Distributed Detection of Information Flows in Chaff

Ting He and Lang Tong
School of Electrical and Computer Engineering,
Cornell University, Ithaca, NY 14853.
Email: {th255, lt35}@cornell.edu.

***Abstract*— Distributed detection of information flows is considered. The detector detects the presence of information flows by collecting timing information from nodes of interest through channels of finite capacity. The information flows are assumed to be perturbed up to a bounded delay and interleaved with chaff. Joint compression and detection schemes are proposed to achieve reliable detection with inaccurate measurements. Detection performance is analytically evaluated by robustness against chaff as functions of the capacity constraints in the data collection. The proposed detectors are proved to be optimal for their corresponding quantizers. A comparison of their performance gives guidelines on quantizer design.**

***Index Terms*— Information flow detection, Robust detection algorithms, Quantizer design, Distributed detection.**

I. Introduction

Consider the wireless ad hoc network illustrated in Fig. 1, where nodes B and C are connected by a subnetwork. Eavesdroppers are deployed to record the timing of transmissions intended for B and C respectively[1]. Let S_1 denote the timing measurements for B, and S_2 for C. Then S_i $(i = 1, 2)$ is a point process defined as

$$S_i = (S_i(1),\ S_i(2),\ S_i(3), \ldots),$$

where $S_i(k)$ is the transmission epoch of the kth packet on S_i. We want to detect whether there are information flows using B as a relay node to reach C by collecting measurements of S_1 and S_2 through channels of finite capacity.

We assume that the transmission of information flows satisfies the following definition.

Definition 1.1: A pair of processes $(F_1,\ F_2)$ forms an *information flow* if for every realization of F_i $(i = 1, 2)$, there exists a bijection[2] $g : \mathcal{F}_1 \to \mathcal{F}_2$ such that $g(s) - s \in [0,\ \Delta]$ for all $s \in \mathcal{F}_1$.

The bijection g is a mapping between the arrival epochs of the same packets at B and C respectively, allowing perturbations and permutations during the relay. The condition that g is a bijection imposes a *packet-conservation* constraint, *i.e.*, every packet at B generates one and only one relay packet at C. The condition $g(s) - s \geq 0$ is a *causality* constraint, which means that a packet cannot arrive at C before it reaches B. The condition $g(s) - s \leq \Delta$ imposes a *bounded delay* constraint that the maximum delay from B to C is bounded by Δ.

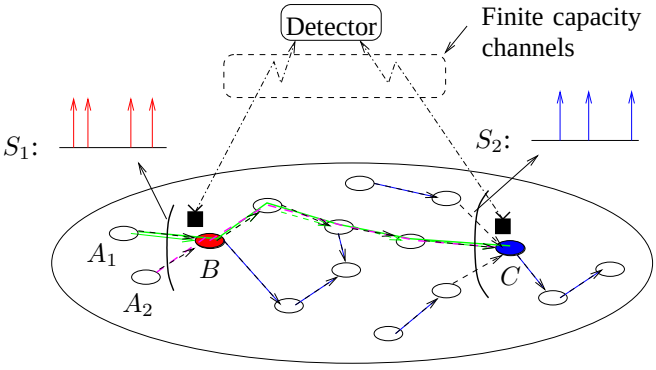

Fig. 1. Detecting information flows along $A_j \to B \to \ldots \to C$ $(j = 1, 2)$ by distributing eavesdroppers (■) to collect timing measurements of the incoming traffic to B and C and report to a detector through channels of finite capacity.

In practice, nodes multiplex their transmissions, or they may introduce dummy packets deliberately. Transmissions that are not part of the information flow are called *chaff*. A pair of processes $(S_1,\ S_2)$ *contains an information flow* if for every realization[3], $S_i = F_i \bigoplus C_i$ $(i = 1, 2)$, where $(F_1,\ F_2)$ is an information flow, and C_i is the subsequence of S_i consisting of chaff packets.

We are interested in testing the following hypotheses:

$$\begin{aligned} \mathcal{H}_0 :&\quad S_1 \text{ and } S_2 \text{ are independent,} \\ \mathcal{H}_1 :&\quad (S_1,\ S_2) \text{ contains an information flow,} \end{aligned}$$

by the distributed detection system illustrated in Fig. 2.

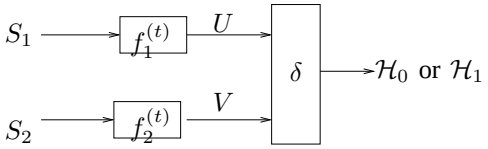

Fig. 2. Distributed detection: the detector makes decisions based on indices $(U,\ V)$ compressed from $(S_1,\ S_2)$.

In a distributed detection system, eavesdroppers observe the part of S_i $(i = 1, 2)$ that falls into the observation window

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, and TRUST (The Team for Research in Ubiquitous Secure Technology) sponsored by the National Science Fundation under award CCF-0424422. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

[1]We assume receiver code, and the eavesdroppers know the code of B and C. The same formulation applies to transmitter code.

[2]Given a realization of a point process F_i, $\mathcal{F}_i$ denotes the set of elements in this realization.

[3]For increasing sequences $(a_1,\ a_2, \ldots)$ and $(b_1,\ b_2, \ldots)$, define $(a_i)_{i=1}^{\infty} \bigoplus (b_i)_{i=1}^{\infty} \stackrel{\Delta}{=} (c_i)_{i=1}^{\infty}$, where $c_1 \leq c_2 \leq \ldots$ and $\{a_i\}_{i=1}^{\infty} \cup \{b_i\}_{i=1}^{\infty} = \{c_i\}_{i=1}^{\infty}$.

$[t_0,\ t_0+t]$ (t_0 is the starting time and t is the duration of the observation) and compress the measurements by mappings $f_i^{(t)}$ to indices U and V in finite sets of size $||f_i^{(t)}||$. Under capacity constraints $(R_1,\ R_2)$, the sets have to satisfy[4]

$$\limsup_{t\to\infty}\frac{1}{t}\log||f_i^{(t)}||\le R_i,\qquad i=1,\ 2. \tag{1}$$

Then the detector makes a decision based on $(U,\ V)$.

Assume that the detector knows Δ but not t_0 or traffic before t_0, and S_i ($i=1,\ 2$) have the same marginal distributions under both hypotheses[5]. For detailed analysis, we assume that S_i ($i=1,\ 2$) are marginally Poisson processes. Note that we do not impose parametric models for the joint distribution of S_1 and S_2 under $\mathcal{H}_1$.

A. Related Work

The detection of information flows is a timing analysis problem, which is a special case of traffic analysis [1]. Centralized detection of information flow detection has been studied as a countermeasure to stepping-stone attacks. Since Zhang and Paxson [2] first studied the problem from the perspective of timing analysis, various timing-based detection schemes have been proposed to deal with perturbations such as random delays and reordering; see [3]–[5]. The results of [5]–[7] show that reliable detection can be achieved even with actively injected chaff packets, although they all impose constraints on how chaff packets can be inserted[6]. For arbitrary chaff insertion, we proposed the first timing-based algorithms that achieve consistent detection in the presence of a number of chaff packets increasing proportionally to the total traffic size [8].

Distributed detection of information flows belongs to statistical inference under multiterminal data compression [9]. In their survey paper [9], Han and Amari presented results on the performance of distributed detection based on the assumption that measurements are *i.i.d.* over time, and hypotheses are parametric. Our problem is fundamental different in that the timing measurements are generally not *i.i.d.* over time, and the hypotheses are not parametric.

The subproblem of compressing timing measurements modelled as Poisson processes has been studied previously. Rubin in [10] and Verdú in [11] derived the rate-distortion functions for absolute-error distortion and an asymmetric distortion measure, respectively. One of our quantization schemes is the same as that of Rubin, but our quantizers are evaluated by the overall detection performance rather than the fidelity in reconstructing the processes.

B. Summary of Results and Organization

This paper addresses the distributed detection of information flows by timing analysis. Our goal is to develop joint compression and detection schemes to achieve consistent detection for the broadest class of information flows under given capacity constraints.

We divide the detection scheme into three stages: compression, data transmission, and detection. The compression stage compresses timing measurements to satisfy the capacity constraints; the data transmission stage transmits the compressed measurements to the detector; the detection stage makes decisions based on the received measurements.

For compression, we propose a slotted quantizer which quantizes epochs to multiples of a slot length and a one-bit quantizer which further compresses the slotted epochs to indicators of nonempty slots. For each quantizer, we develop a threshold detector based on the minimum number of chaff packets required to embed information flows. Detection performance is evaluated by the maximum amount of chaff such that the detector is Chernoff-consistent. The proposed detectors are optimal for their corresponding quantizers in that they achieve Chernoff-consistent detection against the maximum amount of chaff. Analytical comparison of their performance suggests a correlation between traffic rate and the performance of detection schemes.

The rest of the paper is organized as follows. Section II defines the quantizers, based on which Section III presents detection algorithms and consistency analysis. Optimality of the algorithms is proved in Section IV, followed by an analytical comparison of the algorithms. Then Section V concludes the paper with remarks on its contributions.

II. Quantizer Design

In this section, we decompose the mapping $f_i^{(t)}$ ($i=1,\ 2$) into a quantizer and a lossless encoder and introduce two simple quantizers invariant with t.

Definition 2.1: Given a point process S, a *slotted quantizer*[7] with slot length T is defined as $\gamma(S)\stackrel{\Delta}{=}(Z_1,\ Z_2,\ldots)$, where Z_j is the number of points in S in the jth slot (*i.e.,* the interval $[(j-1)T,\ jT)$).

Quantization by a slotted quantizer is called *slotted quantization*. It is easy to see that the above definition is equivalent to the point-by-point quantizer $\tilde{\gamma}(t)=\lfloor t/T\rfloor$, where $t\in\mathbb{R}^+$.

The quantization results of a slotted quantizer can be further compressed by the following quantizer.

Definition 2.2: Given a point process S, a *one-bit quantizer* with slot length T is defined as $\hat{\gamma}(S)\stackrel{\Delta}{=}(Z_1,\ Z_2,\ldots)$, where Z_j is the indicator that the jth slot is nonempty.

Quantization by a one-bit quantizer is called *one-bit quantization*.

Let $\mathbf{X}^N=(X_1,\ldots,X_N)$ be the result of slotted or one-bit quantization of S_1 in $[t_0,\ t_0+NT]$. By the source coding theory [12], the condition in (1) is achievable if

$$\limsup_{NT\to\infty}\frac{H(\mathbf{X}^N)}{NT}\le R_1. \tag{2}$$

[4]All the logarithm in this paper is natural log.

[5]This is to avoid the case that a single eavesdropper can make decisions based only on its measurements.

[6]For example, in [6], [7], chaff is only inserted in S_2, and in [5], the number of chaff packets in every predefined number of packets is bounded.

[7]The same quantizer was used in [10] and was shown to approximate the optimal compression performance under the absolute-error fidelity critetion.

If S_1 is a Poisson process, then X_j's are *i.i.d.* , and the condition in (2) is simplified to

$$\limsup_{T\to\infty} \frac{H(X_1)}{T} \le R_1. \tag{3}$$

The same holds for $\mathbf{Y}^N$, the quantization result of S_2.

III. Consistent Detection Algorithms

In this section, we will present detection algorithms for each of the quantizers defined in Section II and analyze their performance.

Due to the nonparametricness of information flows, we propose a performance measure based on the allowable amount of chaff. We introduce the following definition.

Definition 3.1: Given a realization of an information flow (S_1, S_2), its *chaff-to-traffic ratio* (CTR) in an interval $[t_0, t_0+t]$ is defined as[8]

$$\mathrm{CTR}(t;\, t_0) = \frac{\sum_{i=1}^{2} |\mathcal{C}_i \cap [t_0,\, t_0+t]|}{\sum_{i=1}^{2} |\mathcal{S}_i \cap [t_0,\, t_0+t]|}.$$

Using CTR to measure the amount of chaff, we introduce the following performance measure.

Definition 3.2: The *r-consistency* of a detection system $(f_1^{(t)},\, f_2^{(t)},\, \delta_t)$ is the supremum of γ ($\gamma \in [0,\, 1]$) such that

1) $\lim_{t\to\infty} P_F(\delta_t) = 0$;
2) $\sup_{(S_i)_{i=1}^2 \in \mathcal{P}} \lim_{t\to\infty} P_M(\delta_t) = 0$, where[9]

$$\mathcal{P} = \{(S_i)_{i=1}^2 :\ (S_i)_{i=1}^2 \text{ contains an information flow, and } \limsup_{t\to\infty} \mathrm{CTR}(t;\, t_0) \le \gamma \text{ a.s.}\}.$$

That is, the r-consistency is the maximum fraction of chaff such that δ_t is always Chernoff-consistent no matter how the chaff is inserted. Our goal is to design a system with the maximum r-consistency under given capacity constraints.

The idea behind the notion of r-consistency is that since the detector does not know how information flows are perturbed and chaff packets are inserted, it tries to make sure that it requires a sufficiently large amount of chaff to evade detection. In the sequel, symmetric capacity constraints (*i.e.,* $R_1 = R_2 = R$) will be considered for simplicity.

A. Detection under Slotted Quantization

Suppose that slotted quantizers with slot length T are used to compress S_i ($i = 1,\, 2$). The idea of detection is to compute the minimum CTR required to generate the quantized measurements, and make a decision by comparing the computed CTR with a threshold.

To compute the minimum CTR, we borrow the idea of an algorithm called "Bounded-Greedy-Match" (BGM) proposed by Blum *et al.* in [5]. Given a delay bound Δ, BGM sequentially matches every point s in S_1 with the first unmatched point in $[s,\, s+\Delta]$ in S_2. Then the matched pairs form an information flow, and the unmatched points become chaff. It is shown in [5] that BGM inserts the minimum number of chaff packets for any realization of $(S_1,\, S_2)$.

Given quantized measurements $(\mathbf{X}^N,\, \mathbf{Y}^N)$, we need to reconstruct the processes of transmission epochs such that the required CTR is minimized. The detector, denoted by δ_{l}, works as follows:

1) construct point processes $\hat{S}_i$ ($i = 1,\, 2$) as bursts of X_j (or Y_j) simultaneous points at $(j-1)T$ for $j \ge 1$ (as illustrated in Fig. 3);
2) run BGM on $(\hat{S}_1, \hat{S}_2)$ with delay bound $\lceil \frac{\Delta}{T} \rceil T$; let C_N be the number of unmatched points, excluding[10] unmatched points in the first $\lceil \frac{\Delta}{T} \rceil$ slots in $\hat{S}_2$;
3) if $C_N/N \le \tau_{\mathrm{l}} - \epsilon$, return $\mathcal{H}_1$; otherwise, return $\mathcal{H}_0$.

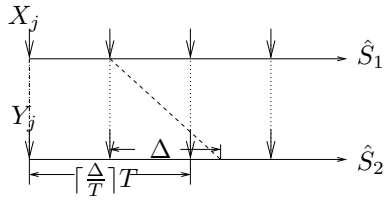

Fig. 3. δ_{l}: construct batched point processes $(\hat{S}_1, \hat{S}_2)$ from $(\mathbf{X}, \mathbf{Y})$; compute the minimum number of chaff packets in $(\hat{S}_1,\, \hat{S}_2)$ by greedy matching; make decisions by comparing the calculated number of chaff packets with a threshold.

Under $\mathcal{H}_1$, the actual number of chaff packets is lower bounded by C_N. This is because that a packet in slot j can only be matched to packets in slots $j, \ldots,\, j + \lceil \frac{\Delta}{T} \rceil$, which corresponds to a delay bound $\lceil \frac{\Delta}{T} \rceil T$, and BGM computes the minimum chaff for given delay bound. Therefore, δ_{l} has no miss detection for any realization of information flow with up to $\tau_{\mathrm{l}} - \epsilon$ chaff packets per slot.

Under $\mathcal{H}_0$, the performance of δ_{l} is characterized by the following theorem.

Theorem 3.3: If S_1 and S_2 are independent Poisson processes of equal rate λ, T is large, and $\tau_{\mathrm{l}} = c_1\sqrt{\lambda T}e^{-\lambda T/6}$ where $c_1 = 0.0014$, then for any $\epsilon > 0$, the false alarm probability of δ_{l} decays exponentially with N.

Proof: See Appendix. ∎

Remark: The theorem gives a sufficient condition on τ_{l} to guarantee vanishing false alarm probability. Combining this result with the arguments on miss detection yields that the r-consistency of δ_{l} is lower bounded by $r_{\mathrm{l}}(T) \triangleq \tau_{\mathrm{l}}/(2\lambda T) = c_1/(2\sqrt{\lambda T})e^{-\lambda T/6}$.

Under Poisson assumption, it is know that the rate to reliably deliver $\mathbf{X}^N$ and $\mathbf{Y}^N$ for large N is $R_{\mathrm{l}}(T) \triangleq H(\mathrm{Poi}(\lambda T))/T$, where $H(\mathrm{Poi}(\lambda T))$ is the entropy of Poisson distribution with mean λT. Therefore, δ_{l} achieves the rate-consistency pair $(R_{\mathrm{l}}(T),\, r_{\mathrm{l}}(T))$; since they are both decreasing functions of T, they define a consistency-rate function by $r_{\mathrm{l}}(R_{\mathrm{l}}^{-1}(R))$.

[8] We use $\mathcal{C}_i$ and $\mathcal{S}_i$ to denote the sets of elements in the realizations of C_i and S_i, respectively.

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

[10] This adjustment is needed because these packets may be the relay of packets in S_1 sent before the detector starts.

B. Detection under One-Bit Quantization

Suppose that S_i ($i = 1, 2$) are compressed by one-bit quantizers with slot length T. The detection algorithm in Section III-A is still applicable, but the threshold will be different. Specifically, the detector under one-bit quantization, denoted by δ_{II}, is defined the same as δ_{I} except that the threshold is changed to $\tau_{\text{II}} - \epsilon$.

Under $\mathcal{H}_1$, it is easy to see that it minimizes the need of chaff to put only one packet in a nonempty slot. Combining this observation with the arguments in Section III-A yields that C_N is a lower bound on the actual number of chaff packets, and thus δ_{II} has no miss detection for any realization of information flow with the average number of chaff packets per slot bounded by $\tau_{\text{II}} - \epsilon$.

Under $\mathcal{H}_0$, we have the following theorem on the false alarm probability.

Theorem 3.4: If S_1 and S_2 are independent Poisson processes of equal rate λ, T is large, and $\tau_{\text{II}} = e^{-2\lambda T}(1 - e^{-\lambda T})$, then for any $\epsilon > 0$, the false alarm probability of δ_{II} decays exponentially with N.

Proof: See Appendix. ■

Remark: By Theorem 3.4 and the arguments on miss detection, we see that the r-consistency of δ_{II} is at least $r_{\text{II}}(T) \triangleq \tau_{\text{II}}/(2\lambda T)$.

Since it is known that the rate needed to transmit $\mathbf{X}^N$ and $\mathbf{Y}^N$ reliably is[11] $R_{\text{II}}(T) \triangleq h(e^{-\lambda T})/T$, we can obtain a lower bound on the consistency-rate function of δ_{II} by $r_{\text{II}}(R_{\text{II}}^{-1}(R))$.

Note that for the same T, $r_{\text{II}}(T)$ decays 12 times faster than $r_{\text{I}}(T)$, reflecting the information loss due to further compression. It is, however, not clear that slotted quantization is better than one-bit quantization because one-bit quantization can afford to use a much smaller T under the same capacity constraints.

IV. Optimality and Comparison

We have divided the entire detection scheme into three stages—compression, data transmission, and detection. In this section, we will show that the proposed detectors are optimal for their corresponding quantizers and then discuss the heuristics on quantizer design.

A. Optimality of Detectors

In [13], we show that the proposed detectors are optimal for their corresponding quantizers in terms of r-consistency, as stated in the following theorem.

Theorem 4.1: Assume that S_1 and S_2 are independent Poisson processes under $\mathcal{H}_0$. For any detector under slotted quantization, there exists τ_{I} and ϵ such that the r-consistency of δ_{I} is no smaller than the r-consistency of that detector[12]. Similar result holds for δ_{II} and one-bit quantization.

[11]Function $h(p)$ is the binary entropy function defined by $h(p) = -p \log p - (1-p) \log (1-p)$.

[12]Note that by the definition of r-consistency, we require the detector to have vanishing miss probability for all the chaff insertion methods. This is different from other related work which only addresses a certain insertion method.

Remark: The theorem says that the proposed detectors achieve the maximum r-consistency under the corresponding quantization schemes. The values of τ_i ($i =$ I, II) in Theorem 3.3 and 3.4 are lower bounds on the optimal thresholds.

B. Comparison of Quantizers

Having established the optimality of the detectors, we can compare the proposed quantizers by the corresponding detection performance. Specifically, we compare their consistency-rate functions $r_i(R_i^{-1}(R))$ ($i =$ I, II) for different traffic rates[13] over a range of R such that $R_i^{-1}(R) \geq 2\Delta$; see Fig. 4–6.

The plots provide the following observations: i) for light traffic (Fig. 4), slotted quantization is better than one-bit quantization, whereas for traffic not too light (Fig. 5–6), one-bit quantization performs better; ii) the performance deteriorates quickly as traffic rate increases (*e.g.*, when λ increases 10 times from Fig. 4 to Fig. 5, the vertical scale drops by a factor of 10^{-2}).

These observations yield several insights into the problem. Observations (i) suggest that heavy traffic requires significant compression to efficiently utilize the capacity, whereas light traffic should be reported to more detail. Furthermore, if we normalize the maximum delay by the average interarrival time, the normalized maximum delay is $\lambda\Delta$, which suggests that higher traffic rate relaxes the delay constraint and therefore makes the detection more difficult, as is confirmed by observation (ii).

V. Conclusion

This paper proves that consistent detection of information flows is achievable even if the information flows are perturbed and mixed with chaff, and the measurements are limited in accuracy. Analysis of the proposed detection schemes shows that detectors are consistent against chaff processes of positive rate under arbitrarily small capacities. Although Poisson assumption has been made for traffic under the null hypothesis, it can be shown that it requires more chaff noise to embed information flows into real-world traffic, and thus the results in this paper serve as lower bounds on the detection performance in practice [13].

VI. Appendix

A. Proof of Theorem 3.3

Let $\tilde{C}_{2i}$ ($i = 1, 2, \ldots$) be the number of chaff packets in the $(2i)$th slot if Step 2) of δ_{I} is only performed on even slots. Obviously, $C_N \geq \sum_{i=1}^{\lfloor N/2 \rfloor} \tilde{C}_{2i}$, and the false alarm probability satisfies

$$P_F(\delta_{\text{I}}) = \Pr\{C_N/N \leq \tau\} \leq \Pr\{\frac{2}{N} \sum_{i=1}^{\lfloor N/2 \rfloor} \tilde{C}_{2i} \leq 2\tau\},$$

where $\tau = \tau_{\text{I}} - \epsilon$. It is easy to see that $\tilde{C}_2, \tilde{C}_4, \tilde{C}_6, \ldots$ are *i.i.d.* . By Cramer's Theorem [14], we can prove that $P_F(\delta_{\text{I}})$ decays exponentially if we show that $\mathbb{E}[\tilde{C}_2] \geq 2\tau_{\text{I}}$.

[13]The traffic rate λ is the rate of S_i ($i = 1, 2$), whereas R is the capacity of the uplink channels.

The robustness-rate functions of δ_i (i = I, II) under different traffic rates

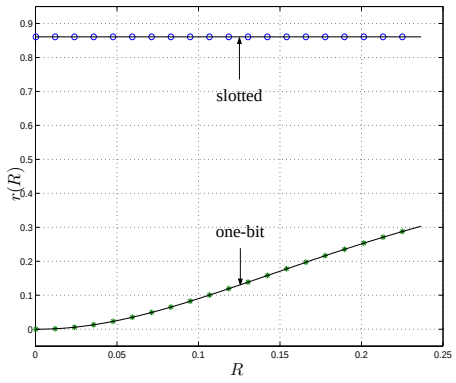

Fig. 4. $\lambda = 0.1$, $\Delta = 1$.

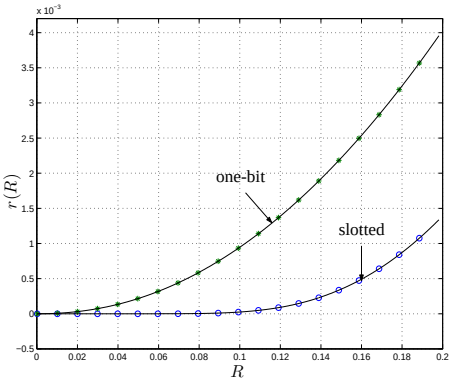

Fig. 5. $\lambda = 1$, $\Delta = 1$.

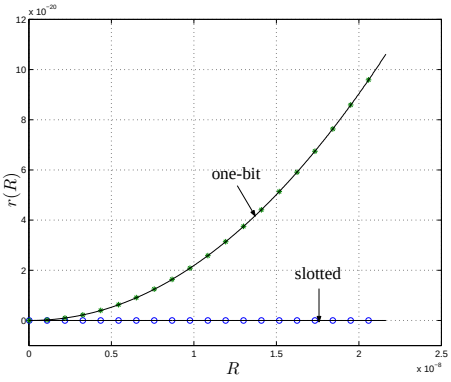

Fig. 6. $\lambda = 10$, $\Delta = 1$.

Since $\mathbb{E}[\tilde{C}_2] = \mathbb{E}[\max(Y_2 - X_1 - X_2,\ 0) + \max(X_2 - Y_2 - Y_3,\ 0)]$, and by Gaussian approximation, $(Y_2 - X_1 - X_2)$, $(X_2 - Y_2 - Y_3) \sim \mathcal{N}(-\lambda T,\ 3\lambda T)$ for large T, we have

$$\begin{aligned}\frac{1}{2}\mathbb{E}[\tilde{C}_2] &\approx \int_0^\infty \frac{z}{\sqrt{6\pi\lambda T}} e^{-(z+\lambda T)^2/(6\lambda T)} dz \\ &= \sqrt{\frac{3\lambda T}{2\pi}} e^{-\lambda T/6} - \lambda T Q\left(\sqrt{\frac{\lambda T}{3}}\right) \\ &\approx c_1 \sqrt{\lambda T} e^{-\lambda T/6} = \tau_{\text{I}}, \end{aligned} \tag{4}$$

where (4) is obtained by the approximation of $Q(\cdot)$ in [15]. ■

B. Proof of Theorem 3.4

Let T_i ($i \geq 1$) denote the number of slots between the $(i-1)$th and the ith chaff packets (including the slot with the ith chaff packet) found by δ_{II}. Then the false alarm probability can be written as

$$P_F(\delta_{\text{II}}) = \Pr\{\frac{C_N}{N} \leq \tau\} = \Pr\{\frac{1}{N\tau} \sum_{i=1}^{N\tau+1} T_i > \frac{1}{\tau}\},$$

where $\tau = \tau_{\text{II}} - \epsilon$. Now let $\tilde{T}_i$ be the number of slots between chaff packets if Step 2) of δ_{II} is only performed on even slots. Then $T_i \leq \tilde{T}_i$, and we have $P_F(\delta_{\text{II}}) \leq \Pr\{\frac{1}{N\tau} \sum_{i=1}^{N\tau+1} \tilde{T}_i > \frac{1}{\tau}\}$. Since $\tilde{T}_i$'s are *i.i.d.* , by Cramer's Theorem, we can prove the exponential decay of $P_F(\delta_{\text{II}})$ if $\mathbb{E}[\tilde{T}_1] \leq 1/\tau_{\text{II}}$.

Note that if Step 2) of δ_{II} is only applied to even slots, the event $\{\exists$ chaff in slot $2j\}$ is equivalent to

$$A_{2j} \stackrel{\Delta}{=} \{X_{2j-1} + X_{2j} < Y_{2j},\ \text{or}\ X_{2j} > Y_{2j} + Y_{2j+1}\},$$

which has probability $\rho = 2e^{-2\lambda T}(1 - e^{-\lambda T})$, and is *i.i.d.* for $j = 1, 2, \ldots$. Let $Z = \inf\{j \geq 1 : A_{2j} \text{ occurs}\}$. Then Z is the number of slot pairs until the first chaff packet, and $\tilde{T}_1 = 2Z$. It is easy to see that Z has the geometric distribution

$$\Pr\{Z = n\} = (1 - \rho)^{n-1}\rho, \quad n \geq 1.$$

Therefore, $\mathbb{E}[\tilde{T}_1] = 2\mathbb{E}[Z] = 2/\rho = 1/\tau_{\text{II}}$.

■

REFERENCES

[1] N. Ferguson and B. Schneier, *Practical Cryptography*. Indianapolis,IN: John Wiley & Sons, Inc., 2003.

[2] Y. Zhang and V. Paxson, "Detecting stepping stones," in *Proc. the 9th USENIX Security Symposium*, pp. 171–184, August 2000.

[3] 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.

[4] X. Wang and D. Reeves, "Robust correlation of encrypted attack traffic through stepping stones by manipulation of inter-packet delays," in *Proc. of the 2003 ACM Conference on Computer and Communications Security*, pp. 20–29, 2003.

[5] 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.

[6] P. Peng, P. Ning, D. Reeves, and X. Wang, "Active Timing-Based Correlation of Perturbed Traffic Flows with Chaff Packets," in *Proc. 25th IEEE International Conference on Distributed Computing Systems Workshops*, (Columbus, OH), pp. 107–113, June 2005.

[7] L. Zhang, A. Persaud, A. Johson, and Y. Guan, "Stepping Stone Attack Attribution in Non-cooperative IP Networks," in *Proc. of the 25th IEEE International Performance Computing and Communications Conference (IPCCC 2006)*, (Phoenix, AZ), April 2006.

[8] 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.

[9] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2300–2324, Oct. 1998.

[10] I. Rubin, "Information Rates and Data-Compression Schemes for Poisson Processes," *IEEE Transactions on Information Theory*, vol. 20, pp. 200–210, March 1974.

[11] S. Verdú, "The Exponential Distribution in Information Theory," *Problems of Information Transmission*, vol. 32, no. 1, pp. 86–95, 1996.

[12] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[13] T. He and L. Tong, "Distributed Detection of Information Flows," Tech. Rep. ACSP-TR-04-07-01, Cornell University, April 2007. `http://acsp.ece.cornell.edu/pubR.html`.

[14] F. den Hollander, *Large Deviations (Fields Institute Monographs, 14)*. American Mathematical Society, 2000.

[15] N. Kingsbury, "Approximation formulae for the Gaussian error integral Q(x)," Tech. Rep. m11067, Connexions, June 2005. `http://cnx.org/content/m11067/latest/`.