ON SECURITY-AWARE TRANSMISSION SCHEDULING

Ting He , Ameya Agaskar, and Lang Tong

ABSTRACT

The problem of interest is to characterize to what extent nodes independently following certain transmission schedules can be hijacked to relay flows of information packets. Information flows can be embedded in given transmission schedules by properly adding delays and inserting dummy packets. Such hidden flows are usually indicators of network intrusion, and it is of interest to know their rates. The maximum rate of information flow that can be transmitted without causing the transmission activities to deviate from given transmission schedules is used to measure the covert capacity under these schedules. Based on the assumption that information flows have bounded delays, a theoretical framework is constructed to quantitively analyze the covert capacity under transmission schedules modeled by renewal processes. Explicit solution is obtained for Poisson processes. The results suggest a close correlation between the covert capacity and the traffic burstiness.

Keywords: Covert capacity, Information flow, Transmission scheduling.

1. INTRODUCTION

Hidden flows of information-carrying packets are usually indicators of network intrusion. For example, in stepping-stone attacks ([1]), an attacker tries to protect his identity by tunneling through multiple hosts before attacking the target. To secure the network, it is crucial to detect hidden information flows.

The task of detecting information flows faces multiple challenges. For example, as illustrated in Fig. 1, consider detecting the flow through nodes R_1 and R_2 in a wireless ad hoc network. Suppose that traffic is encrypted so that no correlation can be deduced from packet content or length. Moreover, the nodes can hide the correlation in timing by delaying the transmission of information packets and multiplexing them with packets of intersecting flows as well as dummy packets. Such multiplexed packets, which are not part of the flow of interest, cause noise in detection and are thus referred to as *chaff noise*. With proper perturbation and sufficient chaff noise, an information flow can be disguised as traffic of arbitrary pattern. In particular, the flow can appear identical to independent traffic following certain transmission schedules. Therefore, every transmission schedule has certain capacity of being utilized to transmit information flows covertly.

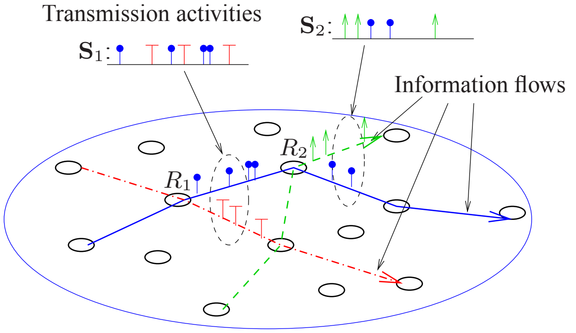

Fig. 1. In a wireless ad hoc network, an information flow through R_1, R_2 (solid line) can be hidden among intersecting flows (dashed and dash-dot lines) and dummy packets. ○: wireless node; $\mathbf{S}_i$: process of transmission timestamps.

1.1. Related Work

This work is motivated by the detection of stepping-stone attacks first studied in [1], where the goal is to detect flows of attacking traffic by correlating traffic characteristics. To deal with encrypted traffic, timing characteristics are used in detection. In particular, Donoho *et al.* in [2] proposed a flow model based on bounded delay constraint, and a parallel model based on bounded memory constraint was proposed in [3]. Chaff noise was briefly addressed in [2], with the claim that it is impossible to hide information flows in independent traffic if the perturbation is bounded and the chaff noise is independent of the flow. This argument, however, breaks down when the chaff noise and the flow can be correlated.

For arbitrary chaff noise, [4] presented the first timing-based detector that achieves consistent detection even if chaff noise grows proportionally with the traffic size; other detectors only handle a limited number of chaff packets (see references in [4]). Furthermore, in [5], it is shown that there exists a threshold on the noise level beyond which the flows can be completely undetectable and below which detectable by a single detector. Analytical characterization of the thresholds are derived, showing that perfect detection can be achieved as the flow path grows. The work in [4] has been extended to distributed detection, where achievability results are obtained for detecting flows with quantized measurements; see [6, 7].

1.2. Summary of Results and Organization

We want to analyze how much information flow can be transmitted covertly through nodes following certain transmission schedules. Our goal is to provide a quantifiable measure for covert information flows and study its relationship with statistical properties of the transmission schedules.

T. He (the@us.ibm.com) is with IBM T. J. Watson Research Center, Hawthorne, NY 10532.

A. Agaskar and L. Tong ({apa22,lt35}@cornell.edu) are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853.

Most of the work was completed when Ting He was at Cornell. This work is supported in part by the National Science Foundation under award CCF-0635070, the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program DAAD19-01-2-0011.

With arbitrary chaff noise, it is always possible to embed information flows into given transmission schedules by properly adding delays and mixing chaff noise. Once embedded, the flows become undetectable and thus covert. The maximum rate of such covert flows normalized by the overall traffic rate gives us a natural measure of the capacity of covert information flows under the corresponding schedules. Based on this measure, we present a framework for analyzing the covert capacity under transmission schedules modeled by renewal processes. Specifically, we use an algorithm proposed in [8] which can embed the most information packets into any transmission schedules subject to a strict delay constraint. By modeling the behavior of this algorithm as a discrete-time, continuous-state Markov process, we convert the problem to one of computing the limiting probabilities of the Markov process, which can then be solved either analytically or numerically. Through simulations, we calculate the covert capacities under various interarrival distributions. Our results suggest that the covert capacity tends to increase as the interarrival tailweight decreases and the traffic becomes less bursty. In particular, Poisson traffic has a much higher covert capacity than real traces.

The rest of the paper is organized as follows. Section 2 defines the problem. Section 3 presents theoretical results, followed by simulations in Section 4. Then Section 5 concludes the paper.

2. PROBLEM STATEMENT

Let the transmission schedule[1] of node R_i ($i = 1,\ 2$) be denoted by a point process $\mathbf{S}_i$, *i.e.,*

$$\mathbf{S}_i = (S_i(1),\ S_i(2),\ S_i(3), \ldots),\ \ i = 1,\ 2,$$

where $S_i(k)$ ($k \geq 1$) is the transmission timestamp of the kth packet at R_i (assume no simultaneous transmissions). We say that $(\mathbf{S}_1,\ \mathbf{S}_2)$ *contains an information flow* $(\mathbf{F}_1,\ \mathbf{F}_2)$ if it can be decomposed[2] into processes $(\mathbf{F}_i)_{i=1}^2$ and $(\mathbf{W}_i)_{i=1}^2$:

$$\mathbf{S}_i = \mathbf{F}_i \oplus \mathbf{W}_i,\ \ i = 1,\ 2, \tag{1}$$

where $(\mathbf{W}_i)_{i=1}^2$ is called *chaff noise* and $(\mathbf{F}_i)_{i=1}^2$ an *information flow* by the following definition.

Definition 2.1 *A pair of processes* $(\mathbf{F}_1,\ \mathbf{F}_2)$ *is an* information flow *if for every realization*[3] $(\mathbf{f}_1,\ \mathbf{f}_2)$*, there exists a bijection* $g : \mathcal{F}_1 \rightarrow \mathcal{F}_2$ *such that* $g(s) - s \in [0,\ \Delta]$ *for all* $s \in \mathcal{F}_1$.

The bijection g is a mapping between the timestamps of the same packets at R_1 and R_2. The condition that g is a bijection ensures *packet-conservation*. The condition $g(s) - s \in [0,\ \Delta]$ implies *causality* as well as a *maximum delay* Δ (this condition was first used by Donoho *et al.* in [2]). Assume that Δ is known.

Given a transmission schedule, we measure the level to which this schedule can contain information flow as follows.

Definition 2.2 *Given transmission schedule* $(\mathbf{S}_1,\ \mathbf{S}_2)$*, the* relative capacity of covert information flows *(referred to as* covert capacity*) under this schedule is defined as*[4]

$$\begin{aligned} C(\mathbf{S}_1,\ \mathbf{S}_2) \ \triangleq\ & \sup\{r \in [0,\ 1] : \exists (\mathbf{F}_i)_{i=1}^2 \textit{ such that:} \\ & \textit{1) } (\mathbf{S}_i)_{i=1}^2 \textit{ contains an information flow } (\mathbf{F}_i)_{i=1}^2; \\ & \textit{2) } \liminf_{t\rightarrow\infty} \frac{\sum_{i=1}^2 |\mathcal{F}_i \cap [0,\ t]|}{\sum_{i=1}^2 |\mathcal{S}_i \cap [0,\ t]|} \geq r \textit{ a.s.}\}. \end{aligned}$$

Intuitively, the covert capacity is the maximum asymptotic fraction of information packets that can be transmitted (under the causality and the delay constraints) through two nodes using the given schedule[5]. In the sequel, we will focus on schedules that start simultaneously and have the same rate and distribution. All distributions are continuous unless explicitly stated otherwise.

3. COMPUTING COVERT CAPACITY

It is difficult to compute the covert capacity directly by Definition 2.2 because it involves an optimization over all the possible ways of embedding information flows. In this section, we will present a systematic approach to computing the covert capacity. The idea is to find the optimal algorithm that can embed the most information packets and then analyze this algorithm.

3.1. Optimal Embedding Algorithm

The optimal embedding algorithm is called "Bounded Greedy Match" (BGM), proposed by Blum *et al.* in [8]. Given realizations of transmission schedules $(\mathbf{s}_1,\ \mathbf{s}_2)$, BGM does the following:

1. sequentially match every packet at s in $\mathbf{s}_1$ with the first unmatched packet in $[s,\ s + \Delta]$ in $\mathbf{s}_2$;
2. the matched packets form (a realization of) an information flow and the unmatched ones chaff noise.

It was shown in [8] that BGM embeds the most information packets for any $(\mathbf{s}_1,\ \mathbf{s}_2)$.

3.2. Analytical Results

The optimality of BGM allows us to use it to compute the covert capacity. Let us first take a closer look at BGM. In the jth step of BGM, let y_j be the delay in the next pair of unmatched packets $(s_1(m),\ s_2(n))$, *i.e.,* $y_j = s_2(n) - s_1(m)$. Then as illustrated in Fig. 2, if $y_j < 0$ (*e.g.,* y_1), then $s_2(n)$ will be marked as chaff noise, and the next pair will be $(s_1(m),\ s_2(n + 1))$; if $y_j > \Delta$ (*e.g.,* y_3), then $s_1(m)$ will be chaff noise and the next pair $(s_1(m + 1),\ s_2(n))$; otherwise (*e.g.,* $y_2,\ y_4$), the pair $(s_1(m),\ s_2(n))$ is successfully matched, and the next pair will be $(s_1(m+1),\ s_2(n+1))$. Based on the above observation, we derive the following property of BGM.

Proposition 3.1 *If* $\mathbf{S}_i$ *(*$i = 1,\ 2$*) are renewal processes and* Y_j *(*$j \geq 1$*) defined as above, then* $(Y_j)_{j=0}^\infty$ *(*$Y_0 \triangleq 0$*) is a Markov process.*

[1] In general, a transmission schedule can be stochastic, with each realization denoting a specific sequence of transmission timestamps.

[2] Given nondecreasing sequences $(a_1,\ a_2, \ldots)$ and $(b_1,\ b_2, \ldots)$, $\bigoplus$ is defined as $(a_k)_{k=1}^\infty \oplus (b_k)_{k=1}^\infty = (c_k)_{k=1}^\infty$, where $c_1 \leq c_2 \leq \ldots$ and $\{a_k\}_{k=1}^\infty \cup \{b_k\}_{k=1}^\infty = \{c_k\}_{k=1}^\infty$.

[3] We use lower case letters for realizations and script letters for the sets of timestamps in the realizations.

[4] Here "a.s." means almost surely.

[5] Note that the covert capacity is a deterministic quantity; it is the largest almost-sure lower bound of the asymptotic fraction of information packets.

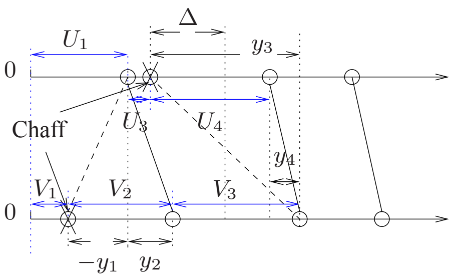

Fig. 2. BGM: sequentially match packets subject to causality and bounded delay. Chaff packets are in 1-1 correspondence with delays less than 0 or greater than Δ.

Proof: From the above discussion, one can derive that Y_j has the following update:

$$Y_j = \begin{cases} Y_{j-1} + V_j & \text{if } Y_{j-1} < 0 \\ Y_{j-1} - U_j & \text{if } Y_{j-1} > \Delta \\ Y_{j-1} + V_j - U_j & \text{o.w.,} \end{cases} \tag{2}$$

where U_j, V_j denote the interarrivals before the next unmatched packets in the jth step, as shown in Fig. 2. Since the interarrivals are independent[6], $(Y_j)_{j=0}^{\infty}$ is Markovian. ■

The significance of this Markovian property is that it gives us a convenient way of computing the covert capacity. As illustrated in Fig. 2, each Y_j within the interval $[0,\ \Delta]$ corresponds to a pair of matched packets, whereas each Y_j outside this interval corresponds to a chaff packet. Thus, the problem of computing the covert capacity is reduced to one of calculating the limiting probabilities of $(Y_j)_{j=0}^{\infty}$, as stated in the following theorem.

Theorem 3.2 *If* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are i.i.d. renewal processes with interarrival* probability density function (pdf) $f(x)$ *(*$x \geq 0$*), suppose* $\exists$ *a nondecreasing, right-continuous function* $H(x)$ *(*$x \in \mathbb{R}$*) satisfying* $\lim_{x\to-\infty} H(x) = 0$, $\lim_{x\to\infty} H(x) = 1$, *and*

$$H(x) = L(x) + \int_{-\infty}^{0} H(y)f(x-y)dy + \int_{0}^{\Delta} H(y)g(x-y)dy + \int_{\Delta}^{\infty} H(y)f(y-x)dy, \tag{3}$$

where $g(x)$ *is the convolution of* $f(x)$ *and* $f(-x)$*, defined as* $g(x) \triangleq \int_0^{\infty} f(y)f(y-x)dy$*, and*

$$L(x) \triangleq [F(x) - G(x)]\, H(0) + [G(x-\Delta) + F(\Delta - x) - 1]\, H(\Delta),$$

where $F(x), G(x)$ *are the* cumulative distribution functions (cdf's) *of* $f(x)$*,* $g(x)$*, respectively. Then the fraction of packets matched by BGM converges a.s., and the limit (i.e., the covert capacity) is given by*

$$C(\mathbf{S}_1,\ \mathbf{S}_2) = \frac{2-2q}{2-q}, \tag{4}$$

where $q = 1 + H(0) - H(\Delta)$*.*

Proof: See Appendix. ■

Theorem 3.2 provides an analytical way of computing the covert capacity for renewal processes. The theorem says that for continuous interarrivals, the process $(Y_j)_{j=0}^{\infty}$ in Proposition 3.1 is ergodic if invariant probability measure exists. Here $H(x)$ is the limiting cdf of $(Y_j)_{j=0}^{\infty}$, and q is the asymptotic fraction of time that Y_j's fall outside $[0,\ \Delta]$. Equation (3) does not always have a closed-form solution; for Poisson processes, however, we have a closed-form solution as follows.

Corollary 3.3 *If* $\mathbf{S}_i$ *(*$i = 1,\ 2$*) are independent Poisson processes of rate* λ*, then*

$$C(\mathbf{S}_1,\ \mathbf{S}_2) = \lambda\Delta/(1 + \lambda\Delta). \tag{5}$$

Proof: In Theorem 3.1 in [4], it was shown that the fraction of chaff noise inserted by BGM is $1/(1+\lambda\Delta)$. Since the covert capacity is the complement of this fraction, the desired result holds. ■

We note that the covert capacity in (5) has an interpretation by queueing theory: by Little's law, the average number of packets buffered at the relay node is $\lambda\Delta$.

4. SIMULATIONS

In this section, we study the covert capacities under several types of interarrival distributions, aiming at revealing the relationship between covert capacity and statistical properties of the distribution. Let the mean interarrival time be $1/\lambda$. Using Poisson traffic (*i.e.,* exponential interarrival) as a benchmark, we simulate uniform distribution on $[0,\ 2/\lambda]$, Pareto distribution with pdf[7]

$$f(x) = \beta a^{\beta} x^{-\beta-1}, \quad a,\ \beta \geq 0, \quad x \geq a,$$

and the shifted Pareto distribution (the distribution of $X - a$ for Pareto random variable X). These distributions represent traffic both less bursty and burstier than Poisson traffic. The covert capacities are computed by simulating BGM on pairs of renewal processes independently generated according to the above distributions.

We first plot the covert capacities as functions of the traffic rate λ, as shown in Fig. 3. We see that all the covert capacities increase with λ, which is because as λ increases, the delay bound becomes relatively larger compared with the interarrival times, verifying the intuition that it is easier to hide information flows in heavier traffic. Moreover, we see a trend that steady traffic has higher covert capacity than bursty traffic, *e.g.,* uniform, exponential, and shifted Pareto distributions have increasing tailweights, resulting in more burstiness and smaller covert capacities. This observation, however, does not hold for Pareto distribution, for which the covert capacity can be either higher (*e.g.,* $\beta = 2$) or lower (*e.g.,* $\beta = 1.1$) than exponential.

Next, to further understand the trend of change, we plot the covert capacities with respect to the shape parameter β; see Fig. 4. The figure shows that the covert capacity under shifted Pareto distribution converges to that under exponential distribution as $\beta \to \infty$, which is as expected because the distributions converge. For

[6] Note that each interarrival is only added (or subtracted) once.

[7] Here, a is called the position parameter and β the shape parameter of Pareto distribution.

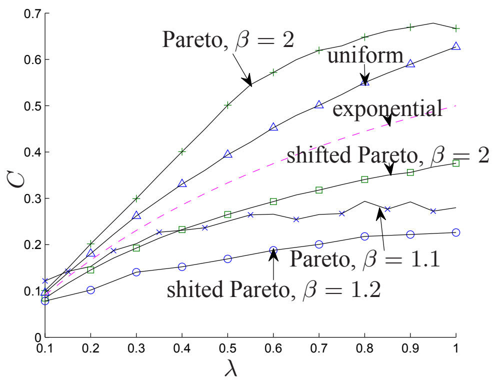

Fig. 3. Covert capacities vs. traffic rate λ ($\Delta = 1$, 10^5 packets per process).

Pareto distribution, there exists a threshold $\beta^* \approx 1.4$ on β below which the covert capacity is lower than exponential and above which it is higher. To explain this phenomenon, note that as $\beta \to \infty$, Pareto distribution becomes $\delta(x - 1/\lambda)$ with covert capacity 1, whereas as $\beta \to 1$, it becomes the shifted Pareto distribution. Since Pareto distribution with $\beta \approx 0.9$ fits the interarrivals of TELNET traces ([9]), we expect the covert capacity under practical transmission schedules to be much lower than that under Poisson schedule. A lesson learned from these simulations is that although within the same family of distributions, the covert capacity decreases with the tailweight, the statement may not hold for different types of distributions because the "shape" of the distribution matters.

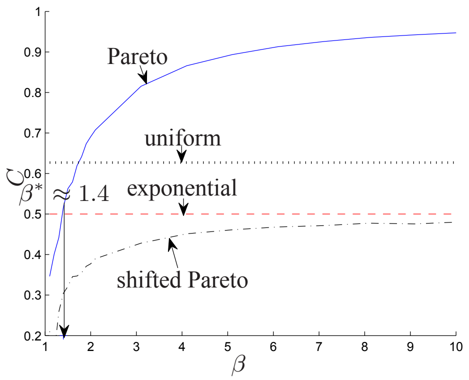

Fig. 4. Covert capacities vs. shape parameter β ($\lambda = 1$, $\Delta = 1$, 10^5 packets per process).

5. CONCLUSION

In this paper, we study the capacity of transmitting information flows covertly under transmission schedules with predetermined distributions. We define a mathematical framework for analyzing the covert capacity, based on which both theoretical and numerical results are obtained for transmission schedules with *i.i.d.* interpacket delays.

6. APPENDIX

6.1. Proof of Theorem 3.2

It suffices to show that the long-term frequency for Y_j to fall outside $[0,\ \Delta]$ converges a.s. to q. Then since each Y_j corresponds to two packets if within $[0,\ \Delta]$ but only one packet otherwise, the asymptotic fraction of matched packets, *i.e.,* the covert capacity, is given by $2(1-q)/(2-q)$.

We now prove the convergence. It suffices to consider states reachable from 0. By Theorem 17.1.7 in [10], the long-term frequency of visiting $[0,\ \Delta]^c$ converges a.s. to its probability in the stationary distribution if $(Y_j)_{j=0}^{\infty}$ is positive Harris. Positivity follows from the fact that $H(x)$ is the cdf of an invariant probability measure of $(Y_j)_{j=0}^{\infty}$. Harris recurrence follows from that: (i) $(Y_j)_{j=0}^{\infty}$ is ψ-irreducible for a maximal irreducibility measure ψ, (ii) $[0,\ \Delta]$ is a petite set, and (iii) $[0,\ \Delta]$ is a.s. accessible from all the states (details omitted due to space constraint).

■

7. REFERENCES

[1] S. Staniford-Chen and L. Heberlein, "Holding intruders accountable on the internet," in *Proc. the 1995 IEEE Symposium on Security and Privacy*, (Oakland, CA), pp. 39–49, May 1995.

[2] D. Donoho, A. Flesia, U. Shankar, V. Paxson, J. Coit, and S. Staniford, "Multiscale stepping-stone detection: Detecting pairs of jittered interactive streams by exploiting maximum tolerable delay," in *5th International Symposium on Recent Advances in Intrusion Detection, Lecture Notes in Computer Science 2516*, 2002.

[3] T. He and L. Tong, "Detecting Encrypted Stepping-Stone Connections," *IEEE Transactions on Signal Processing*, vol. 55, pp. 1612–1623, May 2007.

[4] T. He and L. Tong, "Detecting Information Flows: Improving Chaff Tolerance by Joint Detection," in *Proc. 2007 Conference on Information Sciences and Systems*, (Baltimore, MD), March 2007.

[5] T. He and L. Tong, "Detection of Information Flows." submitted to IEEE Trans. on Information Theory, 2007.

[6] T. He and L. Tong, "Distributed Detection of Information Flows in Chaff," in *Proc. 2007 IEEE International Symposium on Information Theory*, (Nice, France), June 2007.

[7] T. He and L. Tong, "Distributed Detection of Information Flows with Side-Information," in *Proc. 2007 Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA), November 2007.

[8] A. Blum, D. Song, and S. Venkataraman, "Detection of Interactive Stepping Stones: Algorithms and Confidence Bounds," in *Conference of Recent Advance in Intrusion Detection (RAID)*, (Sophia Antipolis, French Riviera, France), September 2004.

[9] V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, June 1995.

[10] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. London, UK: Springer-Verlag, 1993.