

Distributed Detection of Information Flows

Ting He, *Member, IEEE*, and Lang Tong, *Fellow, IEEE*

Abstract—Distributed detection of information flows is considered in which traffic sensors at different locations of a network observe transmission epochs. The traffic sensors communicate their measurements to a fusion center via channels with rate constraints, and the fusion center performs hypothesis testing for information flow detection. Under a nonparametric flow model where relayed packets can be perturbed up to bounded delays and multiplexed with chaff noise, flow detectability is characterized through a notion called consistency-rate function that shows the level of detectable flows under capacity constraints on the fusion channels. Achievability results are presented by constructing detection systems consisting of quantization, data transmission, and detection subsystems. In particular, slot-by-slot quantization schemes at the local sensors and threshold detection schemes at the fusion center are proposed to provide consistent detection with quantifiable performance.

Index Terms—Distributed detection, information flow, information-theoretic limits, traffic analysis.

I. INTRODUCTION

WE investigate distributed detection of information flows in a wireless network where, as illustrated in Fig. 1, traffic sensors are deployed to monitor transmission activities of wireless nodes. We assume that traffic sensors are simple devices that are unable to decode the transmissions of monitored nodes; they merely record the transmission epochs and communicate their measurements to a fusion center over bandwidth-constrained channels. The constraints on fusion capacity dictates the use of local quantization at the individual sensors, and the fusion center has to detect the presence of information flows using quantized measurements.

The problem of detecting information flows arises from applications in information forensics and network security. For example, in the so-called stepping-stone attack [1], an intruder compromises a number of nodes in a network and uses these nodes to relay attacking commands. The detection and tracing of such unauthorized relays is an important yet challenging task

Manuscript received October 11, 2007; revised June 13, 2008. Published August 13, 2008 (projected). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Muriel Medard. This work was supported in part by the National Science Foundation under Award CCF-0635070 and in part by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program DAAD19-01-2-0011. Much of this work was completed when the first author was with Cornell University. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon. Part of this work was published in IEEE ISIT 2007 and IEEE Asilomar 2007.

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

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ece.cornell.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2008.928537

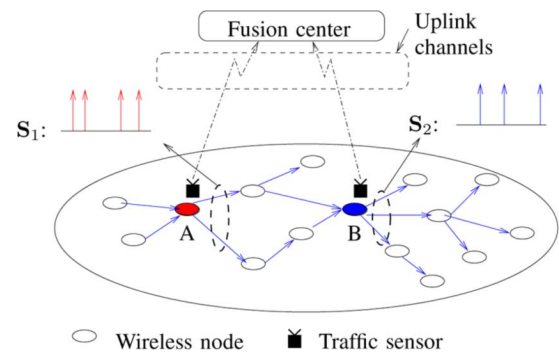


Fig. 1. In a wireless network, traffic sensors are deployed to collect transmission activities of nodes A and B , denoted by S_i ($i = 1, 2$), which are then sent to a detector at the fusion center to detect information flows through A and B .

because attackers can hide their activities by adding random delays, inserting dummy packets, reshuffling transmissions, etc. Also related is the surveillance of wireless transmissions using distributed wireless sensors. In this case, the detection problem is even more difficult because different information flows multiplex at the intersecting nodes, causing the measurements to contain packets belonging to multiple flows. For the flow of interest, all of the measurements that are not part of this flow, including multiplexed packets, dummy packets, dropped packets, etc., act as noise and are therefore referred to as chaff noise.

A. Related Work

Information flow detection is a special case of timing analysis which, in turn, belongs to the family of traffic analysis [2]. The problem has been studied as a countermeasure to stepping-stone attacks in the context of the Internet, where the key task is to reconstruct the intrusion path by analyzing link traffic. Since first studied by Staniford and Heberlein [1], the problem has evolved to allow various manipulations by the attacker, including encryption, padding, perturbations, and insertion of chaff noise. Specifically, Zhang and Paxson [3] first considered timing-based techniques to deal with encrypted traffic. Donoho *et al.* [4] were the first to consider active perturbations. They showed that if the perturbations have a bounded delay, then there will be a distinguishable difference between information flows and independent traffic. In [5], we proposed a parallel perturbation model under the assumption of bounded memory at relay nodes. Recent work shows that reliable detection can be achieved even in the presence of maliciously injected chaff noise [6]–[8]. These detectors, however, can only handle a limited number of chaff packets. The only timing-based detector so far that provides vanishing error probabilities even if the amount of chaff noise grows proportionally to the total traffic size was proposed in [9] and extended in [10]. As a predecessor of the current paper, the approach in [10] is centralized, where there are no capacity constraints on the fusion channels.

The problem of distributed detection of information flows, to our best knowledge, has not been reported in the literature, and such investigation is most relevant for wireless sensors that have to deliver the measured traffic information to the fusion center in some economic form. While the problem of distributed detection has been studied extensively, the standard model considered in classical distributed detection does not apply to information flow detection. Specifically, existing work on distributed inference under capacity constraints (e.g., [11]) typically assumes that samples obtained at different sensors, although maybe correlated spatially, are i.i.d. in time. In particular, Ahlswede and Csiszár [12] studied the problem testing against independence, which bears considerable similarity to the problem of detecting information flows formulated in Section II. The lack of temporal independence in information flows, however, fundamentally changes the problem, and existing techniques in [11] do not apply.

B. Summary of Results, Limitations, and Organization

In this paper, we formulate the problem of information flow detection as one of partially nonparametric binary hypothesis testing. Under the null hypothesis \mathcal{H}_0 that models the normal traffic, we assume that epochs measured by the local sensors are statistically independent, whereas under the alternative \mathcal{H}_1 , a flow of packets exists relayed by the monitored nodes. The presence of information flow imposes certain constraints on the transmission epochs. Specifically, packets that are part of the flow are relayed causally and within a certain delay constraint.

We first address the fundamental issue of detectability: given that an information flow is considered detectable if it has a Chernoff-consistent detector [13] (i.e., a detector with both false alarm and miss probabilities going to zero as the observation length increases), are all the flows detectable? In other words, are there cases in which with none-zero probability, the observations correspond to distributions under both hypotheses? Indeed, it is not hard to see that information flows of very low rate may not be detectable. For example, given a pair of independent processes of transmission epochs, one can always find some epochs that can be used to schedule relay transmissions, and such relay transmissions are not detectable. On the other hand, it is generally impossible to match all of the epochs of two independent processes in such a way that the matching is causal and bounded in delay. Therefore, we expect that there is some rate of information flow below which detection is impossible and some (possibly different) rate above which it is possible.

In the context of the distributed detection of information flows, our goal is to characterize the so-called consistency-rate function. Specifically, we are interested in that given a set of capacity constraints on the fusion channels, to what level the detection can be asymptotically accurate as the observation length increases. To this end, we present some achievability results. Our solution consists of four different scenarios. In particular, we consider cases when the fusion center is colocated with one of the local sensors, which is referred to as distributed detection with full side-information. In this case, measurements at the fusion center do not need to be quantized. Alternatively, the fusion center is not colocated with any sensor, and all of the sensors must quantize their measurements.

For local sensors, we propose several slot-by-slot quantization schemes. Specifically, we divide the observation window into slots and let sensors transmit the number of epochs in each slot to the fusion center. Sufficiently large slot lengths are chosen to satisfy the rate constraints. Alternatively, we also propose a slotted one-bit quantization scheme where only the presence or the absence of transmissions in each slot is reported.

For the fusion center, we propose, for each scenario, a detector that provides Chernoff-consistent detection for all of the flows with chaff noise bounded by a certain level. Our schemes are generalizations of the centralized detection scheme proposed in [10], where one assumes that the fusion rates are unconstrained.

A few words on the limitations of our approach are in order. We note here that the separated design of quantization and detection is, in general, suboptimal. Thus, the results presented in this paper are mostly achievability results. Our model is partially nonparametric, and our approach does not depend on the specific traffic distributions. To obtain analytical characterization of the consistency-rate function, we have assumed that the observed processes by individual sensors are Poisson under both hypotheses. It is well known that Poisson modeling is not accurate for many applications. While acknowledging the limitation of the Poisson assumption, we note, however, that the centralized algorithms developed under this assumption have worked reasonably well for actual Internet traces [14]. Finally, our results are presented for the case of two sensors. The approach has a natural generalization to the multisensor case, preliminary results of which have been presented in [15].

The rest of this paper is organized as follows. Section II formulates the problem. Section III gives the performance criteria and a general converse result. Sections IV–VI are dedicated to the design of specific detection systems, among which Section IV defines two slot-based quantizers, Section V presents the corresponding detectors, and Section VI analyzes and compares the performance of the proposed systems. Then, Section VII concludes with remarks.

II. PROBLEM FORMULATION

A. Notation

We use the convention that uppercase letters denote random variables, lowercase letters denote realizations, boldface letters denote vectors, and plain letters denote scalars. For example, we denote a point process by \mathbf{S} , its realization \mathbf{s} , the k th epoch $S(k)$, and realization of the k th epoch $s(k)$. Given a realization \mathbf{s} 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, \dots) and (b_1, b_2, \dots) , \oplus is the superposition operator defined as $(a_k)_{k=1}^{\infty} \oplus (b_k)_{k=1}^{\infty} = (c_k)_{k=1}^{\infty}$, where $c_1 \leq c_2 \leq \dots$ and $\{a_k\}_{k=1}^{\infty} \cup \{b_k\}_{k=1}^{\infty} = \{c_k\}_{k=1}^{\infty}$.

B. Flow Model and Hypotheses

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

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

where $S_i(k)$ ($k \geq 1$) is the k th transmission epoch¹ of A (or B).

¹Assume no simultaneous transmission.

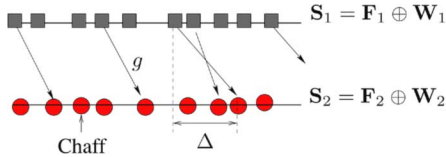


Fig. 2. Information flow $(\mathbf{F}_1, \mathbf{F}_2)$ in chaff noise $\mathbf{W}_i (i = 1, 2)$; g : the causal bijection between \mathbf{F}_1 and \mathbf{F}_2 with delays bounded by Δ .

We first introduce the definition of information flow.

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

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.

Next, we say that $(\mathbf{S}_1, \mathbf{S}_2)$ contains an information flow if, as illustrated in Fig. 2, $\mathbf{S}_i (i = 1, 2)$ can be partitioned into a subsequence \mathbf{F}_i of information-carrying packets and a subsequence \mathbf{W}_i of the so-called chaff (noise) packets, i.e.,

$$\mathbf{S}_i = \mathbf{F}_i \oplus \mathbf{W}_i, \quad i = 1, 2 \quad (2)$$

where $(\mathbf{F}_1, \mathbf{F}_2)$ is an information flow. Here, the chaff processes $\mathbf{W}_i (i = 1, 2)$ model all of the transmissions that are not part of the information flow. They do not need to satisfy any constraints and can be correlated with the flow.

We are interested in testing the following hypotheses at the fusion center:

$$\begin{aligned} \mathcal{H}_0 : \mathbf{S}_1, \mathbf{S}_2 \text{ are independent} \\ \mathcal{H}_1 : (\mathbf{S}_1, \mathbf{S}_2) \text{ contains an information flow} \end{aligned} \quad (3)$$

for known delay constraint Δ by observing compressed measurements sent by local sensors.

We assume that the marginal distributions of $\mathbf{S}_i (i = 1, 2)$ are known, and they are the same under both hypotheses (detailed analysis is done for Poisson processes). Otherwise, a sensor can independently make a decision based on its own measurements (e.g., by the Anderson–Darling test [16]) and send the result (a 1-b message) to the fusion center, and the error probabilities can be made arbitrarily small if there are enough measurements. This is partially nonparametric hypothesis testing² because no statistical assumptions are imposed on the correlation of \mathbf{S}_1 and \mathbf{S}_2 under \mathcal{H}_1 .

C. Architecture of Distributed Detection Systems

The capacity constraints in the uplink channels make it necessary to employ quantizers $q_i^{(t)}$ ($i = 1, 2$) at the sensors, where t is the duration of the observation. As illustrated in Fig. 3, the

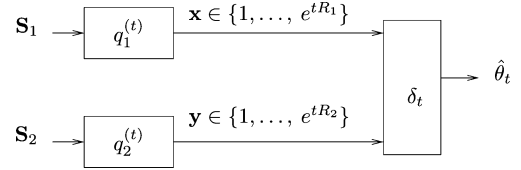


Fig. 3. Distributed detection system. This system consists of two quantizers $q_1^{(t)}$ and $q_2^{(t)}$ and a detector δ_t .

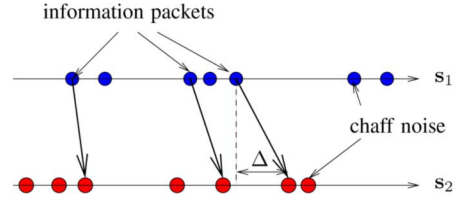


Fig. 4. Information flow hidden in independent processes: s_1, s_2 : realizations of independent processes; each matched pair denotes the transmission epochs of a packet in the flow.

processes $\mathbf{S}_i (i = 1, 2)$ are compressed into $q_i^{(t)}(\mathbf{S}_i)$ which are delivered to the fusion center, and then the detector makes a decision in the form of

$$\hat{\theta}_t = \delta_t \left(q_1^{(t)}(\mathbf{S}_1), q_2^{(t)}(\mathbf{S}_2) \right)$$

where³ $\hat{\theta}_t \in \{0, 1\}$. The capacity constraints are expressed as⁴

$$\|q_i^{(t)}\| \leq e^{tR_i}, \quad i = 1, 2 \quad (4)$$

for sufficiently large t , where $\|q_i^{(t)}\|$ is the alphabet size of the output of $q_i^{(t)}$. Generally, $R_1, R_2 < \infty$, but if the detector is located at one of sensors (e.g., the sensor at node B), then $R_2 = \infty$, which is the so-called case of full side information.

III. DETECTABILITY OF INFORMATION FLOWS

Since chaff noise can be arbitrarily correlated with information flows, as mentioned earlier, not all flows are detectable. Fig. 4 illustrates realizations of independent point processes under \mathcal{H}_0 within which some epochs are used to relay packets of an information flow.

In this section, we formalize this intuition into some quantifiable measure of the detectability of information flows.

A. Performance Measure

Intuitively, if the rate of information flow is low, or equivalent, the rate for chaff noise is high, detection is not possible. Thus, the flow detectability is tied intimately to the notion of chaff-to-traffic ratio (CTR) that is defined, which is reminiscent of the more familiar term signal-to-noise ratio.

Definition 3.1: Given realizations $(\mathbf{f}_1, \mathbf{f}_2)$ and $(\mathbf{w}_1, \mathbf{w}_2)$ of an information flow and its chaff noise, the CTR of (s_1, s_2) , where $s_i = \mathbf{f}_i \oplus \mathbf{w}_i (i = 1, 2)$ is defined as

$$\text{CTR} \triangleq \limsup_{t \rightarrow \infty} \frac{\sum_{i=1}^2 |\mathfrak{W}_i \cap [0, t]|}{\sum_{i=1}^2 |\mathfrak{S}_i \cap [0, t]|}. \quad (5)$$

\square

²Note that this notion is different from seminonparametric tests [17], in which an unknown distribution function is approximated by a series of known functions.

³The value 0 denotes \mathcal{H}_0 , and 1 denotes \mathcal{H}_1 .

⁴The unit of $R_i (i = 1, 2)$ is nats per unit time.

As the name suggests, CTR ($0 \leq \text{CTR} \leq 1$) is the asymptotic fraction of chaff noise in traffic that contains information flow. We will use CTR to measure the amount of chaff noise that a detector can handle. Specifically, we use Chernoff consistency [13] to define the consistency level of a detector as follows.

Definition 3.2: A detector δ_t is r -consistent ($r \in [0, 1]$) if it is Chernoff consistent for all of the information flows with CTR bounded by r a.s.⁵, that is, the false alarm probability $P_F(\delta_t)$ and the misprobability $P_M(\delta_t)$ satisfy

- 1) $\lim_{t \rightarrow \infty} P_F(\delta_t) = 0$ for any $(\mathbf{S}_i)_{i=1}^2$ under \mathcal{H}_0 ;
- 2) $\sup_{(\mathbf{S}_i)_{i=1}^2 \in \mathcal{P}_0} \lim_{t \rightarrow \infty} P_M(\delta_t) = 0$, where $\mathcal{P}_0 = \{(\mathbf{S}_i)_{i=1}^2 : (\mathbf{S}_i)_{i=1}^2 \text{ contains an information flow with CTR} \leq r \text{ a.s.}\}$.

The consistency level of δ_t is the supremum of r such that δ_t is r -consistent. \square

Intuitively, the consistency level of a detector is the maximum fraction of chaff noise so that both false alarm and miss probabilities vanish with the observation length. Note that since the distribution of traffic is unknown when there is a flow (\mathcal{H}_1), one cannot explicitly characterize the misprobability. The consistency level, however, provides a unified way to measure how much chaff noise a detector can tolerate, regardless of the distributions of the flow and the noise. In the sequel, we will use the consistency level to evaluate detection performance, where a higher consistency level means better robustness against chaff noise.

B. Achievability and Converse on Flow Detectability

Using consistency level as the performance measure, we can formalize the detectability of information flows in terms of achievable consistency levels and the converse. Moreover, since the overall performance of distributed detection relies on the capacities of data fusion, our characterization needs to be functions of these capacities. Specifically, we characterize the optimal achievable performance of information flow detection by the highest consistency level as a function of the capacity constraints as follows.

Definition 3.3: Given capacity constraints (R_1, R_2) , we define the consistency-rate function, denoted by $\alpha(R_1, R_2)$, as

$$\alpha(R_1, R_2) \triangleq \sup\{r \in [0, 1] : \exists (q_1^{(t)}, q_2^{(t)}, \delta_t) : \begin{aligned} &1) \delta_t \text{ is } r\text{-consistent;} \\ &2) \limsup_{t \rightarrow \infty} \frac{1}{t} \log \|q_i^{(t)}\| \leq R_i, \quad i = 1, 2. \end{aligned}\} \quad (6)$$

\square

The consistency-rate function maps capacity constraints to the maximum achievable consistency level among all of the detection systems satisfying the constraints, providing the optimal tradeoff between communication cost and detection performance. Similarly, if the aforementioned optimization is only among a certain family of quantizers or detection systems, then the result is called the consistency-rate function of this type of quantizers or detection system, which can then be used to evaluate the system performance.

⁵Here ‘‘a.s.’’ means almost surely.

We now consider the converse. We have seen from Fig. 4 that it is not always possible to detect small flows. In particular, the minimum CTR (Definition 3.1) required to mimic \mathcal{H}_0 in distribution establishes a threshold on the noise level beyond which the flows can be made undetectable, as stated in the following definition.

Definition 3.4: Given capacity constraints (R_1, R_2) , the undetectability-rate function is defined as

$$\phi(R_1, R_2) \triangleq \max_{\mathcal{P}_1} \phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y}) \quad (7)$$

where (\mathbf{X}, \mathbf{Y}) are auxiliary random variables from a set \mathcal{P}_1 defined by⁶

$$\mathcal{P}_1 = \{(\mathbf{X}, \mathbf{Y}) : \mathbf{X} \rightarrow \mathbf{S}_1 \rightarrow \mathbf{S}_2 \rightarrow \mathbf{Y}, \limsup_{t \rightarrow \infty} \frac{1}{t} I(\mathbf{S}_1; \mathbf{X}) \leq R_1, \limsup_{t \rightarrow \infty} \frac{1}{t} I(\mathbf{S}_2; \mathbf{Y}) \leq R_2\},$$

and⁷

$$\begin{aligned} &\phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y}) \\ &\triangleq \inf\{r \in [0, 1] : \\ &\quad \exists(\mathbf{F}_i, \mathbf{W}_i)_{i=1}^2 \text{ and some } (\mathbf{S}_i)_{i=1}^2 \text{ under } \mathcal{H}_0 : \\ &\quad 1) \mathbf{F}_i \oplus \mathbf{W}_i \stackrel{d}{=} \mathbf{S}_i (i = 1, 2) \text{ and } (\mathbf{X}, \mathbf{Y})|_{(\mathbf{F}_i \oplus \mathbf{W}_i)_{i=1}^2} \\ &\quad \stackrel{d}{=} (\mathbf{X}, \mathbf{Y})|_{(\mathbf{S}_i)_{i=1}^2}; \\ &\quad 2) (\mathbf{F}_1, \mathbf{F}_2) \text{ is an information flow;} \\ &\quad 3) \text{CTR} \leq r \text{ a.s.}\} \end{aligned} \quad (8)$$

In (8), $(\mathbf{X}, \mathbf{Y})|_{(\mathbf{S}_i)_{i=1}^2}$ is the marginal of (\mathbf{X}, \mathbf{Y}) in $(\mathbf{X}, \mathbf{Y}, \mathbf{S}_1, \mathbf{S}_2)$ specified by the distribution of $(\mathbf{S}_1, \mathbf{S}_2)$ and the conditional distribution⁸ $Q_1(\mathbf{X}|\mathbf{S}_1)Q_2(\mathbf{Y}|\mathbf{S}_2)$. \square

The undetectability-rate function maps capacity constraints to the minimum level of chaff noise required to make an information flow mimic \mathcal{H}_0 in distribution under the optimal quantization. Here, the conditional distributions $Q_1(\mathbf{X}|\mathbf{S}_1), Q_2(\mathbf{Y}|\mathbf{S}_2)$ specify the quantizers of \mathbf{S}_1 and \mathbf{S}_2 , and (8) calculates the minimum noise level [condition (3)] for an information flow (condition (2)) to appear identical with some traffic under \mathcal{H}_0 in both marginal distributions before quantization and the joint distribution after quantization (condition (1)). The overall level of undetectability is then defined as the maximum of such a noise level over all quantizers satisfying given capacity constraints. Contrary to the consistency-rate function, the undetectability-rate function provides a converse on flow detectability, where for noise levels above this function, flows can always be camouflaged as traffic under \mathcal{H}_0 by perturbations and chaff insertion. It is easy to see the following relationship between these functions.

Theorem 3.5: For any $R_i \geq 0 (i = 1, 2)$, $\alpha(R_1, R_2) \leq \phi(R_1, R_2)$.

⁶Note that \mathcal{P}_1 is well defined because $\mathbf{S}_i (i = 1, 2)$ have the same distributions under both hypotheses.

⁷The notation ‘‘ $\stackrel{d}{=}$ ’’ means equal in distribution.

⁸We can write the conditional distribution $Q(\mathbf{X}, \mathbf{Y}|\mathbf{S}_1, \mathbf{S}_2)$ in product form because the quantization of the two processes is independent.

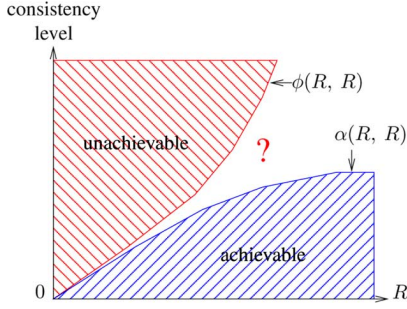


Fig. 5. Achievable and unachievable regions of distributed information flow detection in the consistency-rate space ($R_1 = R_2 = R$). By time-sharing arguments, it can be shown that $\alpha(R_1, R_2)$ is concave whereas $\phi(R_1, R_2)$ is convex with respect to (R_1, R_2) . The plot is not from computation and is only for illustration purposes.

Proof: With (8), an information flow exists ($\mathbf{F}_1, \mathbf{F}_2$) and chaff noise ($\mathbf{W}_1, \mathbf{W}_2$) constituting $\phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y})$ fraction of the total traffic such that the mixture $\mathbf{F}_i \oplus \mathbf{W}_i$ ($i = 1, 2$) appears statistically identical with traffic under \mathcal{H}_0 to the sensors and the detector. No detector can detect this flow consistently and, thus, the consistency level under these quantizers is bounded by $\phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y})$. Therefore, maximizing $\phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y})$ over all of the quantizers under (R_1, R_2) yields an upper bound on the overall highest consistency level. ■

Through the notions in Definition 3.3 and 3.4, we have introduced a partition of the consistency-rate space: as illustrated in Fig. 5, the region below $\alpha(R_1, R_2)$ is achievable in that any flow with noise level (CTR) below $\alpha(R_1, R_2)$ can be detected consistently under capacity constraints (R_1, R_2) ; the region above $\phi(R_1, R_2)$ is unachievable in that with the noise level above $\phi(R_1, R_2)$, a flow can appear statistically identical with normal traffic under any quantization that satisfies the corresponding capacity constraints and is thus completely undetectable. It has been proven in [10] that for centralized detection (i.e., $R_1 = R_2 = \infty$) and Poisson processes, the two functions are equal

$$\alpha(\infty, \infty) = \phi(\infty, \infty) = \frac{1}{1 + \lambda\Delta}$$

where λ is the rate of \mathbf{S}_i ($i = 1, 2$). Under finite capacities, the results are unknown.

A few remarks are in order at the conclusion of this section:

- The undetectability-rate function is conceptually similar to the distortion-rate function of point processes in the following sense: under rate constraints, the distortion-rate function gives the minimum distortion between the original and the reconstructed processes, whereas the undetectability-rate function calculates the maximum “deviation” between the two hypotheses in terms of the amount of chaff noise required to make them equal.
- While achievability results can be obtained by designing specific detection systems, the converse is more challenging because the optimal quantizers are unknown. For Poisson traffic, however, we have $\alpha(R_1, R_2) \leq 1/(1 + \lambda\Delta)$ since $\phi(\cdot)$ is monotone increasing.

In the sequel, we will focus on achievability results by constructing explicit detection systems.

IV. QUANTIZERS

The design of quantizers $q_i^{(t)}$ ($i = 1, 2$) is complicated by the dependency on t . To simplify the design, we partition the observation into n slots of equal length T ($T = t/n$) and use independent and identical quantization in each slot. We consider the following quantizers.

Definition 4.1: Given a point process \mathbf{S} , a slotted counter with slot length T is defined as $\gamma(\mathbf{S}) \triangleq (Z_1, Z_2, \dots)$, where Z_j ($j \geq 1$) is the number of points in the j th slot (i.e., the interval $[(j-1)T, jT)$) of \mathbf{S} . □

The slotted counter was first used to compress Poisson processes by Rubin in [18], where combined with proper reconstruction methods, it was shown to approximate the optimal performance predicted by the rate distortion function under the single-letter absolute-error fidelity criterion. This result, however, does not imply that the slotted counter is (near) optimal in our problem because our fidelity criterion is different. We refer to the quantization by a slotted counter as slotted quantization. It is easy to see that the aforementioned definition is equivalent to the point-wise quantizer $\tilde{\gamma}(t) = \lfloor t/T \rfloor$, where $t \in \mathbb{R}^+$.

For applications requiring an extremely low rate, it may be desirable to further compress the results of slotted quantization. To this end, we propose the following quantizer.

Definition 4.2: Given a point process \mathbf{S} , a one-bit quantizer is a binary quantization of the output of a slotted counter, defined as $\hat{\gamma}(\mathbf{S}) = (I_{\{Z_j > 0\}})_{j=1}^{\infty}$, where $\mathbf{Z} = \gamma(\mathbf{S})$, and $I_{\{\cdot\}}$ is the indicator function. □

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

Hereafter, we will refer to the quantization results of \mathbf{S}_1 and \mathbf{S}_2 by $\mathbf{X}^n = (X_j)_{j=1}^n$ and $\mathbf{Y}^n = (Y_j)_{j=1}^n$, respectively, the meaning of which will depend on the quantizers used. For the full side-information case (i.e., $R_2 = \infty$), we use $Y(s, t)$ to denote the number of epochs in \mathbf{S}_2 in the interval $[s, t)$.

If \mathbf{S}_i ($i = 1, 2$) are Poisson processes, then X_j 's and Y_j 's are i.i.d., and they can be delivered almost perfectly (for sufficiently large n) under the capacity constraints in (4) if and only if

$$\frac{H(X_1)}{T} \leq R_1, \quad \frac{H(Y_1)}{T} \leq R_2. \quad (9)$$

V. DETECTORS

In this section, we will present detectors for each quantization scheme proposed in Section IV. The detectors first compute the minimum fraction of chaff noise needed to generate the received measurements (assuming \mathcal{H}_1) and then declare \mathcal{H}_1 if this fraction is suspiciously small. In the rest of this section, we will discuss the following four cases:

- Case 1) \mathbf{S}_1 is under slotted quantization and \mathbf{S}_2 fully available;⁹
- Case 2) both \mathbf{S}_1 and \mathbf{S}_2 are under slotted quantization;
- Case 3) \mathbf{S}_1 is under one-bit quantization and \mathbf{S}_2 fully available;
- Case 4) both \mathbf{S}_1 and \mathbf{S}_2 are under one-bit quantization. Since the optimal detection performance in high capacity

⁹The detector when \mathbf{S}_2 is compressed and \mathbf{S}_1 fully available is analogous.

regime is already known [10], our analysis will focus on the low capacity (i.e., large slot length) regime.

A. Case I: Slotted Quantization With Full Side Information

Consider the case when \mathbf{S}_1 is under slotted quantization, and \mathbf{S}_2 is fully available. Then, the detector's observations are realizations \mathbf{x}^n and \mathbf{s}_2 .

To compute the minimum chaff noise in these realizations, 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: 1) $\mathbf{x}^n = \gamma(\mathbf{f}_1 \oplus \mathbf{w}_1)$, $\mathbf{s}_2 = \mathbf{f}_2 \oplus \mathbf{w}_2$; 2) the CTR is minimized. If both \mathbf{s}_1 and \mathbf{s}_2 are given, then the minimum chaff can be computed by an algorithm called "Bounded-Greedy-Match" (BGM) proposed by Blum *et al.* in [7]. For delay bound Δ , BGM works as follows:

- 1) sequentially match every epoch s in \mathbf{s}_1 with the first unmatched epoch in $[s, s + \Delta]$ in \mathbf{s}_2 ;
- 2) the matched epochs form $\mathbf{f}_i (i = 1, 2)$ and the unmatched epochs \mathbf{w}_i .

A self-explanatory pseudocode implementation of BGM is given in Algorithm 1.

Algorithm 1 Bounded Greedy Match (BGM)

Require: Realizations of point processes $(\mathbf{s}_1, \mathbf{s}_2)$; maximum delay Δ .

Ensure: Return the minimum number of chaff packets in given realizations.

- 1) initialize indices m, n , and counter C ;
- 2) **while** m, n are valid indices in $\mathbf{s}_1, \mathbf{s}_2$ **do**;
- 3) **if** $s_1(m) > s_2(n)$ **then** {If the pair $(s_1(m), s_2(n))$ is noncausal,};
- 4) $C \leftarrow C + 1, n \leftarrow n + 1$ {then $s_2(n)$ is chaff noise.};
- 5) **else if** $s_1(m) < s_2(n) - \Delta$ **then** {If the delay exceeds Δ ,};
- 6) $C \leftarrow C + 1, m \leftarrow m + 1$ {then $s_1(m)$ is chaff noise.};
- 7) **else** {Otherwise, the pair is matched,};
- 8) $m \leftarrow m + 1, n \leftarrow n + 1$ {and the indices are advanced.};
- 9) **return** C { C is the total number of chaff packets found by BGM.}.

Now that we only know \mathbf{x}^n and \mathbf{s}_2 , the idea is to reconstruct \mathbf{s}_1 from \mathbf{x}^n and apply BGM on the reconstructed processes. Based on this idea, we develop a chaff-inserting algorithm called "Slotted-Full Greedy Match" (SF-GM) as follows. Given $(\mathbf{x}^n, \mathbf{s}_2)$, SF-GM does the following:

- 1) construct a point process $\hat{\mathbf{s}}_1$ as bursts of x_j simultaneous epochs at $(j - 1)T$ ($j \geq 1$), as illustrated in Fig. 6;
- 2) run BGM on $(\hat{\mathbf{s}}_1, \mathbf{s}_2)$ with delay bound $T + \Delta$.

The optimality of SF-GM is provided by the following proposition.

Proposition 5.1: Under the quantization in Case I, SF-GM inserts the minimum number of chaff packets in transmitting an

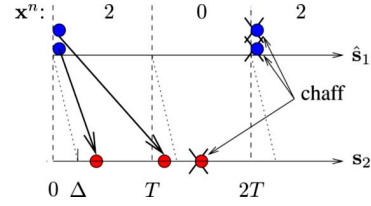


Fig. 6. SF-GM: greedy match between a batched process $\hat{\mathbf{s}}_1$ and \mathbf{s}_2 . Epochs not matched within delay $T + \Delta$ are considered to be chaff noise.

information flow according to the (quantized) schedule $(\mathbf{x}^n, \mathbf{s}_2)$ for any $(\mathbf{x}^n, \mathbf{s}_2)$.

Proof: See the Appendix. ■

Since SF-GM is optimal, we can compute the minimum number of chaff packets to mimic the measurements $(\mathbf{x}^n, \mathbf{s}_2)$ using SF-GM. This idea leads to the following detector.

Given $(\mathbf{x}^n, \mathbf{s}_2)$, define a detector δ_I as

$$\delta_I(\mathbf{x}^n, \mathbf{s}_2) = \begin{cases} 1, & \text{if } C_I/N \leq \tau_I \\ 0, & \text{o.w.} \end{cases}$$

where $\tau_I \in [0, 1]$ is a predetermined threshold, $N = \sum_{j=1}^n x_j + |\mathfrak{S}_2|$, and C_I is the number of chaff packets found by SF-GM in $(\mathbf{x}^n, \mathbf{s}_2)$, excluding chaff packets in $^{10} \mathfrak{S}_2 \cap [0, \Delta)$ (i.e., if $(\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2)$ is the chaff noise found by SF-GM for input $(\mathbf{x}^n, \mathbf{s}_2)$, then $C_I = |\mathfrak{W}_1| + |\mathfrak{W}_2 \setminus [0, \Delta)]$).

Under \mathcal{H}_1 , the actual number of chaff packets has to be at least C_I . Thus, δ_I has vanishing misprobability for all of the information flows with CTR bounded by τ_I a.s. To guarantee vanishing false alarm probability, on the other hand, τ_I cannot be too large. The overall performance of δ_I is characterized 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 the consistency-rate function (see Definition 3.3) of δ_I (combined with quantizers in Case I) is lower bounded by $\alpha_I(R_I^{-1}(R))$, where¹¹

$$\alpha_I(T) \triangleq \frac{1}{2\sqrt{\pi\lambda T}} - \frac{\Delta}{8T} \approx \frac{1}{2\sqrt{\pi\lambda T}} \quad (10)$$

$$R_I(T) \triangleq H(\text{Poi}(\lambda T))/T. \quad (11)$$

Furthermore, for any $\tau_I < \alpha_I(T)$, the false alarm probability decays exponentially with n .

Proof: See the Appendix. ■

The theorem contains the following results. For slot length T

1) the consistency level of δ_I is τ_I for any $\tau_I < \alpha_I(T)$;

2) the rate $R_I(T)$ suffices to deliver \mathbf{X}^n reliably for large n .

Therefore, by using a proper τ_I and the optimal encoding of \mathbf{X}^n , we can make δ_I achieve Chernoff-consistent detection for the noise level bounded by $\alpha_I(R_I^{-1}(R))$ under a transmission rate R . The performance guarantee in the other cases can be interpreted similarly.

¹⁰This is because packets in this interval may be relays of packets transmitted before the detector starts taking observations.

¹¹Here, $H(\text{Poi}(\lambda T))$ is the entropy of Poisson distribution with mean λT .

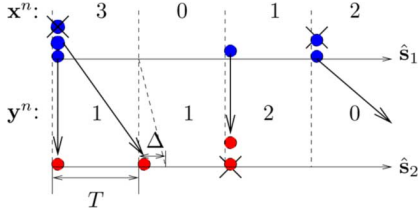


Fig. 7. SS-GM: match between batched processes \hat{s}_1 and \hat{s}_2 . Epochs not matched within delay $\lceil (\Delta/T) \rceil T$ are considered chaff noise.

B. Case II: Symmetric Slotted Quantization

Suppose that \mathbf{S}_1 and \mathbf{S}_2 are both compressed by slotted counters of slot length¹² T . We follow the procedure in Case I to develop a detector in this scenario.

To compute the minimum chaff noise, we develop a chaff-inserting algorithm called ‘‘Slotted-Slotted Greedy Match’’ (SS-GM) which is also based on BGM. Given $(\mathbf{x}^n, \mathbf{y}^n)$, SS-GM 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 $j \geq 1$;
- 2) run BGM on (\hat{s}_1, \hat{s}_2) with delay bound $\lceil (\Delta/T) \rceil T$, as illustrated in Fig. 7.

Algorithm SS-GM is optimal in minimizing the number of chaff packets, as stated in the following proposition.

Proposition 5.3: Under the quantization in Case II, SS-GM inserts the minimum number of chaff packets in transmitting an information flow according to any given schedule $(\mathbf{x}^n, \mathbf{y}^n)$.

Proof: See the Appendix. ■

Algorithm SS-GM provides a method to compute the minimum amount of chaff noise in the measurements, based on which we design a detector as follows.

Given $(\mathbf{x}^n, \mathbf{y}^n)$, define a detector δ_{II} as

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

where $\tau_{II} \in [0, 1]$ is a predetermined threshold, $N = \sum_{j=1}^n (x_j + y_j)$, and C_{II} is the number of chaff packets found by SS-GM in $(\mathbf{x}^n, \mathbf{y}^n)$, except for chaff packets in $\hat{\mathcal{G}}_2 \cap [0, \lceil (\Delta/T) \rceil T)$.

The optimality of SS-GM implies that the actual number of chaff packets under \mathcal{H}_1 is no smaller than C_{II} . Therefore, δ_{II} has vanishing misprobability for all of the information flows with CTR bounded by τ_{II} a.s. The consistency of δ_{II} is guaranteed by the following theorem.

Theorem 5.4: If \mathbf{S}_1 and \mathbf{S}_2 are Poisson processes of maximum rate λ , and T is large, then the consistency-rate function of δ_{II} is lower bounded by $\alpha_{II}(R_{II}^{-1}(R))$, where

$$\alpha_{II}(T) \triangleq \frac{c_1}{2\sqrt{\lambda T}} e^{-\lambda T/6}, \quad c_1 = 0.0014 \quad (12)$$

¹²The slot length T is chosen to be large enough to satisfy both capacity constraints.

¹³As in the computation of C_I , this adjustment is needed because packets at the beginning of \hat{s}_2 may be the relays of packets transmitted before the detector starts.

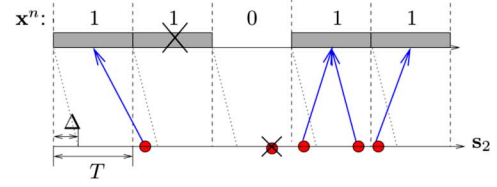


Fig. 8. OF-GM: Greedy matching starting from \mathbf{s}_2 . Each epoch in \mathbf{s}_2 is matched to the first unmatched nonempty slot (i.e., $x_j = 1$) that is no more than Δ earlier.

and $R_{II}(T) \triangleq R_I(T)$. Moreover, its false alarm probability decays exponentially with n for any $\tau_{II} < \alpha_{II}(T)$.

Proof: See the Appendix. ■

Note that as T increases, $\alpha_{II}(T)$ decays exponentially at the rate $O(e^{-\lambda T/6})$. Compared with the $O(1/\sqrt{T})$ decay of $\alpha_I(T)$, the results suggest that the consistency level decays much faster due to the quantization of \mathbf{S}_2 .

C. Case III: One-Bit Quantization With Full Side Information

Consider the scenario when \mathbf{S}_1 is compressed by one-bit quantization, and \mathbf{S}_2 is fully available.

This case is similar to Case I in Section V-A except that the observations are indicators instead of the exact counts. Clearly, more information is lost after one-bit quantization because when $x_j = 1$, there can be one or more epochs in slot j , which prevents a direct reconstruction of \mathbf{s}_1 . To overcome this difficulty, we use a greedy matching starting from epochs in \mathbf{s}_2 . Specifically, we develop a chaff-inserting algorithm called one-bit-full greedy match (OF-GM) which works as follows. Given $(\mathbf{x}^n, \mathbf{s}_2)$, OF-GM:

- 1) match every epoch in \mathbf{s}_2 with the earliest unmatched nonempty slot within delay Δ , as illustrated in Fig. 8;
- 2) unmatched epochs become chaff; each unmatched nonempty slot contains a chaff packet.

Algorithm OF-GM is the optimal chaff-inserting algorithm in Case III, as stated in the following proposition.

Proposition 5.5: Under the quantization in Case III, OF-GM inserts the minimum number of chaff packets in transmitting an information flow according to any given schedule $(\mathbf{x}^n, \mathbf{s}_2)$.

Proof: See the Appendix. ■

Based on OF-GM, we develop the following detector. Given $(\mathbf{x}^n, \mathbf{s}_2)$, the detector δ_{III} is defined as

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

where $\tau_{III} \in [0, 1]$ is a fixed threshold, C_{III} is the number of chaff packets found by OF-GM in $(\mathbf{x}^n, \mathbf{s}_2)$, excluding chaff packets in $\mathcal{G}_2 \cap [0, \Delta)$, and $\hat{N}_1 = -\log(1 - \bar{x})$ for $\bar{x} = (1/n) \sum_{j=1}^n x_j$. Here, \hat{N}_1 is the maximum-likelihood estimate of the mean number of epochs per slot in \mathbf{S}_1 based on the assumption that \mathbf{S}_1 is Poisson.

Proposition 5.5 guarantees that the actual number of chaff packets under \mathcal{H}_1 is no smaller than C_{III} . Moreover, under the Poisson assumption, \hat{N}_1 converges to the average traffic size per slot in \mathbf{S}_1 a.s. Thus, the statistic $C_{III}/(n\hat{N}_1 + |\mathcal{G}_2|)$ is upper bounded by the actual CTR a.s. as $n \rightarrow \infty$, implying that δ_{III}

has vanishing misprobability for CTR bounded by τ_{III} a.s. The overall performance of δ_{III} is guaranteed by the following theorem.

Theorem 5.6: If \mathbf{S}_1 and \mathbf{S}_2 are independent Poisson processes of maximum rate λ , and T is large, then δ_{III} has a consistency-rate function that is no smaller than $\alpha_{\text{III}}(R_{\text{III}}^{-1}(R))$, where¹⁴

$$\alpha_{\text{III}}(T) \triangleq \frac{1}{4}e^{-\lambda T} \quad (13)$$

$$R_{\text{III}}(T) \triangleq \begin{cases} \log 2/T & \text{if } \lambda T \geq \log 2 \\ h(e^{-\lambda T})/T, & \text{o.w.,} \end{cases} \quad (14)$$

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

Proof: See the Appendix. ■

Note that as T increases, $\alpha_{\text{III}}(T)$ decays exponentially with exponent λ , which is much faster than the $O(1/\sqrt{T})$ decay of $\alpha_{\text{I}}(T)$, indicating that for the same slot length, one-bit quantization significantly reduces consistency compared with slotted quantization. However, it does not imply that slotted quantization is better because the one-bit quantizer can use a much smaller slot length under the same capacity constraint.

D. Case IV: Symmetric One-Bit Quantization

Suppose that one-bit quantizers with the same slot length T are used for \mathbf{S}_1 and \mathbf{S}_2 . To compute the minimum chaff noise, we observe that if slot j in \mathbf{s}_1 is nonempty, then the corresponding slots within the delay bound (i.e., slots $j, \dots, j + \lceil \Delta/T \rceil$) in \mathbf{s}_2 cannot be all empty; otherwise, slot j in \mathbf{s}_1 must contain chaff noise. Similar arguments hold for \mathbf{s}_2 . Based on this observation, we develop an algorithm called ‘‘One-Bit-One-Bit–Greedy Match’’ (OO–GM) which works as follows. Given $(\mathbf{x}^n, \mathbf{y}^n)$, OO–GM inserts a chaff packet in slot j if

$$x_j > \sum_{k=j}^{j+\lceil \Delta/T \rceil} y_k, \text{ or } y_j > \sum_{k=j-\lceil \Delta/T \rceil}^j x_k$$

for $j = 1, \dots, n$.

Algorithm OO–GM computes the minimum amount of chaff noise as stated in the following proposition.

Proposition 5.7: Under the quantization in Case IV, OO–GM inserts the minimum number of chaff packets in transmitting an information flow according to any given schedule $(\mathbf{x}^n, \mathbf{y}^n)$.

Proof: See the Appendix. ■

Based on OO–GM, we develop a detector δ_{IV} as follows. Given $(\mathbf{x}^n, \mathbf{y}^n)$, the detector is defined as

$$\delta_{\text{IV}}(\mathbf{x}^n, \mathbf{y}^n) = \begin{cases} 1, & \text{if } C_{\text{IV}}/[n(\hat{N}_1 + \hat{N}_2)] \leq \tau_{\text{IV}}, \\ 0, & \text{o.w.} \end{cases}$$

where $\tau_{\text{IV}} \in [0, 1]$ is a predetermined threshold, C_{IV} is the number of chaff packets inserted by OO–GM in $(\mathbf{x}^n, \mathbf{y}^n)$, excluding chaff packets in $\tilde{\mathfrak{S}}_2 \cap [0, \lceil (\Delta)/(T) \rceil T)$, and \hat{N}_i ($i = 1, 2$) are defined as in δ_{III} as functions of \mathbf{x}^n and \mathbf{y}^n , respectively.

Detector δ_{IV} has vanishing misprobability as long as the CTR is bounded by τ_{IV} a.s. because of arguments similar to those in

¹⁴Here, $h(p)$ is the binary entropy function defined as $h(p) = -p \log p - (1-p) \log(1-p)$.

Section V-C. Its achievable consistency-rate function is given in the following theorem.

Theorem 5.8: If \mathbf{S}_1 and \mathbf{S}_2 are independent Poisson processes of maximum rate λ and T is large, then the consistency-rate function of δ_{IV} is at least $\alpha_{\text{IV}}(R_{\text{IV}}^{-1}(R))$, where

$$\alpha_{\text{IV}}(T) \triangleq \frac{(1 - e^{-\lambda T})}{2\lambda T} e^{-2\lambda T} \quad (15)$$

and $R_{\text{IV}}(T) \triangleq R_{\text{III}}(T)$. Furthermore, its false alarm probability decays exponentially if $\tau_{\text{IV}} < \alpha_{\text{IV}}(T)$.

Proof: See the Appendix. ■

For the same T , $\alpha_{\text{IV}}(T)$ decays 12 times faster than $\alpha_{\text{II}}(T)$. Again, it does not mean that slotted quantization is better because the slot lengths under different quantization schemes are different.

VI. ANALYSIS AND COMPARISON

We have designed detection systems by dividing the detection procedure into three steps:

- 1) quantization;
- 2) data transmission;
- 3) detection.

Although such separation has greatly reduced the complexity, it is still difficult to find the optimal design, especially for the quantizers. Instead, we will try to compare the detection performance under each quantization scenario in Section V to derive some heuristics.

A. Performance Analysis

Under the assumption that \mathbf{S}_i ($i = 1, 2$) are Poisson processes of maximum rate λ , we have derived performance lower bounds for the proposed detectors in Section V. The following result enables us to characterize the exact performance.

Lemma 6.1: The minimum CTR of j -GM ($j = \text{SF, SS, OF, OO}$) under \mathcal{H}_0 converges a.s. to a constant, denoted by $\alpha_i^*(T)$ ($i = \text{I, \dots, IV}$).

Proof: See the Appendix. ■

By Lemma 6.1, it can be shown that the consistency-rate function of detector δ_i ($i = \text{I, \dots, IV}$) is equal to $\alpha_i^*(R_i^{-1}(R))$ [for $R_i(\cdot)$ defined in (11) and (14)]. The computation of $\alpha_i^*(T)$ is rather involved [see (19)–(22) for their expressions]; instead, we resort to closed-form lower bounds, which leads to $\alpha_i(T)$ in (10), (12), (13), and (15).

Let the highest consistency level in Case i ($i = \text{I, \dots, IV}$) be $\tilde{\alpha}_i(T)$ (i.e., $\tilde{\alpha}_i(T) \triangleq \sup_{\delta} \{\text{consistency level of } \delta\}$ over all detectors δ under the quantization in Case i). We can bound $\tilde{\alpha}_i(T)$ as follows.

Theorem 6.2: Under the quantization schemes in Case i ($i = \text{I, \dots, IV}$), the optimal consistency level is bounded by

$$\alpha_i^*(T) \leq \tilde{\alpha}_i(T) \leq (E[|X - Y|])/(2\lambda T)$$

where X and Y are independent Poisson random variables with mean λT .

Proof: See the Appendix. ■

The theorem contains achievability and converse results regarding the performance under the proposed quantizers. Specifically, the proposed detectors may not be optimal because

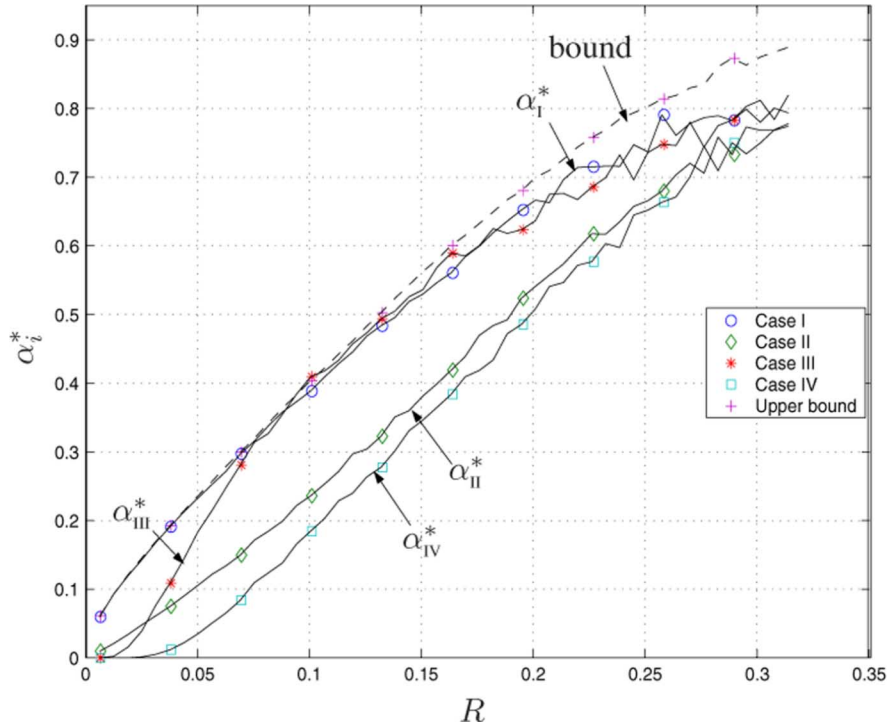


Fig. 9. Consistency-rate functions of $\delta_I \dots \delta_{IV}$ for various traffic rates: $\Delta = 1, 10^4$ slots. The overall consistency-rate function is bounded between the largest $\alpha_i^*(\cdot)$ and $1/(1+\lambda\Delta)$ (not shown). $\lambda = 0.1$.

the theorem only considers detection at the fusion center; the marginal distributions at sensors (before quantization) under the two hypotheses are still different and can be used for detection. Moreover, the minimum level of chaff noise needed to mimic \mathcal{H}_0 [in the sense of (8)] under the proposed quantization is upper bounded by $\mathbb{E}[|X - Y|]/(2\lambda T)$ (in the proof, an explicit method is given to mimic the distributions under \mathcal{H}_0).

Our results on performance analysis so far can be summarized as follows: for $T = R_i^{-1}(R)$

$$\alpha_i(T) \leq \alpha_i^*(T) \leq \tilde{\alpha}_i(T) \leq \phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y}) \leq \frac{\mathbb{E}[|X - Y|]}{2\lambda T}$$

where $\phi(\mathcal{H}_0; \mathbf{X}, \mathbf{Y})$ is given by (8) for the quantizers in Case i .

B. Numerical Comparison

In this section, we present a numerical comparison of the proposed detection systems from theoretical and operational perspectives.

1) *Consistency-Rate Functions*: Given capacity R , we compare the consistency-rate functions¹⁵ $\alpha_i^*(R_i^{-1}(R))$ ($i = I, \dots, IV$) together with the upper bound¹⁶ in Theorem 6.2. For the simplicity of implementation, we choose the range of R to guarantee that $R_i^{-1}(R) \geq \Delta$ (see Figs. 9–11).

The plots yield several observations. First, for small λ (Fig. 9), the detectors under slotted and one-bit quantization

¹⁵Since $\alpha_i^*(\cdot)$ is not in closed form, we calculate it by simulating the algorithms SF, SS, OF, and OO-GM on quantized independent Poisson processes of rate λ .

¹⁶Only the upper bound for $T = R_I^{-1}(R)$ is plotted because the bound for $T = R_{III}^{-1}(R)$ is much looser.

(i.e., δ_I versus δ_{II} and δ_{II} versus δ_{IV}) have similar performance; as λ increases (Figs. 10 and 11), the detectors under slotted quantization (δ_I and δ_{II}) outperform their counterparts. This observation clearly suggests that the quantization to use should depend on the traffic rate. For very light traffic, one-bit quantization achieves similar performance as slotted quantization at a lower complexity, whereas for heavier traffic, slotted quantization is much better.

Second, for fixed R , the consistency levels of all the detectors decreases with the increase of λ . This observation implies that it is more difficult to detect information flows in heavy traffic. Intuitively, this is because the relative maximum delay (i.e., the maximum delay Δ normalized by the average interarrival time $1/\lambda$) is an increasing function of the traffic rate λ and, thus, the delay constraint is relatively loose for heavy traffic.

Finally, $\alpha_I^*(R_I^{-1}(R))$ is close to the upper bound, especially at small R . Therefore, δ_I is near optimal for detection under slotted quantization and full side information.

2) *Simulations*: Besides the theoretical comparison, we also simulate the proposed detectors under the following traffic model. Under \mathcal{H}_0 , \mathbf{S}_1 and \mathbf{S}_2 are independent Poisson processes of rate λ . Under \mathcal{H}_1 , $\mathbf{S}_i = \mathbf{F}_i \oplus \mathbf{W}_i$ ($i = 1, 2$), where \mathbf{F}_2 is generated by adding i.i.d. uniform delays in $[0, \Delta]$ to a Poisson process \mathbf{F}_1 of rate $(1 - f_c)\lambda$ ($f_c \in (0, 1)$), and \mathbf{W}_i 's are independent Poisson processes (also independent of \mathbf{F}_i) of rate $f_c\lambda$. The parameter f_c is the CTR.

Assuming full side information,¹⁷ we compare the receiver operating characteristics (ROCs) of the detectors under slotted

¹⁷It can be shown that the quantization results of \mathbf{S}_2 under \mathcal{H}_1 are not i.i.d., which complicates the calculation of the communication rate. Thus, we only simulate the detectors with full side information.

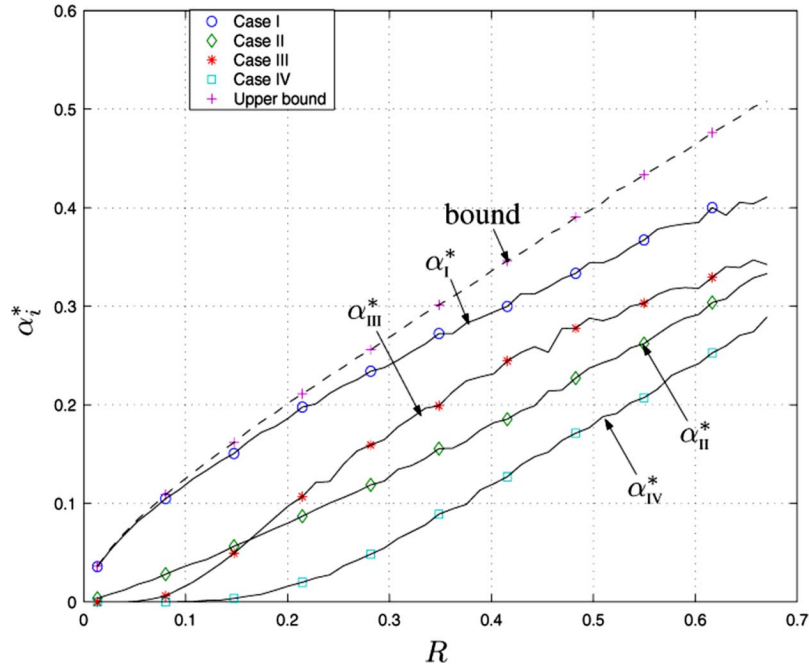


Fig. 10. Consistency-rate functions of $\delta_{I} \dots \delta_{IV}$ for various traffic rates: $\Delta = 1, 10^4$ slots. The overall consistency-rate function is bounded between the largest $\alpha_i^*(\cdot)$ and $1/(1+\lambda\Delta)$ (not shown). $\lambda = 0.5$.

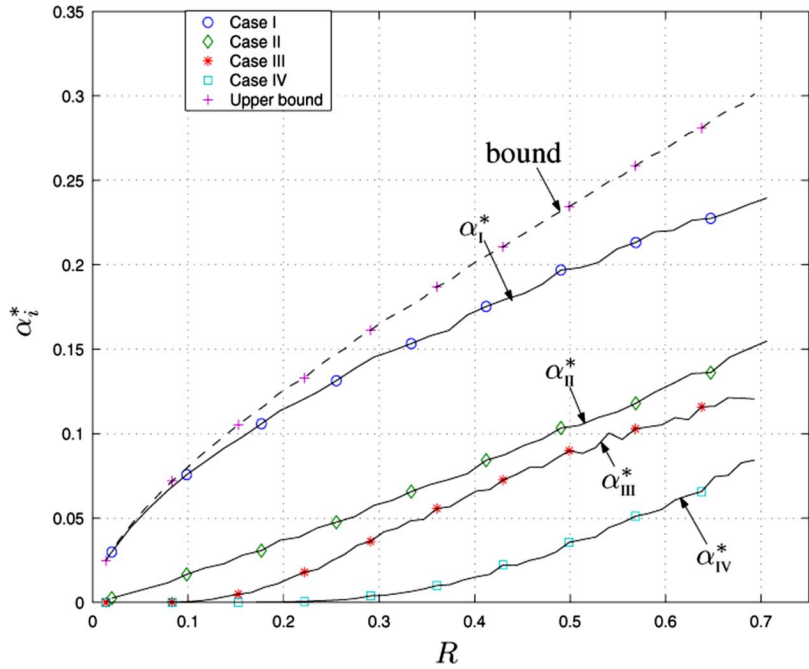


Fig. 11. Consistency-rate functions of $\delta_{I} \dots \delta_{IV}$ for various traffic rates: $\Delta = 1, 10^4$ slots. The overall consistency-rate function is bounded between the largest $\alpha_i^*(\cdot)$ and $1/(1+\lambda\Delta)$ (not shown). $\lambda = 1$.

and one-bit quantization together with those of the centralized detector in [9]; see Fig. 12. Specifically, under capacity $R = 0.5$, we simulate the proposed detectors with various thresholds under both hypotheses and plot the simulated false alarm and detection probabilities. We use a sample size of 100 packets per process and repeat the experiments for 10^4 times. The plot clearly shows the advantage of slotted quantization over one-bit quantization. Further simulations show that both ROCs will approach the upper left corner as the capacity increases.

VII. CONCLUSION

This paper addresses the distributed detection of information flows under capacity constraints. Our contributions are twofold: we introduce the notions of consistency-rate and undetectability-rate functions to characterize the fundamental limits of distributed information flow detection; moreover, we obtain achievability results by developing detection systems which are simple, yet effective. Our results show that slotted quantization and threshold detectors based on noise-level estimation

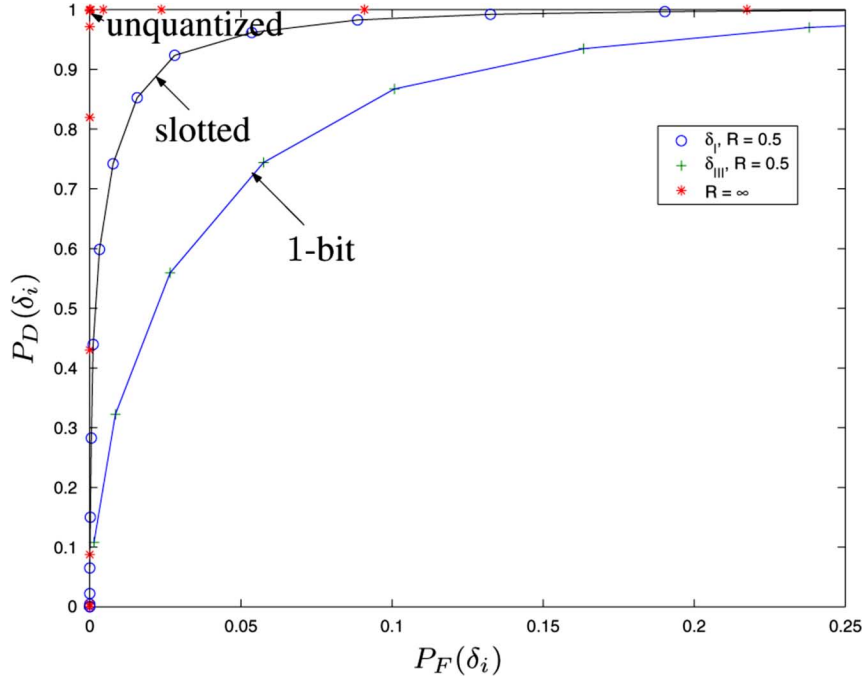


Fig. 12. ROCs of δ_I and δ_{III} with and without capacity constraints ($P_D(\delta_i)$ denotes the detection probability): $\lambda = 1$, $\Delta = 1$, $f_c = 0.2$, $R = 0.5$ (if finite capacity), 100 packets per process, 10 000 Monte Carlo runs.

provide competitive performance. This work, combined with the previous work on centralized detection in [10], establishes a mathematical framework for statistical information-flow detection.

APPENDIX

A. Proof of Proposition 5.1

First, we show that the matched pairs found by SF-GM indeed form a realization of an information flow. Let \mathbf{x}^m be the numbers of matched epochs in $\hat{\mathbf{s}}_1$, and $\hat{\mathbf{f}}_2 = (t_1, t_2, \dots)$ be the sequence of matched epochs in \mathbf{s}_2 . We construct a sequence $\hat{\mathbf{f}}_1 = (s_j)_{j \geq 1}$ as follows. For an epoch t_1 in $\hat{\mathbf{f}}_2$ matched to the same slot, we construct an epoch at t_1 in $\hat{\mathbf{f}}_1$; for an epoch t_2 matched to the previous slot, we construct an epoch in $\hat{\mathbf{f}}_1$ at the end of that slot. Such construction guarantees that slotted quantization of $\hat{\mathbf{f}}_1$ yields \mathbf{x}^m , and $(\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2)$ forms a realization of the information flow.

Then, we show that SF-GM is optimal. Since BGM is optimal, it remains to show that our choice of $\hat{\mathbf{s}}_1$ and the delay bound $T + \Delta$ minimizes chaff noise. Given $x_j (j = 1, \dots, n)$, the maximum interval for the x_j packets to be relayed through \mathbf{s}_2 is $[(j-1)T, jT + \Delta)$, and SF-GM allows all such matching. Therefore, SF-GM inserts the minimum chaff noise for any $(\mathbf{x}^n, \mathbf{s}_2)$. ■

B. Proof of Theorem 5.2

It is known that for Poisson processes of maximum rate λ , the rate $R_I(T)$ suffices to deliver \mathbf{X}^n (for large n). We only need to show that the consistency level of δ_I can approximate $\alpha_I(T)$. By our argument on the misprobability before Theorem 5.2, the desired result holds if we can prove that the false alarm

probability decays exponentially for any $\tau_I < \alpha_I(T)$. It suffices to consider the worst case where $\mathbf{S}_i (i = 1, 2)$ have equal rate λ .

First, for $\beta \triangleq \tau_I / \alpha_I(T)$, note that a necessary condition for false alarm is that $C_I/n \leq \sqrt{\beta} 2\lambda T \alpha_I(T)$ or $N/n \geq (2\lambda T) / \sqrt{\beta}$. By union bound, we have

$$P_F(\delta_I) \leq \Pr \left\{ \frac{C_I}{n} \leq \sqrt{\beta} 2\lambda T \alpha_I(T) \right\} + \Pr \left\{ \frac{N}{n} \geq \frac{2\lambda T}{\sqrt{\beta}} \right\}.$$

By Cramer's Theorem [19], the second term $\Pr\{N/n \geq 2\lambda T / \sqrt{\beta}\}$ decays exponentially with n for all $\tau_I < \alpha_I(T) (\beta < 1)$. The exponential decay of the first term remains to be shown.

Let C_k be the number of chaff packets inserted in the k th slot. Then, C_k is correlated with C_{k-1} and C_{k+1} . If, however, we run δ_I on every other slot, then the numbers of chaff packets $\check{C}_2, \check{C}_4, \dots$ will be i.i.d. Obviously, $C_I \geq \sum_{i=1}^{n/2} \check{C}_{2i}$. Thus, we have

$$\Pr \left\{ \frac{C_I}{n} \leq \eta \right\} \leq \Pr \left\{ \frac{2}{n} \sum_{i=1}^{n/2} \check{C}_{2i} \leq 2\eta \right\}. \quad (16)$$

Again by Cramer's Theorem, the right-hand side of (16) decays exponentially as long as $2\eta < \mathbb{E}[\check{C}_{2i}]$. Since $\check{C}_{2i} \stackrel{d}{=} (Y(0, T) - X_1)_+ + (X_1 - Y(0, T + \Delta))_+$, where $(x)_+ \triangleq \max(x, 0)$ by Gaussian approximation of Poisson random variables, we can approximate $(Y(0, T) - X_1)$ by a Gaussian random variable with zero mean and variance $2\lambda T$ and $(X_1 - Y(0, T + \Delta))$ with mean $-\lambda\Delta$ and variance $\lambda(2T + \Delta)$. Thus

$$\mathbb{E}[\check{C}_{2i}] \approx 2\sqrt{\frac{\lambda T}{\pi}} - \frac{1}{2}\lambda\Delta = 4\lambda T \alpha_I(T).$$

Therefore, (16) decays exponentially if $\eta < 2\lambda T \alpha_I(T)$, which is satisfied if $\tau_I < \alpha_I(T)$. This completes the proof. ■

C. Proof of Proposition 5.3

By the construction method in the Proof of Proposition 5.1, we see that the epochs matched by SS–GM can indeed be used to transmit information flow. Moreover, by the constraints of causality and bounded delay, packets in slot j can only be matched to packets in slots $j, \dots, j + \lceil (\Delta/T) \rceil$, and SS–GM allows all such matching. Combining this argument with the fact that BGM is optimal yields the optimality of SS–GM. ■

D. Proof of Theorem 5.4

Following the Proof of Theorem 5.2, it suffices to show that $\Pr\{C_{\text{II}}/n \leq \eta\}$ decays exponentially for any $\eta < 2\lambda T\alpha_{\text{II}}(T)$ for processes of equal rate λ . As before, we bound this probability by $\Pr\{(2/n)\sum_{i=1}^{n/2}\tilde{C}_{2i} \leq 2\eta\}$, where \tilde{C}_{2i} is the number of chaff packets in slot $2i$ if we only run δ_{II} on even slots. Now that

$$\begin{aligned} \mathbb{E}[\tilde{C}_2] &= \mathbb{E}[(Y_2 - X_1 - X_2)_+ + (X_2 - Y_2 - Y_3)_+] \\ &\text{and } (Y_2 - X_1 - X_2), (X_2 - Y_2 - Y_3) \sim \mathcal{N}(-\lambda T, 3\lambda T) \text{ for large } T, \text{ we have} \\ \frac{1}{2}\mathbb{E}[\tilde{C}_2] &\approx \sqrt{\frac{3\lambda T}{2\pi}}e^{-\lambda T/6} - \lambda T Q\left(\sqrt{\frac{\lambda T}{3}}\right) \\ &\approx c_1\sqrt{\lambda T}e^{-\lambda T/6} = 2\lambda T\alpha_{\text{II}}(T) \end{aligned} \quad (17)$$

where $c_1 = 0.0014$, and (17) is obtained by the approximation of $Q(\cdot)$ in [20] $Q(x) \approx (e^{-x^2/2})/(1.64x + \sqrt{0.76x^2 + 4})$. Since \tilde{C}_{2i} 's are i.i.d., by Cramer's Theorem, the probability $\Pr\{(2/n)\sum_{i=1}^{n/2}\tilde{C}_{2i} \leq 2\eta\}$ decays exponentially if $\eta < \mathbb{E}[\tilde{C}_2]/2 \approx 2\lambda T\alpha_{\text{II}}(T)$. ■

E. Proof of Proposition 5.5

By the construction in the Proof of Proposition 5.1, we can construct an information flow based on the matching found by OF–GM. Since OF–GM is a bounded greedy match starting from \mathbf{s}_2 , by symmetry, its optimality can be proved following the same arguments that prove the optimality of BGM (see [7]). ■

F. Proof of Theorem 5.6

The proof follows similar steps as in the Proof of Theorem 5.2. It is easy to see that for Poisson processes of rate bounded by λ , the rate $R_{\text{III}}(T)$ suffices to deliver the one-bit quantization results. For consistency results, it suffices to show that \hat{N}_1 converges to λT exponentially, and $\Pr\{C_{\text{III}}/n \leq \eta\}$ decays exponentially for any $\eta < 2\lambda T\alpha_{\text{III}}(T)$. For the former, since X_k 's are i.i.d., \bar{X} converges to its mean $1 - e^{-\lambda T}$ exponentially (by Cramer's Theorem), implying the exponential convergence of \hat{N}_1 . For the latter, by Cramer's Theorem, it is reduced to showing that $\mathbb{E}[\tilde{C}_{2i}] \geq 4\lambda T\alpha_{\text{III}}(T)$, where \tilde{C}_{2i} is the number of chaff packets per slot if δ_{III} is only run on even slots. If $X_{2i} = 0$ (with probability $e^{-\lambda T}$), then all of the epochs in $[(2i-1)T + \Delta, 2iT]$ in \mathbf{S}_2 will be chaff; if $X_{2i} = 1$ and $Y((2i-1)T, 2iT + \Delta) = 0$, then there will be at least one chaff packet in \mathbf{S}_1 in slot $2i$. Thus

$$\begin{aligned} \mathbb{E}[\tilde{C}_{2i}] &\geq e^{-\lambda T}\lambda(T - \Delta) + (1 - e^{-\lambda T})e^{-\lambda(T+\Delta)} \\ &\approx 4\lambda T\alpha_{\text{III}}(T). \end{aligned} \quad \blacksquare$$

G. Proof of Proposition 5.7

To construct an information flow, we use the following variation of BGM: match every $x_j = 1$ with all $y_k = 1$ for $k \in \{j, \dots, j + \lceil \Delta/T \rceil\}$ (both x_j and y_k can be matched repeatedly). Each pair of matched slots corresponds to a pair of matched epochs. It is easy to see that such matching generates (a realization of) an information flow.

On the other hand, if OO–GM inserts a chaff packet, then we must have a nonempty slot such that all of the corresponding slots in the other process within the delay and the causality bounds are empty. Thus, any other chaff-inserting algorithm would have to insert a chaff packet in that slot as well, which implies that OO–GM is optimal. ■

H. Proof of Theorem 5.8

For δ_{IV} , we take an approach slightly different from the others. Let $T_i (i \geq 1)$ denote the number of slots between the $(i-1)$ th and the i th chaff packets (including the latter). Then, we can write the average number of chaff packets as

$$\left\{ \frac{C_{\text{IV}}}{n} \leq \tau \right\} \stackrel{\text{d}}{=} \left\{ \frac{1}{n\tau} \sum_{i=1}^{n\tau} T_i \geq \frac{1}{\tau} \right\}. \quad (18)$$

To bound the probability of this event, we note that $T_i \leq \tilde{T}_i$ for i.i.d. random variables \tilde{T}_i representing the number of slots between chaff packets if chaff can only appear in even slots. We claim that $\tilde{T}_i \stackrel{\text{d}}{=} 2Z$, where Z has the geometric distribution $\Pr\{Z = n\} = (1-\rho)^{n-1}\rho (n \geq 1)$ for $\rho = 2e^{-2\lambda T}(1 - e^{-\lambda T})$. Here, Z is the number of slot pairs until the event A_{2i} occurs, which is given by

$$A_{2i} \triangleq \{X_{2i-1} + X_{2i} < Y_{2i}, \text{ or } X_{2i} > Y_{2i} + Y_{2i+1}\}$$

equivalent to the event $\{\exists \text{ chaff in slot } 2i\}$. Event A_{2i} is i.i.d. for $i = 1, 2, \dots$ with probability ρ . Based on the aforementioned results, we have

$$\Pr\{C_{\text{IV}}/n \leq \tau\} \leq \Pr\left\{ \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \tilde{T}_i \geq \frac{1}{\tau} \right\}$$

where $\tilde{n} = n\tau$. Now that $\mathbb{E}[\tilde{T}_i] = 2\mathbb{E}[Z] = 2/\rho = 1/(2\lambda T\alpha_{\text{IV}}(T))$, by Cramer's Theorem, $\Pr\{C_{\text{IV}}/n \leq \tau\}$ decays exponentially for any $\tau < 2\lambda T\alpha_{\text{IV}}(T)$. This result, coupled with the exponential convergence of $\hat{N}_i (i = 1, 2)$ (as in the Proof of Theorem 5.6), completes the proof. ■

I. Proof of Lemma 6.1

It is easy to see that the CTR is minimized when \mathbf{S}_1 and \mathbf{S}_2 have equal rate λ .

1) *Case I:* Let the k th interarrival time in \mathbf{S}_2 be $V_k \triangleq S_2(k) - S_2(k-1) (k \geq 1)$. Let $Z_j (j \geq 0)$ be the starting time for finding matches in the $(j+1)$ th slot of \mathbf{S}_2 ($Z_0 = 0$ by definition). For each $X_j (j \geq 1)$, SF–GM matches the X_j reconstructed epochs with epochs in \mathbf{S}_2 . Let $K_j (j \geq 1)$ be the index of the last epoch in \mathbf{S}_2 that is matched or assigned as chaff after the X_j epochs are matched, and $K_0 = 0$. Then, $Z_j (j \geq 1)$ satisfies the following

recursion:

$$Z_j = \min \left(\max \left(Z_{j-1} + \tilde{V}_{K_{j-1}+1} + \sum_{k=K_{j-1}+2}^{K_{j-1}+X_j} V_k - T, 0 \right), \Delta \right)$$

where $\tilde{V}_{K_{j-1}+1} = S_2(K_{j-1} + 1) - (j - 1)T - Z_{j-1}$ is a truncated inarrival time. Since V_k 's are i.i.d. exponential random variables, by the memoryless property of exponential distribution, we have $\{Z_j\}_{j=0}^{\infty}$ which is a random walk with reflecting barriers at 0 and Δ . The steps have the same distribution as $\sum_{k=1}^{X_1} V_k - T$, where X_1 is a Poisson variable with mean λT , V_k 's ($k \geq 1$) are i.i.d. exponential variables with mean $1/\lambda$, and V_k 's are independent of X_1 . It is easy to check that $\{Z_j\}_{j=0}^{\infty}$ is an ergodic process and, thus, its limiting distribution exists.

Let C_j ($j \geq 1$) denote the number of chaff packets in slot j . Then

$$C_j = \max(Y((j-1)T + Z_{j-1}, jT) - X_j, X_j - Y((j-1)T + Z_{j-1}, jT + \Delta), 0).$$

It is clear that C_j 's only depend on each other through Z_j . Now that $\{Z_j\}_{j=0}^{\infty}$ is ergodic, $(\sum_{j=1}^n C_j)/n$ converges a.s. Since the average traffic size per slot converges a.s., their ratio, which is the CTR, also converges a.s. The limit is given by

$$\alpha_{\text{I}}^*(T) \triangleq \frac{\mathbb{E}[\max(Y(Z, T) - X_1, X_1 - Y(Z, T + \Delta), 0)]}{2\lambda T} \quad (19)$$

where Z is distributed by the limiting distribution of $\{Z_j\}_{j=0}^{\infty}$.

2) *Case II:* Considering $T \geq \Delta$ for simplicity; the proof can be generalized to $T < \Delta$. Let Z_j ($j \geq 1$) be the number of packets in slot j in \mathbf{S}_2 which are matched to slots before j ($Z_1 = 0$ by definition). It can be shown that Z_j satisfies the following recursion:

$$Z_{j+1} = \min(\max(Z_j + X_j - Y_j, 0), Y_{j+1}).$$

Note that $\{Z_j\}_{j=1}^{\infty}$ is not Markovian because given Z_j, Z_{j+1} still depends on Z_{j-1} through Y_j . We can solve this problem by including Y_j in the state. Specifically, it can be shown that $\{(Z_j, Y_j)\}_{j=1}^{\infty}$ is a Markov chain and is ergodic. The number of chaff packets can be written as

$$C_j = \max(Y_j - Z_j - X_j, X_j - Y_j + Z_j - Y_{j+1}, 0).$$

Since $\{(Z_j, Y_j)\}_{j=1}^{\infty}$ is ergodic, and X_j 's and Y_j 's are i.i.d., we see that $(\sum_{j=1}^n C_j)/n$ converges a.s., which implies that the CTR converges a.s. The limiting CTR can be computed by

$$\alpha_{\text{II}}^*(T) \triangleq \frac{\mathbb{E}[\max(Y - Z - X_1, X_1 - Y + Z - Y_1, 0)]}{2\lambda T} \quad (20)$$

where (Z, Y) is distributed by the limiting distribution of $\{(Z_j, Y_j)\}_{j=1}^{\infty}$, X_1 and Y_1 are independent Poisson variables with mean λT , and (Z, Y) is independent of (X_1, Y_1) .

3) *Case III:* Let U_j ($j \geq 0$) is the starting time for finding matches in the $(j+1)$ th slot in \mathbf{S}_2 if $X_{j+1} = 0$ (define $U_0 = \Delta$); similarly, let L_j ($j \geq 0$) be the starting time if $X_{j+1} = 1$ (define $L_0 = 0$). Then, U_j and L_j satisfy the following recursions:

$$U_j = \begin{cases} \max(U_{j-1}, jT), & \text{if } X_j = 0 \\ jT + \Delta, & \text{o.w.} \end{cases}$$

$$L_j = \begin{cases} \max(L_{j-1}, jT), & \text{if } X_j = 0 \\ S_2(K), & \text{if } X_j = 1 \\ \max(S_2(K+1), jT), & \text{o.w.} \end{cases}$$

where $K \triangleq \inf\{k : S_2(k) \geq L_{j-1}\}$. It is easy to see that $\{(U_j, L_j)\}_{j=0}^{\infty}$ is a Markov process. Moreover, it can be shown that the process $\{(U_j - jT, L_j - jT)\}_{j=0}^{\infty}$ is ergodic. Then, we can write the number of chaff packets in slot j as

$$C_j = \begin{cases} Y(U_{j-1}, jT), & \text{if } X_j = 0 \\ I_{\{Y(L_{j-1}, jT + \Delta) = 0\}}, & \text{o.w.} \end{cases}$$

By the ergodicity of $\{(U_j - jT, L_j - jT)\}_{j=0}^{\infty}$ and the homogeneity of the Poisson processes, one can show that $(\sum_{j=1}^n C_j)/n$ converges a.s., and so does the CTR. The limit is given by (21), shown at the bottom of the page, where (U, L) is distributed by the limiting distribution of $\{(U_j - jT, L_j - jT)\}_{j=0}^{\infty}$.

4) *Case IV:* Consider the case $T \geq \Delta$. It is easy to check that the process $\{(X_{j-1}, X_j, Y_j, Y_{j+1})\}_{j=1}^{\infty}$ (where $X_0 \equiv 1$) is an ergodic Markov chain. Since the number of chaff packets in slot j is computed by

$$C_j = \max(Y_j - X_j - X_{j-1}, X_j - Y_j - Y_{j+1}, 0)$$

we see that $(\sum_{j=1}^n C_j)/n$ converges a.s. Therefore, the CTR converges a.s. The limit is given by

$$\alpha_{\text{IV}}^*(T) \triangleq \frac{\mathbb{E}[\max(Y - X - X_1, X - Y - Y_1, 0)]}{2\lambda T} \quad (22)$$

where (X_1, X, Y, Y_1) has the limiting distribution of $\{(X_{j-1}, X_j, Y_j, Y_{j+1})\}_{j=1}^{\infty}$. The proof can be generalized to $T < \Delta$. ■

$$\alpha_{\text{III}}^*(T) \triangleq \frac{\mathbb{E}[Y(U, T)]e^{-\lambda_1 T} + \Pr\{Y(L, T + \Delta) = 0\}(1 - e^{-\lambda_1 T})}{2\lambda T} \quad (21)$$

J. Proof of Theorem 6.5

The lower bound is by definition. For the upper bound, it suffices to consider Case I because the observations in the other cases are further compression. We give the following method to mimic distributions under \mathcal{H}_0 . Given realizations $(\mathbf{x}^n, \mathbf{s}_2)$, where \mathbf{x}^n is the result of slotted quantization, we construct a sequence $\hat{\mathbf{s}}_1$ as follows. For $j = 1, \dots, n$, if $x_j \leq y((j-1)T, jT)$, then randomly select x_j epochs from the j th slot of \mathbf{s}_2 ; otherwise, select all of the epochs in the j th slot of \mathbf{s}_2 and pick $x_j - y((j-1)T, jT)$ more epochs i.i.d. uniformly from $[(j-1)T, jT)$. The overall method is the following:

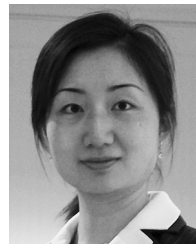
- 1) generate i.i.d. Poisson random variables $X_j (j \geq 1)$ with mean λT and an independent Poisson process \mathbf{S}_2 of rate λ ;
- 2) construct a process $\hat{\mathbf{S}}_1$ as described before;
- 3) use BGM with delay bound Δ to decompose $(\hat{\mathbf{S}}_1, \mathbf{S}_2)$ into an information flow and chaff noise.

Traffic generated by this method is equal to traffic under \mathcal{H}_0 in marginal distributions and the joint distribution after quantization. Moreover, there are at least $\min(X_j, Y((j-1)T, jT))$ pairs of matched epochs in slot j (i.e., the average number of chaff packets per slot is upper bounded by $\mathbb{E}[|X - Y|]$), where X and Y are independent Poisson variables with mean λT . Therefore, the CTR $\mathbb{E}[|X - Y|]/(2\lambda T)$ suffices to mimic \mathcal{H}_0 under the quantization in Case I and, thus, is an upper bound on the corresponding highest consistency level. ■

REFERENCES

- [1] S. Staniford-Chen and L. Heberlein, "Holding intruders accountable on the internet," in *Proc. IEEE Symp. Security and Privacy*, Oakland, CA, May 1995, pp. 39–49.
- [2] N. Ferguson and B. Schneier, *Practical Cryptography*. New York: Wiley, 2003.
- [3] Y. Zhang and V. Paxson, "Detecting stepping stones," in *Proc. 9th USENIX Security Symp.*, Aug. 2000, pp. 171–184.
- [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 *Proc. 5th Int. Symp. Recent Advances in Intrusion Detection*, 2002, pp. 17–35, Lect. Notes Comput. Sci. 2516.
- [5] T. He and L. Tong, "A signal processing perspective to stepping-stone detection," presented at the Conf. Information Sciences and Systems, Princeton, NJ, Mar. 2006.
- [6] P. Peng, P. Ning, D. Reeves, and X. Wang, "Active timing-based correlation of perturbed traffic flows with chaff packets," in *Proc. IEEE Int. Conf. Distributed Computing Systems Workshops*, Columbus, OH, Jun. 2005, pp. 107–113.
- [7] A. Blum, D. Song, and S. Venkataraman, "Detection of interactive stepping stones: Algorithms and confidence bounds," presented at the Conf. Recent Advance in Intrusion Detection (RAID), Sophia Antipolis, French Riviera, France, Sep. 2004.
- [8] L. Zhang, A. Persaud, A. Johnson, and Y. Guan, "Stepping stone attack attribution in non-cooperative IP networks," presented at the 25th IEEE Int. Performance Computing and Communications Conf., Phoenix, AZ, Apr. 2006.
- [9] T. He and L. Tong, "Detecting information flows: Improving chaff tolerance by joint detection," presented at the Conf. Information Sciences and Systems, Baltimore, MD, Mar. 2007.
- [10] T. He and L. Tong, "Detection of information flows," *IEEE Trans. Inf. Theory*, 2008, accepted for publication.

- [11] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.
- [12] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 4, pp. 533–542, Jul. 1986.
- [13] J. Shao, *Mathematical Statistics*. New York: Springer, 1999.
- [14] T. He and L. Tong, "Detecting encrypted stepping-stone connections," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1612–1623, May 2007.
- [15] A. Agaskar, T. He, and L. Tong, "A distributed scheme for detection of information flows in chaff," presented at the 42nd Conf. Information Sciences and Systems, Princeton, NJ, Mar. 2008.
- [16] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, Jun. 1995.
- [17] H. P. Boswijk, A. Lucas, and N. Taylor, "A comparison of parametric, semi-nonparametric, adaptive, and nonparametric cointegration tests," Feb. 1999, Tinbergen Inst. Discussion Papers 99-012/4, Tinbergen Inst.
- [18] I. Rubin, "Information rates and data-compression schemes for poisson processes," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 200–210, Mar. 1974.
- [19] F. den Hollander, *Large Deviations (Fields Institute Monographs, 14)*. Providence, RI: Amer. Math. Soc., 2000.
- [20] N. Kingsbury, Approximation formulae for the Gaussian error integral $Q(x)$ Tech. Rep. m11067, Jun. 2005. [Online]. Available: <http://cnx.org/content/m11067/latest/>, Connexions.



Ting He (S'04-M'07) received the B.S. degree in computer science from Peking University, Peking, China, in 2003 and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, in 2007.

She joined the Network Technologies Department at IBM T. J. Watson Research Center, Hawthorne, NY, in 2007. At IBM, she was with the International Technology Alliance (ITA) and on IBM projects on network technologies. She was an Undergraduate Research Assistant in the Micro Processor Research and Development Center of Peking University from 2001 to 2003.

Ms. He was a member of the Adaptive Communications & Signal Processing Group (ACSP) from 2003 to 2007 while at Cornell University. She received the Best Student Paper Award at the 2005 International Conference on Acoustic, Speech and Signal Processing (ICASSP). She was an Outstanding College Graduate of Beijing Area and an Outstanding Graduate of Peking University in 2003. She was a winner of the Excellent Student Award of Peking University during 1999-2002 for academic excellence and was a recipient of Canon, Sony, and Yang-Wang Academician scholarships.



Lang Tong (F'05) received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, in 1991.

He joined Cornell University, Ithaca, NY, in 1998, where he is now the Irwin and Joan Jacobs Professor in Engineering. Prior to joining Cornell University, he was with West Virginia University, Morgantown, and the University of Connecticut, Storrs. He also was the 2001 Cor Wit Visiting Professor at the Delft University of Technology, Delft, The Netherlands.

He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, Stanford, CA, in 1991.

Dr. Tong received the Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 best paper award (with M. Dong) from the IEEE Signal Processing Society, and the 2004 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society (with P. Venkitasubramaniam and S. Adireddy). He is also a coauthor of five student paper awards. He received the Young Investigator Award from the Office of Naval Research.