. If the underlying statistical model of the data is consistent with the observations, the values obtained from the mapping are uniformly distributed in the interval $[0, 1]$. To assess the uniformity of the mapped values or equivalently assessing the goodness of the theoretical model an empirical CDF of the mapped values is compared against the theoretical CDF of a uniform distribution, which is a 45-degree reference line. The closer the CDF to this reference line, the greater the evidence that the statistical model captures the underlying phenomenon. The Q–Q plot in Figure 3 shows that the MMPP model for the interactive traffic with parameters (3) provides a good fit for the data and significantly outperforms a simpler Poisson model, or a Pareto distribution that has been previously proposed to fit interactive traffic [11].

3. CODE CONSTRUCTION FOR TRAFFIC SHAPING

As mentioned earlier, we now develop codes that can be used for embedding information in timings during interactive SSH sessions. We consider forcing the *inter-arrival times* to satisfy certain algebraic conditions. We know that if we would like to construct a random variable Z with cumulative distribution function (CDF) $F_Z(z)$, then we can first construct a uniform random variable U on $[0, 1]$ and then construct X as

$$Z = F_Z^{-1}(U). \tag{4}$$

So for the for the discrete-time (DT) and continuous-time (CT) queuing channel scenarios, Markov-modulated memoryless point processes we have:

	CT (exponential)	DT (geometric)
$F_Z(z)$	$1 - e^{-\lambda z}$	$1 - (1-\lambda)^z$
$Z(U)$	$\frac{-\ln(1-U)}{\lambda}$	$\left\lceil \frac{\ln(1-U)}{\ln(1-\lambda)} \right\rceil$

(5)

So by first using the inverse CDF, we can collapse the encoding problem into constructing n i.i.d. uniform $[0, 1]$ random variables. It is well known [9] that the ensemble of random linear codes over $\mathbb{F}_Q$ produces codewords words whose elements are i.i.d. and uniformly distributed over $\mathbb{F}_Q$. By shaping according to the method in the previous section, Gallager showed how using random linear coset codes over finite fields with maximum-likelihood decoding suffices to achieve capacity [9, p. 208] on arbitrary discrete memoryless channels.

We propose using a technique based upon algebraic codes and dithering. Consider some field size $Q = 2^t$. Then we force our X_i's to lie in the finite field $\mathbb{F}_Q$. We consider a matrix H with $m < n$ rows and n columns defined over $\mathbb{F}_Q$ and define the linear coset code

$$\mathcal{C} = \{x : \ H\underline{x} = \underline{w}\}.$$

From here, interpret each $x_i \in \mathbb{F}_Q$ as a member of $\mathbb{R}$ and define the ith inter-arrival time, Z_i, as

$$\begin{aligned} Z_i &= T_i(x_i) && \text{(6a)} \\ &= F_Z^{-1}\left(\left[\frac{x_i}{Q} + U_i\right]_{\bmod 1}\right) && \text{(6b)} \end{aligned}$$

where $(U_1, \ldots, U_n)$, are i.i.d. uniform$[0, 1]$ dithers that have been used extensively in quantization and watermarking [5], and communication on linear Gaussian channels [8]. We note from the Crypto Lemma [8], $\left[\frac{x_i}{Q} + U_i\right]_{\bmod 1}$ will also be uniformly distributed on $[0, 1]$. Secondly, the ensemble of random linear codes, the X_i's will be uniformly distributed over $\mathbb{F}_Q$ and thus the Z_i's will be i.i.d. and distributed according to their target distribution.

Before the interactive traffic begins, the encoder and decoder construct the two shaping function vectors pertaining to $i.i.d.$ inter-arrival times at rates λ_0 and λ_1, respectively:

$$\begin{aligned} \underline{T}^{(0)} &= [T_1^{(0)}(\cdot), T_2^{(0)}(\cdot), \ldots, T_n^{(0)}(\cdot)] \\ \underline{T}^{(1)} &= [T_1^{(1)}(\cdot), T_2^{(1)}(\cdot), \ldots, T_n^{(1)}(\cdot)] \end{aligned}$$

We assume that the encoder and decoder in real time know the states $\phi_1, \phi_2, \ldots, \phi_n$, where each $\phi_i \in \{0, 1\}$ and $P(\phi^n)$ follows from the Markov model defined earlier. So we construct the inter-arrival times as follows:

$$T_i(x_i) = T_i^{(\phi_i)}(x_i).$$

So in short, we a priori define the $2n$ functions

$$\begin{aligned} &\{T_i^{(0)} : \mathbb{F}_Q \to \mathbb{R}, 1 \le i \le n\}, \\ &\{T_i^{(1)} : \mathbb{F}_Q \to \mathbb{R}, 1 \le i \le n\}, \end{aligned}$$

and they are completely characterized by a table of $2nQ$ real numbers known at the encoder and decoder. In practice, we can imagine that the encoder and decoder only need to know the state of a random number generator to generate the $2n$ U_i's, which subsequently defines the $2n$ functions.

3.1. Perfectly Secure Structure of This Architecture

Under the assumption that the SSH session evolves according to a two-state Markov-modulated Poisson process, we iterate here the key properties of our approach that makes it perfectly secure:

- We employ random linear coset codes that enable the inter-arrival times Z_i to be independent across time [9]
- By the Crypto Lemma [8] and our inverse CDF shaping technique (6), each inter-arrival time Z_i is of the desired memoryless distribution with parameter λ_{ϕ_i}.

Thus, our approach is perfectly secure with respect to any Markov-modulated Poisson process testing approach. Moreover, our dither shaping technique (6) allows us to precisely mimick the statistical structure of any point process traffic model.

4. PERFORMANCE ACROSS QUEUING TIMING CHANNELS

Here we demonstrate our code's performance across a discrete-time first-in, first-out queuing system with geometric service times of parameter $\mu = 0.9$. The Markov model parameters reflect the continuous-time parameters given in (3) and are given by $\lambda_0 = 0.45$, $\lambda_1 = 0.1$, and $p00 = 0.96$; The average long-term rate of packets/slot for the arrival process is given by:

$$\lambda = \frac{1}{\frac{\pi_0}{\lambda_0} + \frac{\pi_1}{\lambda_1}},$$

where π_0 and π_1 are given by (2). For a discrete queue with geometric service times, the capacity over all input processes of rate λ is given by [2]:

$$C(\lambda) = H_2(\lambda) - \frac{\lambda}{\mu} H_2(\mu) \text{ (bits/slot)}$$
$$\tilde{C}(\lambda) = \frac{H_2(\lambda)}{\lambda} - \frac{H_2(\mu)}{\mu} \text{ (bits/packet)},$$

where $H_2(\cdot)$ is the binary entropy function, and the capacity-achieving input is a Bernoulli process with i.i.d. geometric inter-arrival times of rate λ. So although our traffic scheme is strictly suboptimal in terms of capacity, we gain in terms of traffic shaping and covertness.

We used the aforementioned encoding process and the decoder architecture given in [6] to test its performance using a $Q = 4$, $n = 1000$ regular LDPC coset code to encode messages and simulate them through a FCFS memoryless queue. The DT performance for a geometric server with service rate $\mu = 0.9$ is given in Figure 4. This demonstrates the effectiveness of this approach, with low symbol error rates near the capacity, while maintaining provably good covertness.

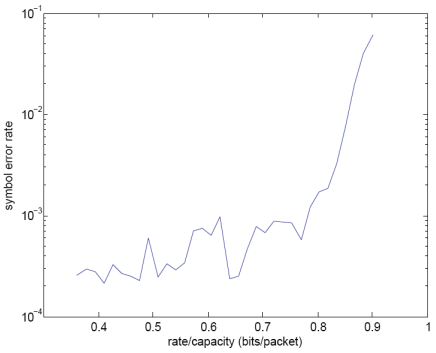

Fig. 4. Symbol Error Rate vs ratio to capacity for the DT queuing channel.

5. REFERENCES

[1] V. Anantharam and S. Verdu. Bits through queues. *IEEE Transactions on Information Theory*, 42(1):4–18, 1996.

[2] A. S. Bedekar and M. Azizoglu. The information-theoretic capacity of discrete-time queues. *IEEE Transactions on Information Theory*, 44(2):446–461, 1998.

[3] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346, 2002.

[4] C. Cachin. An information-theoretic model for steganography. *Information and Computation*, 192(1):41–56, 2004.

[5] B. Chen and G. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Information Theory*, 47(4):1423–1443, 2001.

[6] T. Coleman and N. Kiyavash. Practical codes for queueing channels: An algebraic, state-space, message-passing approach. *Information Theory Workshop (ITW)*, pages 318–322, 2008.

[7] W. Fischer and K. Meier-Hellstern. The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, 18(2):149–171, 1993.

[8] G. D. Forney. On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener. In *Allerton Conference*, 2003.

[9] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, New York, 1968.

[10] M. H. Kang, I. S. Moskowitz, and D. C. Lee. A network pump. *IEEE Transactions on Software Engineering*, 22(5):329–338, May 1996.

[11] V. Paxson and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, June 1995.

[12] P. Peng, P. Ning, and D. S. Reeves. On the secrecy of timing-based active watermarking trace-back techniques. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy*, pages 334–349, 2006.

[13] L. Rabiner. A tutorial on hidden Markov models and selected applications inspeech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[14] A. B. Wagner and V. Anantharam. Information theory of covert timing channels. In *Proceedings of the 2005 NATO/ASI Workshop on Network Security and Intrusion Detection*, 2005.

[15] X. Wang, S. Chen, and S. Jajodia. Network Flow Watermarking Attack on Low-Latency Anonymous Communication Systems. In *Proc. 2007 IEEE Symposium on Security and Privacy, May*, 2007.