Distributed Detection of Information Flows with Side-Information

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

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

***Abstract*— Distributed detection of information flows by timing analysis is considered. Timing measurements are subject to perturbations and the insertion of chaff noise. Moreover, communications among distributed traffic sensors are subject to capacity constraints. With the assumption that the detector is co-located at a point of measurement, the problem is formulated as a distributed hypothesis testing with side-information. The main challenge is that there is no parametric model for the relaying of information flows. Under the assumption that the relaying of flows is subject to bounded delays, distributed detection systems are designed by converting a centralized detector to distributed detectors under slot-based quantization. The proposed systems are evaluated both theoretically and numerically by the proposed performance measure as a function of the capacity in data collection. Numerical evaluation shows that separately designing data compression and detection modules can achieve satisfactory overall performance.**

***Index Terms*— Information flow, Distributed detection, Side-information, Chernoff-consistency.**

I. Introduction

Consider a wireless ad hoc network illustrated in Fig. 1. Given potential routes through nodes A and B, we are interested in knowing whether these routes are being used by eavesdropping the ongoing traffic. Each eavesdropper is equipped with an energy detector to record the transmission epochs of the node of interest. Assume negligible measurement error. Then one of the eavesdroppers, *e.g.,* the one at node A, reports the (compressed) measurements through a channel of limited capacity to the other eavesdropper, where a detector will combine the measurements to make decisions.

This is a problem of detecting information flows by timing analysis [1]. The advantage of timing analysis is that the measurements are easy to obtain because of the shared medium, and the detection is applicable even if all the packets are encrypted and padded. The challenges include perturbations due to delays, permutations, etc., capacity constraint in the communication channel, and *chaff noise*. Chaff noise models unrelated traffic multiplexed at intermediate nodes as well as dummy traffic artificially inserted by the nodes to evade detection. Suppose that the detector is located at one of the eavesdroppers, which implies that the raw measurements of that eavesdropper are perfectly known by the detector, referred to as the case of *full side-information*.

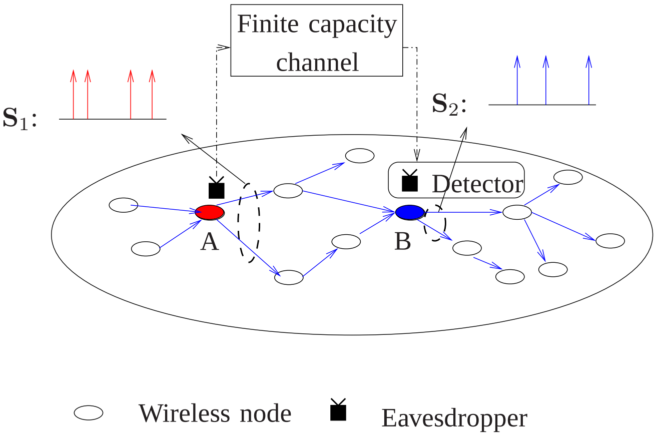

Fig. 1. In a wireless ad hoc network, eavesdroppers are deployed to collect processes of transmission epochs, denoted by $\mathbf{S}_i$ ($i = 1, 2$), from nodes A and B, respectively. Then the eavesdropper at A sends measurements compressed from $\mathbf{S}_1$ to B through a finite capacity channel, and a detector at the eavesdropper at B detects information flows through A and B based on the received measurements and $\mathbf{S}_2$.

Applications of this problem include intrusion detection and network monitoring in wireless ad hoc networks such as wireless sensor networks.

A. Related Work

The problem of information flow detection has been studied in the context of Internet as a countermeasure to stepping-stone attacks (see [2]). Since Zhang and Paxson [3] first took a timing-based approach to stepping-stone detection, the problem has evolved to allow various timing perturbations such as bounded delay perturbation considered in [4] and bounded memory perturbation in [1]. Recent work aims at handling perturbations as well as actively injected chaff noise; see [1], [5], [6]. All these techniques assume that the raw measurements are fully available at the detector.

Under data compression, Ahlswede and Csiszár in [7] considered a related problem of testing the correlation between two random variables by taking temporally *i.i.d.* sample pairs under limited capacities. Our problem is fundamentally different from theirs because timing measurements of flows are not

This work was completed when Ting He was at Cornell. It 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. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

i.i.d. over time. In previous work [8], we consider the detection of information flows under equal capacity constraints. In this paper, we consider the case when the detector is located at a point of measurement, which brings side-information to detection.

B. Summary of Results and Organization

We consider the distributed detection of information flows by timing analysis with side-information. Specifically, by deploying traffic sensors to measure the transmission epochs of two nodes and fusing the measurements at one of the sensors, we want to detect packet flows relayed through these nodes. Based on the assumption that nodes' normal transmission behavior is known, and the relay of packets is subject to strict delay constraint, we formulate the problem as a partially nonparametric hypothesis testing under constraints on the fusion rate.

Our goal is to define a rigorous, analytical framework for the design of distributed flow detection systems and design practical compression and detection schemes within this framework. Generalizing the centralized approach to information flow detection in [1], we take a separated approach, dividing the detection procedure into two stages: data compression and detection. For data compression, we propose a couple of counting-based quantizers based on slotted structure. For detection, we develop threshold detectors based on the optimal chaff-inserting algorithms to detect the flows of sufficiently high rate. We show that the proposed detector can achieve Chernoff-consistent detection for chaff noise of positive rate even if the fusion capacity is very small. In particular, given a fusion capacity, we derive consistency-rate functions which map the capacity to the maximum fraction of chaff noise for which consistent detection is guaranteed. Comparison of the proposed systems shows that slotted quantization, which is known (see [9]) to have good performance in compressing Poisson processes, is also preferable in compressing traffic measurements for the purpose of detection.

The rest of the paper is organized as follows. Section II formulates the problem. Section III briefly summarizes existing results on centralized flow detection. Sections IV and V present a couple of distributed detection systems. Section VI compares the proposed systems numerically. Then Section VII concludes the paper with remarks.

II. The Problem Formulation

We use the convention that uppercase letters denote random variables, lowercase letters realizations, boldface letters vectors, and plain letters scalars. Given a realization of a point process $\mathbf{s}$, we use $\mathcal{S}$ to denote the set of elements in this realization. Given two realizations of point processes $(a_1, a_2, \ldots)$ and $(b_1, b_2, \ldots)$, $\bigoplus$ is the *superposition operator* defined as $(a_k)_{k=1}^{\infty} \bigoplus (b_k)_{k=1}^{\infty} = (c_k)_{k=1}^{\infty}$, where $c_1 \leq c_2 \leq \ldots$ and $\{a_k\}_{k=1}^{\infty} \cup \{b_k\}_{k=1}^{\infty} = \{c_k\}_{k=1}^{\infty}$.

A. Problem Statement

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ denote the transmission activities of nodes A and B, respectively, *i.e.,*

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

where $S_i(k)$ $(k \geq 1)$ is the kth transmission epoch[1] of A (or B).

If the nodes are transmitting an information flow, then there exist subsequences $\mathbf{F}_i$ $(i = 1,\, 2)$ of $\mathbf{S}_i$, which satisfy the following definition.

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

Remarks: The bijection g is a mapping between the transmission epochs of the same packets at the two nodes, allowing permutations during the relay. The condition that g is a bijection imposes a *packet-conservation* constraint, *i.e.,* every information-carrying packet generates one and only one relay packet. The condition $g(s) - s \geq 0$ is the *causality* constraint, which means that a packet cannot leave a node before it arrives. In addition, $g(s) - s \leq \Delta$ imposes a *bounded delay* constraint, meaning that the maximum delay at the relay node is bounded by Δ. The bounded delay constraint, first proposed by Donoho *et al.* in [4], is often implied by reliable communication protocols.

We say that a pair of processes $(\mathbf{S}_1,\, \mathbf{S}_2)$ *contains an information flow* if $\mathbf{S}_i = \mathbf{F}_i \bigoplus \mathbf{W}_i$ $(i = 1, 2)$, where $(\mathbf{F}_1, \mathbf{F}_2)$ is an information flow. The processes $\mathbf{W}_i$ $(i = 1, 2)$ are called *chaff noise*. Note that chaff noise does not have to satisfy Definition 2.1.

Our problem is to test the following hypotheses:

$$\begin{array}{ll} \mathcal{H}_0 : & \mathbf{S}_1,\, \mathbf{S}_2 \text{ are independent}, \\ \mathcal{H}_1 : & (\mathbf{S}_1,\, \mathbf{S}_2) \text{ contains an information flow}, \end{array} \tag{2}$$

by observing measurements compressed from $\mathbf{S}_i$ $(i = 1,\, 2)$. Assume that Δ is known, and the marginal distributions of $\mathbf{S}_i$ $(i = 1,\, 2)$ are known (Poisson processes are assumed in the analysis) and are the same under both hypotheses[2]. Note that the joint distribution of $(\mathbf{S}_1,\, \mathbf{S}_2)$ under $\mathcal{H}_1$ is not fully specified.

B. System Architecture

We consider a two-stage detection system used in multiterminal inference (*e.g.,* see [7]), as illustrated in Fig. 2. In the data collection stage, encoder $q^{(t)}$ compresses the observation $\mathbf{S}_1$ with duration t into $U \in \{1, \ldots,\, e^{tR}\}$ to satisfy the capacity constraint R. In the detection stage, a detector δ_t makes a decision by[3] $\hat{\theta}_t = \delta_t(U,\, \mathbf{S}_2) \in \{0,\, 1\}$.

[1] Assume no simultaneous transmission.

[2] Otherwise, each eavesdropper can detect information flows individually by testing the marginal distribution.

[3] We use the convention that 0 indicates $\mathcal{H}_0$ and 1 $\mathcal{H}_1$.

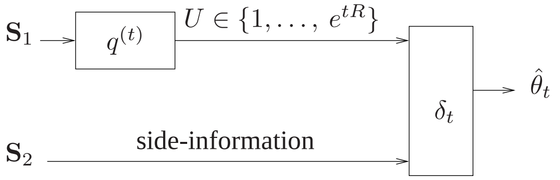

Fig. 2. A distributed detection system which separates the data collection and the decision making stages. $q^{(t)}$: encoder of $\mathbf{S}_1$, δ_t: detector.

C. Performance Measure

Due to the nonparametric model of information flows, the usual performance measure of detection error probabilities cannot be applied directly. Instead, we propose a performance measure based on the amount of chaff noise, measured as follows.

Definition 2.2: Given realizations $(\mathbf{f}_1, \mathbf{f}_2)$ and $(\mathbf{w}_1, \mathbf{w}_2)$ of an information flow and its chaff noise, the *chaff-to-traffic ratio* (CTR) in the interval $[0, t]$ is defined as

$$\mathrm{CTR}(t) \overset{\Delta}{=} \frac{\sum_{i=1}^{2} |\mathcal{W}_i \cap [0, t]|}{\sum_{i=1}^{2} |\mathcal{S}_i \cap [0, t]|}, \qquad \mathrm{CTR} \overset{\Delta}{=} \limsup_{t\to\infty} \mathrm{CTR}(t), \quad (3)$$

where $\mathbf{s}_i = \mathbf{f}_i \bigoplus \mathbf{w}_i$ $(i = 1, 2)$.

Based on the notion of CTR, we borrow the idea of Chernoff-consistency in [10] to introduce the following performance measure.

Definition 2.3: A detector δ_n is *r-consistent* $(r \in [0, 1])$ if it is Chernoff-consistent for all the information flows with CTR bounded by r a.s.[4] That is, the false alarm probability $P_F(\delta_n)$ and miss probability $P_M(\delta_n)$ will vanish if the CTR is bounded by r a.s., *i.e.,*

1) $\lim_{n\to\infty} P_F(\delta_n) = 0$ for any $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_0$;
2) $\sup_{(\mathbf{S}_i)_{i=1}^2 \in \mathcal{P}_0} \lim_{n\to\infty} P_M(\delta_n) = 0$, where

$$\mathcal{P}_0 = \{(\mathbf{S}_i)_{i=1}^2 : (\mathbf{S}_i)_{i=1}^2 \text{ contains an information flow, and } \limsup_{t\to\infty} \mathrm{CTR}(t) \leq r \text{ a.s.}\}.$$

The *consistency* of a detector is the supremum of r such that the detector is r-consistent.

III. Centralized Information Flow Detection

There has been solutions for the case of centralized detection (*i.e.,* $R = \infty$) in [1]. Specifically, it was shown that if $\mathbf{S}_i$ $(i = 1, 2)$ are Poisson processes, then the maximum consistency of a centralized detector is equal to $1/(1 + \lambda\Delta)$, where λ is the maximum rate of $\mathbf{S}_i$. The idea of detection is that detection should be claimed if it is suspiciously easy for an information flow to generate the received measurements. Specifically, the detector estimates the CTR in the measurements and make detection if the estimated CTR is below certain threshold τ. Detection can be guaranteed for flows of rate greater than $\lambda(1 - \tau)$ if the estimated CTR is no larger than the actual CTR.

To calculate the minimum CTR, an optimal chaff-inserting algorithm, "Bounded-Greedy-Match" (BGM) proposed by Blum *et al.* [5] is used. Given realizations $\mathbf{s}_1$, $\mathbf{s}_2$ and delay bound Δ, BGM matches every epoch s in $\mathbf{s}_1$ with the first unmatched epoch in $[s, s+\Delta]$ in $\mathbf{s}_2$. The estimated CTR is the fraction of unmatched epochs among the total measurements. It was shown in [5] that BGM calculates the minimum CTR.

It was shown in [1] that the detector with the maximum consistency has the following form.

Definition 3.1 (Centralized Detector): Given observations $(\mathbf{s}_i)_{i=1}^2$, the detector is defined as

$$\delta_t((\mathbf{s}_i)_{i=1}^2;\, \tau) = \begin{cases} 1 & \text{if } \widehat{\mathrm{CTR}}(t) \leq \tau, \\ 0 & \text{o.w.,} \end{cases}$$

where $\tau \in [0, 1]$ is a predetermined threshold, and $\widehat{\mathrm{CTR}}(t)$ is the CTR in $(\mathbf{s}_i)_{i=1}^2$ estimated by BGM.

It is shown in [1] that there is a single threshold C^* on the CTR that separates detectable and undetectable flows. Actually, this statement holds for general renewal processes, as stated below.

Theorem 3.2: If $\mathbf{S}_i$ $(i = 1, 2)$ are renewal processes, then there exists a constant C^* such that flows with CTR $\geq C^*$ can be embedded in independent processes and are thus undetectable, and all the flows with CTR $< C^*$ can be detected consistently by δ_t.

The proof is omitted due to space limit. The goal of this paper is to characterize C^* for finite capacity constraints.

IV. Quantizer Design

In this section, we present two simple quantizers based on the counting measure.

Definition 4.1: Given a point process $\mathbf{S}$, a *slotted quantizer*[5] with slot length T is defined as $\gamma(\mathbf{S}) \overset{\Delta}{=} (Z_j)_{j=1}^{\infty}$, where Z_j is the number of epochs 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 pointwise 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 4.2: Given a point process $\mathbf{S}$, a *one-bit quantizer* with slot length T is defined as $\hat{\gamma}(\mathbf{S}) \overset{\Delta}{=} (Z_j)_{j=1}^{\infty}$, where Z_j is the indicator that the jth slot is nonempty.

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

Let $\mathbf{X}^n = (X_j)_{j=1}^n$ be the quantization result of $\mathbf{S}_1$. Moreover, use $Y(s, t)$ to denote the number of epochs in $\mathbf{S}_2$ in the interval $[s, t)$. The capacity constraint R can be expressed as

$$\limsup_{n\to\infty} \frac{H(\mathbf{X}^n)}{nT} \leq R, \quad (4)$$

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

[5] The same quantizer was used in [9] and was shown to be near optimal under the absolute-error fidelity critetion.

which, if $\mathbf{S}_1$ is a Poisson process, is reduced to $H(X_1)/T \leq R$. For ergodic source satisfying (4), $\mathbf{X}^n$ (for large n) can be reliably transmitted to the detector. Thus, the problem is reduced to design detectors for observations $\mathbf{X}^n$ and $\mathbf{S}_2$.

V. Detector Design

Following the idea of centralized detection, we develop threshold detectors based on the optimal chaff-inserting algorithms for quantized measurements as follows.

A. Detection Under Slotted Quantization

Given measurements $(\mathbf{x}^n, \mathbf{s}_2)$, where $\mathbf{x}^n$ is the result of slotted quantization of $\mathbf{s}_1$, we want to find realizations of an information flow $(\mathbf{f}_i)_{i=1}^2$ and chaff noise $\mathbf{w}_i$ ($i = 1, 2$) such that i) $\mathbf{x}^n = \gamma(\mathbf{f}_1 \bigoplus \mathbf{w}_1)$, $\mathbf{s}_2 = \mathbf{f}_2 \bigoplus \mathbf{w}_2$, and ii) the CTR is minimized. Now that $\mathbf{s}_1$ is unavailable, we first reconstruct $\mathbf{s}_1$ from $\mathbf{x}^n$ and then apply BGM to the reconstructed processes, as illustrated in Fig. 3. The proposed chaff-inserting algorithm, "Slotted-Greedy-Match" (SGM), works as follows. Given $(\mathbf{x}^n, \mathbf{s}_2)$, SGM does the following:

1) construct a batched process $\hat{\mathbf{s}}_1$ as busts of x_j simultaneous epochs at $(j-1)T$ ($j \geq 1$), as illustrated in Fig. 3;
2) run BGM on $(\hat{\mathbf{s}}_1, \mathbf{s}_2)$ with delay bound $T + \Delta$.

It is shown in [11] that SGM inserts the minimum number of chaff packets for any $(\mathbf{x}^n, \mathbf{s}_2)$.

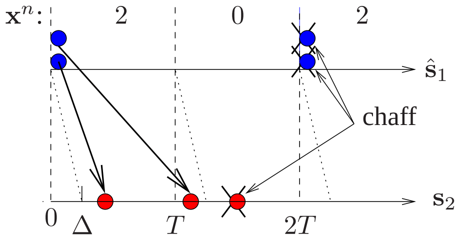

Fig. 3. SGM: greedy matching between a batched process $\hat{\mathbf{s}}_1$ and $\mathbf{s}_2$. Epochs not matched within delay $T + \Delta$ are considered as chaff noise.

Based on SGM, we develop the following detector.

Definition 5.1 (Detector under Slotted Quantization): Given $(\mathbf{x}^n, \mathbf{s}_2)$, define a detector δ_{I} as

$$\delta_{\mathrm{I}}(\mathbf{x}^n, \mathbf{s}_2) = \begin{cases} 1 & \text{if } C_{\mathrm{I}}/N \leq \tau_{\mathrm{I}}, \\ 0 & \text{o.w.}, \end{cases}$$

where $N = \sum_{j=1}^{n} x_j + |\mathcal{S}_2|$, and C_{I} is the number of chaff packets found by SGM in $(\mathbf{x}^n, \mathbf{s}_2)$, excluding chaff packets in[6] $\mathcal{S}_2 \cap [0, \Delta)$.

Intuitively, under $\mathcal{H}_1$, since the actual number of chaff packets has to be at least C_{I}, δ_{I} has vanishing miss probability as long as the CTR is bounded by τ_{I} a.s. Under $\mathcal{H}_0$, δ_{I} will have vanishing false alarm probability if τ_{I} is sufficiently small. The performance of δ_{I} is guaranteed by the following theorem.

Theorem 5.2: If $\mathbf{S}_1$ and $\mathbf{S}_2$ are Poisson processes of rates bounded by λ, and T is large, then δ_{I} can achieve the consistency level $\alpha_{\mathrm{I}}(R_{\mathrm{I}}^{-1}(R))$ under a capacity constraint R, where[7]

$$\alpha_{\mathrm{I}}(T) \overset{\Delta}{=} \frac{1}{\sqrt{\pi \lambda T}} - \frac{\Delta}{4T}, \qquad R_{\mathrm{I}}(T) \overset{\Delta}{=} H(\mathrm{Poi}(\lambda T))/T. \quad (5)$$

Furthermore, for any $\tau_{\mathrm{I}} < \alpha_{\mathrm{I}}(T)$, the false alarm probability decays exponentially with n.

Proof: See the proof of Theorem 5.2 in [11]. ■

Given a capacity R, the function $\alpha_{\mathrm{I}}(R_{\mathrm{I}}^{-1}(R))$ specifies the level of chaff noise under which δ_{I} can achieve consistent detection while satisfying the capacity constraint. This function, referred to as the *consistency-rate function*, is therefore a measure of the performance of distributed flow detection.

B. Detection Under One-Bit Quantization

Under one-bit quantization, the chaff-inserting algorithm also utilizes BGM as in SGM except that now the reconstruction of $\mathbf{s}_1$ is more difficult since the exact number of epochs in a nonempty slot is unknown. This issue can be solved by starting matching from $\mathbf{s}_2$. As illustrated in Fig. 4, if we sequentially match each epoch in $\mathbf{s}_2$ with the first nonempty slot in $\mathbf{s}_1$ that has not been matched, then due to the fact that greedy matching is optimal, one can easily show the optimality of this scheme. Specifically, the proposed algorithm, called "One-bit-Greedy-Match" (OGM), works as following. Given $(\mathbf{x}^n, \mathbf{s}_2)$, OGM:

1) match every epoch in $\mathbf{s}_2$ with the earliest unmatched nonempty slot within delay Δ, as illustrated in Fig. 4;
2) unmatched epochs become chaff; each unmatched nonempty slot contains a chaff packet.

Algorithm OGM inserts the minimum chaff noise, as shown in [11].

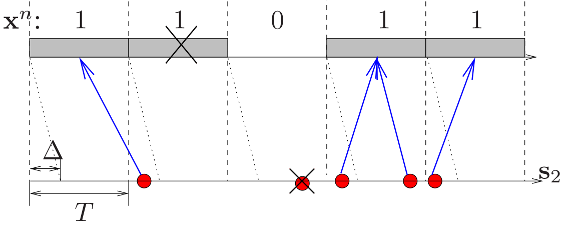

Fig. 4. OGM: Greedy matching starting from $\mathbf{s}_2$. Each epoch in $\mathbf{s}_2$ is matched to the first unmatched nonempty slot (*i.e.*, $x_j = 1$) that is no more than Δ earlier.

The following detector is developed based on OGM.

Definition 5.3 (Detector under One-bit Quantization): Given $(\mathbf{x}^n, \mathbf{s}_2)$, the detector δ_{II} is defined as

$$\delta_{\mathrm{II}}(\mathbf{x}^n, \mathbf{s}_2) = \begin{cases} 1 & \text{if } C_{\mathrm{II}}/(n\hat{N}_1 + |\mathcal{S}_2|) \leq \tau_{\mathrm{II}}, \\ 0 & \text{o.w.}, \end{cases}$$

where C_{II} is the number of chaff packets found by OGM in $(\mathbf{x}^n, \mathbf{s}_2)$, excluding chaff packets in $\mathcal{S}_2 \cap [0, \Delta)$, and $\hat{N}_1 = -\log(1 - \bar{x})$ for $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$. Here $\hat{N}_1$ is the Maximum

[6] This is because packets in this interval may be relays of packets transmitted before the detector starts taking observations.

[7] Here $H(\mathrm{Poi}(\lambda T))$ is the entropy of Poisson distribution with mean λT.

Likelihood estimate of the mean number of epochs per slot in $\mathbf{S}_1$ if $\mathbf{S}_1$ is Poisson.

The structure of δ_{II} is very similar to that of δ_{I}. Not surprisingly, the performance of δ_{II} can also be characterized by a theorem similar to Theorem 5.2.

Theorem 5.4: If $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent Poisson processes of maximum rate λ, and T is large, then δ_{II} can achieve the consistency-rate function $\alpha_{\text{II}}(R_{\text{II}}^{-1}(R))$, where[8]

$$\alpha_{\text{II}}(T) \overset{\Delta}{=} \frac{1}{2}e^{-\lambda T}, \quad R_{\text{II}}(T) \overset{\Delta}{=} \begin{cases} \log 2/T & \text{if } \lambda T \geq \log 2, \\ h(e^{-\lambda T})/T & \text{o.w.,} \end{cases} , \tag{6}$$

and the false alarm probability decays exponentially with n if $\tau_{\text{II}} < \alpha_{\text{II}}(T)$.

Proof: See the proof of Theorem 5.6 in [11]. ■

Comparing Theorems 5.2 and 5.4, we see that as T increases, $\alpha_{\text{II}}(T)$ decays much faster than $\alpha_{\text{I}}(T)$, reflecting the decay of detection performance caused by further quantization. It, however, does not imply that slotted quantization is better because one-bit quantization allows a much smaller slot length under the same capacity.

VI. Numerical Comparison

We compare the proposed detection systems by their consistency-rate functions. It was shown in [11] that $\alpha_i(T)$ (i = I, II) derived in Theorems 5.2 and 5.4 are loose lower bounds on the maximum consistency of δ_i. The actual maximum consistency is equal to a constant $\alpha_i^*(T)$, which is the a.s. limit of the minimum CTR of SGM or OGM under $\mathcal{H}_0$. Furthermore, it was shown in [11] that $u(T) \overset{\Delta}{=} \mathbb{E}[|X - Y|]/(2\lambda T)$, where X and Y are independent Poisson random variables with mean λT, is an upper bound on the consistency of any detector under slotted or one-bit quantization.

We plot the consistency-rate functions $\alpha_i^*(R_i^{-1}(R))$ (i = I, II) together with their upper bounds[9] $u(R_i^{-1}(R))$ under various traffic rates; see Fig. 5. Here $\alpha_i^*(T)$ is computed by simulating SGM or OGM on independent Poisson processes of rate λ. From these plots, we observe: i) slotted quantization outperforms one-bit quantization, and their difference increases with λ; ii) at the same R, the consistency decreases with the increase of λ; iii) the consistency of δ_{I} is close to the upper bound for small R.

Observation (i) implies that besides being a good method for compressing Poisson processes ([9]), slotted quantization also performs well in compressing traffic measurements for detecting flows. Observation (ii) implies that information flows in heavy traffic are more difficult to detect than those in light traffic. This is because as λ increases, the maximum delay normalized by the mean interarrival time (*i.e.*, $\lambda\Delta$) will increase, making the delay constraint relatively loose. Observation (iii) implies that the detector δ_{I} is near optimal under slotted quantization for sufficiently small capacities.

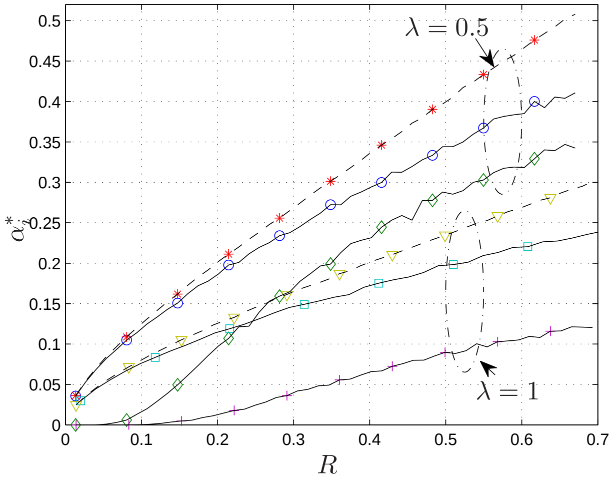

Fig. 5. The consistency-rate functions $\alpha_i^*(R_i^{-1}(R))$ (i = I, II) and their upper bounds ($\Delta = 1$, simulated over 10^4 slots.): For each λ, the three curves in decreasing order are the upper bound, the consistency-rate function under slotted quantization, and the consistency-rate function under one-bit quantization.

VII. Conclusion

This paper has focused on detection in the presence of side-information. Parallel results have also been obtained without side-information [8]. It can be shown that side-information has dominant effect on the detection performance for sufficiently light traffic, whereas quantization scheme matters more for heavier traffic.

References

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

[2] 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.

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

[4] 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.

[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] 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.

[7] R. Ahlswede and I. Csiszar, “Hypothesis testing with communication constraints,” *Information Theory, IEEE Transactions on*, vol. 32, no. 4, pp. 533–542, 1986.

[8] 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.

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

[10] J. Shao, *Mathematical Statistics*. Springer, 1999.

[11] T. He and L. Tong, “Distributed Detection of Information Flows.” submitted to IEEE Trans. on Information Forensics and Security, 2007.

[8] Here $h(p)$ is the binary entropy function defined as $h(p) = -p\log p - (1-p)\log(1-p)$.

[9] The upper bound $u(R_{\text{II}}^{-1}(R))$ is much looser than $u(R_{\text{I}}^{-1}(R))$, and thus only the latter is plotted.