ON THE MAXIMUM THROUGHPUT OF CLANDESTINE SENSOR NETWORKING

Ting He, Lang Tong, and Ananthram Swami

ABSTRACT

We consider the problem of clandestine communications, i.e., communications that are meant to be invisible to third-party eavesdroppers, in the context of wireless sensor networks. Although encryption and anonymous routing protocols can hide the content and the routing information, the transmission activities of sensors on the same route can still reveal the information flow. In this work, a perfectly clandestine scheduling method is developed to hide the desired information flow in a sequence of independent transmission activities resembling those without any flow, while satisfying the resource constraint at the relay nodes in terms of limited buffer size. The proposed method is proved to achieve the maximum throughput, which is characterized analytically for transmission schedules following alternating renewal processes with a closed-form solution for Poisson processes. The analytical results are verified through numerical simulations on synthetic traffic as well as traces.

Keywords: Clandestine communications, Scheduling algorithms, Performance analysis.

1. INTRODUCTION

We consider the problem of clandestine communications in wireless sensor networks. *Clandestine communications* refer to the scenario where the *act* of communication needs to be kept secret from eavesdroppers distributed in the field. In contrast to watermarking techniques which hide secret information in open communications, clandestine communications require the overall act of communication to be hidden. Although various anonymous routing protocols have been developed to hide information in traffic content (including both packet headers and data portions) through encryption and pseudonyms [1], traffic activities can still reveal the presence of communications [2]. In this paper, we propose to enable truly clandestine communications by deploying *clandestine relays*[1], which hide the correlation between incoming and outgoing traffic not only in the content domain (by implementing anonymous routing protocols) but also in the activity domain (by using proper scheduling mechanisms specified later). Clandestine communication is crucial to the success of clandestine military operations where the act is to be kept as invisible as possible. It is also a popular technique in anonymous networking where the sources and/or destinations of flows are hidden from traffic monitors [3]. Moreover, it has significant implications in network security problems such as wormhole attacks [4], where an intruder channels information flows through a tunnel unknown to the source and the destination. Understanding to what degree a clandestine relay can carry information flows without being detected by networked traffic monitors is critical to managing clandestine communications, and to study this "degree" in a rigorous and quantitative manner is at the core of this paper.

We assume omnipresent and fully networked traffic monitors. We will restrict the monitors' observation to timing information only. Other flow information, e.g., addresses, flow types, and packet content, will certainly make the monitors more powerful, but will at the same time limit the scope of the analysis since such information is hidden in many anonymous routing protocols. Obviously, it would not be possible to track a specific packet just based on timing. Resource constraints at the relay node, however, may introduce traceable patterns at the flow level. In this paper, we focus on the buffer size constraint, which implies that the amount of information buffered at a relay node has to be bounded by the maximum buffer size at any time, introducing certain statistical correlation across traces on the same flow path and thus revealing the flow. To hide the flow, nodes have to *embed* the flow transmission into *cover traffic*[2] that follows their normal transmission schedules, a particular type of which is a set of statistically independent schedules. Since not every transmission in the schedule sat-

Research was sponsored in part by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-06-3-0001 and the Army Research Office MURI program under award W911NF-08-1-0238. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence, or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

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

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. Email: lt35@cornell.edu

A. Swami is with the Army Research Laboratory, Adelphia, MD 20783. Email: aswami@arl.army.mil

[1] In contrast, a *covert relay* is a relay node that hides its identity although the presence of flow may be detectable.

[2] Cover traffic is the overall traffic observable to monitors that includes the transmission of both the flow and the chaff traffic. In contrast, relay nodes can see the traffic content and thus distinguish the flow from the chaff.

978-1-4244-5239-2/09/$26.00 ©2009 IEEE

isfies the flow constraints, such embedding may lead to rate loss, and it is our goal to characterize such loss.

1.1. Summary of Contributions and Limitations

The main focus of this paper is on the fundamental limit of clandestine communications. Our contributions are threefold:

Optimal flow-scheduling algorithm: We develop a linear-complexity algorithm that matches the maximum amount of transmissions for a given pair of transmission schedules under a buffer size constraint. The algorithm, using the *First Come, First Serve (FCFS)* principle, is sequential and thus suitable for online scheduling of information flows under predetermined transmission schedules.

Performance analysis: We then analyze the proposed algorithm to characterize its efficiency, measured by the fraction of embedded flow in the cover traffic (i.e., the maximum normalized clandestine throughput). Assuming ON-OFF renewal schedules, we derive an analytical solution based on the limiting distribution of a Markov-modulated random walk constructed via the scheduling algorithm. For small packets, we give a closed-form solution for exponential inter-packet delays, which then provides bounds for other distributions. The analysis shows that the clandestine throughput is negatively related to the level of traffic burstiness.

Simulation studies: We complement the analysis with simulations on both synthetic traffic and real traces. Besides confirming the analytical results, the simulations also show that renewal traffic with power-law inter-packet delays can closely approximate the clandestine throughputs of traces.

As an analytical study, our results are limited by the models and assumptions such as the renewal traffic assumption, although our case study of traces has shown that with proper distributions, renewal processes can model network traffic reasonably well (Section 5.3). We will leave more extensive trace studies to future work.

1.2. Related Work

The problem of characterizing the maximum throughput of clandestine networking has not been formally studied in the past, but problems sharing common concepts have been investigated. The problem of avoiding Internet traffic analysis has been considered in [5], which uses a special relay called *Mix* to mix and re-encrypt packets from multiple users to hide the sources of individual packets; long packet streams can still be correlated. To prevent such *flow correlation*, the method of *cover traffic* is used to pad the actual traffic with dummy packets such that the overall transmission activities conform to predetermined schedules [3], although this method suffers from synchronization and efficiency problems. The most related work is [6], which analyzes the throughput of clandestine communications under strict delay constraints. This work differs from [6] in that: we assume unbounded delay but bounded buffer size, and we use ON-OFF schedules instead of point process schedules. We feel the new models better suit applications in sensor networks because of their stringent resource constraints and bursty communication needs. The results presented in this paper also extend the earlier work in [2, 7] that assumed Poisson transmission schedules.

The rest of the paper is organized as follows. Section 2 defines the problem, and Section 3 presents a flow-scheduling algorithm, which is analyzed in Section 4. Section 5 presents simulation results. Section 6 concludes the paper.

2. PROBLEM STATEMENT

For clarity of presentation, we use uppercase letters to denote random variables, lowercase letters for realizations, boldface letters for vectors, and plain letters for scalars.

2.1. Flow Models

Denote the incoming and outgoing transmission schedules of a relay node by ON-OFF processes $\mathbf{S}_i$ $(i = 1,\ 2)$

$$\mathbf{S}_i \triangleq ([S_i^s(k),\ S_i^t(k)])_{k=1}^{\infty}, \tag{1}$$

where $S_i^s(k)$ is the starting time and $S_i^t(k)$ the terminating time of the kth packet, with a *packet length*[3] $L_i(k) \triangleq S_i^t(k) - S_i^s(k)$. Schedules $(\mathbf{S}_1,\ \mathbf{S}_2)$ specify the generation of cover traffic, which is what the traffic monitor can observe[4].

Under predetermined schedules, the act of relay can be considered a process of embedding an information flow into these schedules. Specifically, as illustrated in Fig. 1, we model such embedding by a decomposition

$$L_i(k) = L_i^{(I)}(k) + L_i^{(C)}(k), \tag{2}$$

where $L_i^{(I)}(k)$ denotes the length of the *information portion* of a packet, defined by the portion that is generated by

[3] Since the problem is defined in the time domain, we measure packet length by the time taken to transmit that packet.

[4] We note that cover traffic may include transmission failures, and the clandestine throughput calculated hereby is thus an optimistic estimate of the actual throughput achieved. From the source's perspective, it can replace the schedules by distributions of successful transmissions to predict the achievable throughput.

the source and will reach the destination, and $L_i^{(C)}(k)$ the length of *chaff noise*. Chaff noise models portions of transmissions that are not relayed from the source to the destination, including dummy packets, dropped packets, superfluous data padded in packets, and multiplexed packets from other flows. Information bits and chaff bits can be mixed in any order within a packet, and $L_i^{(I)}(k)$, $L_i^{(C)}(k)$ denote their total lengths, respectively (either of them can be zero).

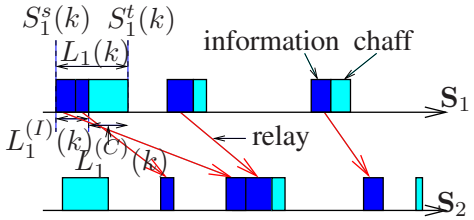

Fig. 1. Decompose each transmission in $\mathbf{S}_i$ ($i = 1, 2$) into an information portion and chaff noise, where the information portions in the two schedules have to be 1-1 matched.

We say that transmission schedules $(\mathbf{S}_1, \mathbf{S}_2)$ contain an *embedded information flow* if they can be decomposed as in (2) such that the following definition holds.

Definition 2.1 *A pair of transmission schedules* $(\mathbf{S}_1, \mathbf{S}_2)$ *with effective packet lengths* $(\mathbf{L}_1^{(I)}, \mathbf{L}_2^{(I)})$ *is a (two-hop)* information flow *if the following conditions hold:*

Flow-conservation: $\sum_{k=1}^{\infty} L_1^{(I)}(k) = \sum_{k=1}^{\infty} L_2^{(I)}(k)$, *i.e., the volume of information-carrying traffic has to be conserved during relay.*

Causality: $\sum_{s_1^t(k)\le t} L_1^{(I)}(k) \ge \sum_{s_2^s(k)\le t} L_2^{(I)}(k)$ *for any* $t > 0$, *i.e., the relay packet can only start transmitting after the original packet is completely received.*

Bounded buffer size: $\sum_{s_1^t(k)\le t} L_1^{(I)}(k) - \sum_{s_2^s(k)\le t} L_2^{(I)}(k) \le b$ *for any* $t > 0$, *i.e., the amount of information bits awaiting relay cannot exceed the maximum buffer size* b *at any time*[5].

The above definition allows packets to be combined, split, delayed, and permuted during relay. The flow-conservation constraint defines a relay operation, the causality constraint ensures temporal order of reception and relay, and the buffer size constraint models the resource limitation at relay nodes. Note that the causality constraint is specially designed to require that information in a packet can only be relayed after the whole packet arrives, which enables packet-level transformation such as decryption and re-encryption. We assume that the constant b is known.

[5]In reality, there will be two other (smaller) buffers to store packets during their reception and transmission, which are not included here.

2.2. Maximum Clandestine Throughput

The constraints in Definition 2.1 imply that not every transmission in given schedules can be used to relay information. We measure the efficiency of relaying information flow under given schedules by maximum clandestine throughput, defined as follows.

Definition 2.2 *Given transmission schedules* $(\mathbf{S}_1, \mathbf{S}_2)$, *the* maximum normalized throughput of a clandestine relay (maximum clandestine throughput) *under these schedules is defined as the maximum asymptotic fraction of embedded information flows, i.e.,*

$$T_b(\mathbf{S}_1, \mathbf{S}_2) \triangleq \sup\{r \in [0, 1] : \exists (\mathbf{L}_i^{(I)})_{i=1}^2 \text{ such that:}$$
$$1)\ (\mathbf{S}_i)_{i=1}^2 \text{ with effective packet lengths } (\mathbf{L}_i^{(I)})_{i=1}^2 \text{ is an information flow};$$
$$2)\ \liminf_{N\to\infty} \frac{\sum_{k=1}^{N} L_1^{(I)}(k) + L_2^{(I)}(k)}{\sum_{k=1}^{N} L_1(k) + L_2(k)} \ge r \text{ a.s.}\}. \quad (3)$$

Under this definition, the maximum clandestine throughput is the long-term fraction of information blocks (in length), maximized over all possible ways of embedding them into the given schedules. Intuitively, certain rate loss will occur if the relay has to embed the flow into given transmission schedules rather than simply forwarding packets as they arrive, and the maximum clandestine throughput is the ratio of flow rates with and without clandestine relay.

3. OPTIMAL FLOW-SCHEDULING ALGORITHM

Given two transmission schedules, there are many ways to embed an information flow in them, some achieving higher throughputs than others. In this section, we aim at developing algorithms that embed flows optimally to achieve the maximum clandestine throughput.

For schedules following general ON-OFF processes, the proposed scheduling algorithm is called "Bounded Buffer Relay" (BBR), presented in Algorithm 1. Algorithm BBR is based on the idea of *First Come, First Serve (FCFS)*: it uses variables $B(n)$ to keep track of the amount of used buffer (lines 5, 7)[6], checks for buffer overflow or underflow after each arrival or departure, and records the superfluous amount as chaff bits in another variable C (lines 9, 12), which is then used to compute the overall fraction of non-chaff bits. The above procedure is illustrated in Fig. 2.

[6]Note that $L(n)$ is the packet length in cover traffic. The actual length of the information portion may be smaller.

Algorithm BBR is an extension of the algorithm "Bounded Memory Relay" (BMR) proposed in [2] for embedding flows into transmission schedules modeled by point processes. Algorithm BMR operates on schedules represented by point processes and thus ignores different packet sizes. It has been shown in [2] that BMR is optimal in that it achieves the maximum throughput under arbitrary realizations of point processes. Similar arguments can also be used to show the optimality of BBR, details omitted due to space limit.

Algorithm 1 Bounded Buffer Relay (BBR)

Require: Realizations of ON-OFF processes $(\mathbf{s}_1,\ \mathbf{s}_2)$.

Ensure: Return the maximum fraction of information bits in $(\mathbf{s}_1,\ \mathbf{s}_2)$ under the constraints in Definition 2.1.

$\mathbf{s} \leftarrow$ merge $\mathbf{s}_1^t,\ \mathbf{s}_2^s$
initial values: $C \leftarrow 0$, $B(0) \leftarrow 0$
for all $s(n)$ in $\mathbf{s}$ **do**
  if $s(n)$ is from $\mathbf{s}_1^t$ **then** {a packet arrives}
    $B(n) \leftarrow B(n-1) + L(n)$ {$L(n)$: packet size}
  else {$s(n)$ is from $\mathbf{s}_2^s$, i.e., a packet departs}
    $B(n) \leftarrow B(n-1) - L(n)$
  if $B(n) > b$ **then** {buffer overflow}
    $C \leftarrow C + B(n) - b$ {count the amount of chaff}
    $B(n) \leftarrow b$
  else if $B(n) < 0$ **then** {buffer underflow}
    $C \leftarrow C - B(n)$
    $B(n) \leftarrow 0$
return $1 - C/\left[\sum_n (L_1(n) + L_2(n))\right]$

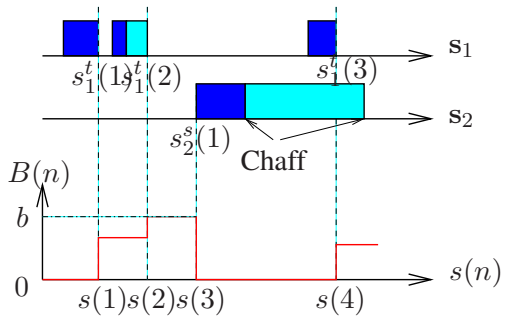

Fig. 2. BBR: Keep track of used buffer size $B(n)$, updating its value after each arrival and before each departure and inserting chaff bits if needed to ensure $B(n) \in [0,\ b]$.

4. THROUGHPUT ANALYSIS

The optimality of BBR allows us to compute the clandestine throughput by applying it to given transmission schedules. Such an algorithmic solution, however, provides little insight into the relationship between the clandestine throughput and external parameters such as statistical properties of the schedules and the maximum buffer size. To this end, we study the maximum clandestine throughput of certain families of schedules and derive analytical characterization accordingly.

4.1. Alternating Renewal Processes

Since the computation of the maximum clandestine throughput is fundamentally an analysis of the asymptotic performance of the optimal scheduling algorithm BBR, it is essential to model the operations of BBR mathematically. Based on Algorithm 1, we see that the key operation is the following update:

$$B(n) = \begin{cases} \min(b,\ B(n-1) + L(n)) & \text{if } S(n) \in \mathbf{S}_1^t, \\ \max(0,\ B(n-1) - L(n)) & \text{o.w.} \end{cases} \tag{4}$$

Intuitively, the size of used buffer $B(n)$ forms a "random walk" on the real axis between 0 and b, increasing or decreasing by a packet length $L(n)$ according to whether the next packet is an arrival or a departure. Since packet lengths are i.i.d. for alternating renewal processes, the absolute value of the step of $B(n)$ is i.i.d. . Endpoints 0 and b are "reflective barriers" of $B(n)$ in the sense that $B(n)$ is constrained at the barrier whenever it tries to escape. More importantly, each escape of $B(n)$ from interval $[0,\ b]$ represents an insertion of chaff bits, and the amount of excess is equal to the amount of chaff bits inserted. Therefore, the process $\{B(n)\}_{n=0}^{\infty}$ directly maps to the maximum clandestine throughput by the formula in (5), where $(x)_+ \stackrel{\Delta}{=} \max(x,\ 0)$, and $L'(n) = L(n)$ if $s(n) \in \mathbf{s}_1^t$ and $L'(n) = -L(n)$ otherwise.

Note that the above intuitive argument has a critical flaw because although the absolute step $L(n)$ is i.i.d. , its sign is not. Actually, the process $\{B(n)\}_{n=0}^{\infty}$ is not even Markovian because $B(n)$ alone is not sufficient to predict future arrivals and departures. We, however, notice that if we enrich it with $(W_i^O(n),\ W_i^F(n))$ $(i = 1,\ 2)$, the elapsed waiting times for the endpoints of the next ON and OFF periods in $\mathbf{S}_i$, then it can be shown that the enriched process $\left\{B(n),\ \left(W_i^O(n),\ W_i^F(n)\right)_{i=1,\,2}\right\}_{n=0}^{\infty}$ is a Markov process. Thus, under ergodicity conditions, we can reduce (5) to a single-letter formula

$$T_b(\mathbf{S}_1, \mathbf{S}_2) = 1 - \mathbb{E}\left[\frac{(B + L - b)_+ + (-B - L)_+}{|L|}\right], \tag{6}$$

where $(B,\ L)$ are random variables with the limiting distribution of $\{(B(n-1),\ L'(n))\}$, assuming it exists. In fact,

$$T_b(\mathbf{S}_1,\,\mathbf{S}_2) = 1 - \limsup_{N\to\infty} \frac{\sum_{n=1}^{N} (B(n-1)+L'(n)-b)_+ + (-B(n-1)-L'(n))_+}{\sum_{n=1}^{N} |L'(n)|}, \tag{5}$$

it can be shown that $\left\{\left(W_i^O(n),\, W_i^F(n)\right)_{i=1,\,2}\right\}_{n=0}^{\infty}$ is already a Markov process, and thus $\{B(n)\}_{n=0}^{\infty}$ is a Markov-modulated random walk. In particular, the ergodicity condition holds for alternating Poisson processes (i.e., ON, OFF distributions are exponential).

4.2. Special Case: Renewal Processes

As pointed out in Section 3, if we ignore the packet lengths (e.g., when ON periods are far smaller than OFF periods), then the transmission schedules are reduced to point processes, and the optimal scheduling algorithm becomes BMR. Computing the maximum clandestine throughput is equivalent to computing the asymptotic fraction of information packets embedded by BMR. Specifically, let $B'(n)$ $(n \geq 0)$ be the number of stored packets after the nth arrival/departure packet ($B'(0) \equiv 0$), then it satisfies the following update

$$B'(n) = \begin{cases} B'(n-1)+1 & \text{if } B'(n-1) < b, S(n) \in \mathbf{S}_1, \\ B'(n-1)-1 & \text{if } B'(n-1) > 0, S(n) \in \mathbf{S}_2, \\ B'(n-1) & \text{o.w.} \end{cases}$$

Each time a self-loop occurs (i.e., $B'(n) = B'(n-1)$), the nth packet becomes chaff. Thus, the problem is reduced to counting self-loops in the process $(B'(n))_{n=0}^{\infty}$.

For i.i.d. Poisson processes, it was shown in [2] that the limiting probability of self-loops is $1/(1+b)$, implying a clandestine throughput of $b/(1+b)$. For general renewal processes, the clandestine throughput is characterized by the following theorem.

Theorem 4.1 *If $\mathbf{S}_i$ $(i = 1,\, 2)$ are i.i.d. renewal processes with absolutely continuous interarrival distribution, and $\exists \epsilon$, $u_0 > 0$ such that for all $u > u_0$,*

$$u \Pr\{U - u < V | U > u\} \geq \mathbb{E}[U] + \epsilon, \tag{7}$$

where U, V are i.i.d. random variables with the interarrival distribution, then the fraction of packets embedded by BMR converges a.s., and the limit (i.e., the maximum clandestine throughput) is bounded as

$$T_b(\mathbf{S}_1,\,\mathbf{S}_2) \gtreqless \frac{b}{1+b} \tag{8}$$

if

$$\Pr\{U - u > V | U > u\} \lesseqgtr \frac{1}{2} \tag{9}$$

for all $u \geq 0$, respectively.

Proof: The key is to construct a Markov-modulated random walk on $B'(n)$ by means similar to that in Section 4.1 and bound the limiting probability of self-loops. See the proof of Theorem 4.5 in [8] for details. ■

This theorem gives a sufficient condition for the convergence of BMR and provides qualitative characterization of the maximum clandestine throughput on point processes. Specifically, it compares the clandestine throughput of general renewal processes with that of Poisson processes, which is known. The comparison provides a lower or upper bound on the former, depending on whether the residual interarrival time $U - u$ is smaller or larger than the original. Intuitively, the smaller the residual interarrival time and thus the more likely a pending packet arrives earlier than a new packet, the higher the maximum clandestine throughput. This is because as the residual interarrival time becomes stochastically smaller than the original interarrival time, arrival and departure packets will interleave more regularly, reducing the probability of buffer underflow or overflow.

Theoretically, it is possible to compute the exact clandestine throughput by a Markovian model as in (6). We, however, choose to focus on characterizations that are easy to verify. The condition in (9) is further simplified as follows.

Corollary 4.2 *Under the assumptions in Theorem 4.1, the clandestine throughput $T_b(\mathbf{S}_1,\,\mathbf{S}_2)$ is lower bounded by, equal to, or upper bounded by $b/(1+b)$ if*

1. *$\Pr\{U - u > V | U > u\}$ is decreasing, independent, or increasing with u, which in turn is implied by that*
2. *$\Pr\{U - u > v | U > u\}$ is decreasing, independent, or increasing with u for all $v \geq 0$.*

Proof: See [8]. ■

Corollary 4.2 provides an easier way of bounding the clandestine throughput because condition 2 involves only one random variable. For example, for the *shifted Pareto distribution* with pdf

$$f^{\text{sPar}}(x) \triangleq \beta a^{\beta}(x+a)^{-\beta-1}, \quad x \geq 0, \tag{10}$$

$\Pr\{U - u > v | U > u\}$ is increasing with u for all v, and hence $T_b^{\text{sPar}} \le b/(1+b)$. For the uniform distribution, $\Pr\{U - u > v | U > u\}$ is decreasing with u, implying that $T_b^{\text{Uni}} \ge b/(1+b)$. Simulations in Section 5.1 have verified these results.

5. SIMULATIONS

5.1. Simulations on Point Processes

We first verify the analytical results in Section 4.2 by simulating the clandestine throughputs of various renewal processes using BMR. Fixing the mean interarrival time at $1/\lambda$, we simulate several types of interarrival distributions including the uniform, the exponential, and the shifted Pareto distributions. Using Poisson traffic as a benchmark, the uniform and the shifted Pareto distributions are selected to represent traffic with lower and higher burstiness, respectively. As discussed after Corollary 4.2, analysis predicts that the clandestine throughputs should decrease in the order of uniform, exponential, and shifted Pareto.

We plot the simulated clandestine throughputs as functions of the maximum buffer size b in Fig. 3. All the throughputs increase with b as expected. Moreover, the simulation results verify the above prediction. In particular, as the parameter β of the shifted Pareto distribution increases, its tailweight and hence burstiness decrease (tail probability $O(x^{-\beta})$), and the throughput increases. In the limit $\beta \to \infty$, the throughput will converge to that of the exponential distribution (not shown), coinciding with the fact that the distributions themselves converge. Furthermore, we have observed that none of the clandestine throughputs are functions of the traffic rate (i.e., λ). This is because the buffer size constraint only specifies the relative order of incoming and outgoing packets, and the actual timestamps are irrelevant.

5.2. Simulations on ON-OFF Processes

We then extend the scope to ON-OFF processes. For OFF periods under the exponential or the shifted Pareto distribution, we generate i.i.d. ON periods (i.e., packet lengths) according to constant, the uniform, and the exponential distributions respectively and plot the clandestine throughputs as functions of the mean packet length; see Fig. 4. The results continue to confirm that burstiness negatively impacts the achievable throughput: the constant, uniform, and exponential ON-period distributions have increasing levels of burstiness and decreasing clandestine throughputs (under the same OFF-period distribution), and similar conclusions can be drawn from the comparison between OFF-period

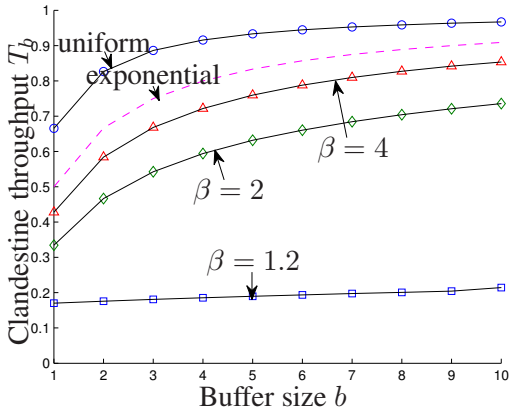

Fig. 3. Clandestine throughput of point processes ($\lambda = 1$, 10^5 packets per process): the uniform, exponential, and shifted Pareto (marked by β values) interarrival distributions.

distributions. The plot shows that the throughputs are not always monotone with the mean packet length, in contrast to the behavior with respect to buffer size (Fig. 3). This is because an increase in mean packet length has two effects: it reduces the burstiness of packet interarrivals[7] (OFF periods), but increases the burstiness of packet lengths (ON periods) and the chances of buffer violations (larger packets are more likely to cause buffer overflows/underflows).

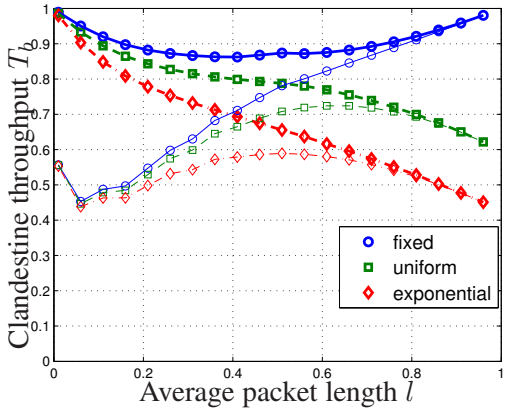

Fig. 4. Clandestine throughput of ON-OFF processes ($\lambda = 1$, $b = 1$, 10^4 packets per process, 100 Monte Carlo runs): the fixed, uniform, or exponential ON-period distribution (marked by legends) combined with the exponential (bold lines) or shifted Pareto (plain lines) OFF-period distribution.

[7] We have fixed the total traffic rate, i.e., the average sum of ON and OFF periods is fixed.

5.3. Simulations on Traces

We simulate the proposed algorithm on network traces to study the clandestine throughput in practice. We use the traces LBL-PKT-4, which contains an hour's worth of wide-area traffic between the Lawrence Berkeley Laboratory and the Internet[8]. As the traces only contain one timestamp per packet, we assume constant packet length per trace which is estimated by the minimum interarrival time. The simulated clandestine throughputs are then compared with those of alternating renewal processes with exponential or shifted Pareto inter-packet delays, as shown in Fig. 5. We find that the shifted Pareto distribution with $\beta = 0.5$ gives a good approximation of the traces (parameter of the exponential distribution does not affect the results), which is consistent with the previous studies in [9] that have claimed these traces to have Pareto-like interarrival distributions. Since $\beta < 1$ implies infinite mean interarrival and hence zero traffic rate, the result suggests that traces have much higher bustiness and lower clandestine throughputs than alternating renewal processes of the same rates.

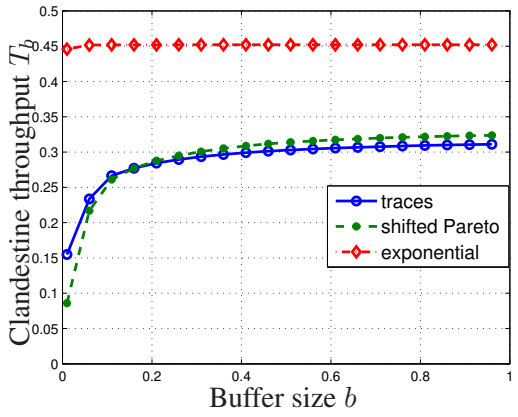

Fig. 5. Clandestine throughput of traces vs. alternating renewal processes with exponential or shifted Pareto inter-packet delays ($\beta = 0.5$, 10^3 packets per process, 17822 process pairs).

6. CONCLUSION

This paper presents an analytical study of the maximum throughput of a clandestine relay under stochastic transmission schedules. Under a buffer size constraint at the relay node, we develop a scheduling algorithm that can embed the maximum amount of relayed traffic into predetermined schedules and analyze the efficiency of this algorithm for schedules following independent ON-OFF renewal processes. Our results provide answers to fundamental questions including how to hide information flows without covert channels and how the rates of such flows are affected by the resource constraints of relay nodes and the statistical properties of transmission schedules.

[8]The traces were collected by Paxson and first used in his paper [9], from which we extract 134 TCP traces of 1000 packets each. The traces can be obtained from `http://ita.ee.lbl.gov/html/contrib/LBL-PKT.html`.

7. REFERENCES

[1] Y. Zhang, W. Liu, W. Lou, and Y. Fang, "MASK: Anonymous On-Demand Routing in Mobile Ad Hoc Networks," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 2376–2385, Sept 2006.

[2] T. He and L. Tong, "Detection of Information Flows," *IEEE Transactions on Information Theory*, vol. 54, pp. 4925–4945, November 2008.

[3] B.Radosavljevic and B. Hajek, "Hiding traffic flow in communication networks," in *Military Communications Conference*, 1992.

[4] Y. Hu, A. Perrig, and D. B. Johnson, "Packet Leashes: A Defense against Wormhole Attacks in Wireless Ad Hoc Networks," in *Proc. IEEE INFOCOM*, (San Francisco, CA), March 2003.

[5] D. Chaum, "Untraceable electronic mail, return addresses and digital pseudonyms," *Communications of the ACM*, vol. 24, pp. 84–88, February 1981.

[6] T. He, A. Agaskar, and L. Tong, "On Security-Aware Transmission Scheduling," in *Proc. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'08)*, (Las Vegas, NV), March 2008.

[7] P. Venkitasubramaniam, T. He, and L. Tong, "Anonymous Networking amidst Eavesdroppers," *IEEE Transactions on Information Theory, Special Issue on Information-Thoeretic Security*, vol. 54, pp. 2770–2784, June 2008.

[8] T. He, L. Tong, and A. Swami, "Maximum Throughput of Clandestine Information Flows." draft, 2009.

[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.