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—Information flow, Distributed detection, Traffic analysis, Information-theoretic limits.

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 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 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 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 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 bounded delay, then there will be 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

---

1Corresponding author.

T. He is with IBM T. J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532. Phone: 914-784-7183. Fax: 914-784-6225. Email: the@us.ibm.com.

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. Phone: 607-255-3900. Fax: 607-255-9072. Email: ltong@ece.cornell.edu.

Majority of this work was completed when Ting He was at Cornell University. It is supported in part by the National Science Foundation under award CCF-0635070 and 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. Part of this work was published in IEEE ISIT 2007 and IEEE Asilomar 2007.
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 do 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 Sec 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 \( H_0 \) that models the normal traffic, we assume that epochs measured by the local sensors are statistically independent, whereas under the alternative \( H_1 \), there exists a flow of packets 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 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 non-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 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 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 co-located 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 co-located with any sensor, and all 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. Suffinely 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 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 multi-sensor case, preliminary results of which have been presented in [15].

The rest of the 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 the paper with remarks.

II. Problem Formulation

A. Notation

We use the convention that uppercase letters denote random variables, lowercase letters realizations, boldface letters vectors, and plain letters scalars. For example, we denote a point process by \( S \), its realization \( s \), the \( k \)th epoch \( S(k) \), and realization of the \( k \)th epoch \( s(k) \). Given a realization \( s \) of a point process \( S \), we use \( S \) to denote the set of elements in this realization. Given two realizations of point processes \( (a_1, a_2, \ldots) \) and \( (b_1, b_2, \ldots) \), \( \oplus \) is the superposition operator defined as \( (a_k)_{k=1}^\infty \oplus (b_k)_{k=1}^\infty = (c_k)_{k=1}^\infty \), where \( c_1 \leq c_2 \leq \ldots \) and \( \{a_k\}_{k=1}^\infty \cup \{b_k\}_{k=1}^\infty = \{c_k\}_{k=1}^\infty \).
B. Flow Model and Hypotheses
Let $S_1$ and $S_2$ denote the transmission activities of nodes $A$ and $B$, respectively, i.e.,

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

(1)

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

We first introduce the definition of information flow.

Definition 2.1: A pair of processes $(F_1, F_2)$ is an information flow if for every realization $(f_1, f_2)$, there exists a bijection $g : \mathcal{F}_1 \to \mathcal{F}_2$ 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 $\Delta$. The bounded delay constraint, first proposed by Donoho et al. in [4], is often implied by reliable communication protocols.

Next, we say that $(S_1, S_2)$ contains an information flow if, as illustrated in Fig. 2, $S_i$ ($i = 1, 2$) can be partitioned into a subsequence $F_i$ of information-carrying packets and a subsequence $W_i$ of the so-called chaff (noise) packets, i.e.,

$$S_i = F_i \oplus W_i, \quad i = 1, 2,$$

(2)

where $(F_1, F_2)$ is an information flow. Here the chaff processes $W_i$ ($i = 1, 2$) model all 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:

$$\mathcal{H}_0 : S_1, S_2 \text{ are independent},$$

$$\mathcal{H}_1 : (S_1, S_2) \text{ contains an information flow},$$

(3)

for known delay constraint $\Delta$ by observing compressed measurements sent by local sensors.

We assume that the marginal distributions of $S_i$ ($i = 1, 2$) 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-bit message) to the fusion center, and the error probabilities can be made arbitrarily small if there are enough measurements. This is a partially nonparametric hypothesis testing$^2$ because no statistical assumptions are imposed on the correlation of $S_1$ and $S_2$ 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 processes $S_i$ ($i = 1, 2$) are compressed into $q_i^{(t)}(S_i)$ which are delivered to the fusion center, and then the detector makes a decision in the form of

$$\hat{\theta}_t = \delta_t(q_1^{(t)}(S_1), q_2^{(t)}(S_2)),$$

where$^3 \hat{\theta}_t \in \{0, 1\}$. The capacity constraints are expressed as

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

(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. Figure 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) defined below, which is reminiscent of the more familiar term signal-to-noise ratio.

Definition 3.1: Given realizations $(f_1, f_2)$ and $(w_1, w_2)$ of an information flow and its chaff noise, the chaff-to-traffic ratio$^4$

$^1$Assume no simultaneous transmission.

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

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

$^4$The unit of $R_i$ ($i = 1, 2$) is nats per unit time.
The consistency-rate function maps capacity constraints to the maximum achievable consistency level among all the detection systems satisfying the constraints, providing the optimal tradeoff between communication cost and detection performance. Similarly, if the above 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 systems, which can then be used to evaluate the system performance.

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 $H_0$ 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

$$\phi(R_1, R_2) \triangleq \max_{P_1} \phi(H_0; X, Y),$$

(7)

where $(X, Y)$ are auxiliary random variables from a set $P_1$ defined by

$$P_1 = \{ (X, Y) : X \rightarrow S_1 \rightarrow S_2 \rightarrow Y, \lim sup_{t \rightarrow \infty} \frac{1}{t} I(S_1; X) \leq R_1, \lim sup_{t \rightarrow \infty} \frac{1}{t} I(S_2; Y) \leq R_2 \}.$$

and

$$\phi(H_0; X, Y) \triangleq \inf_{r \in [0, 1]} : \exists (F_1, W_i)_{i=1}^2 \text{ such that } (S_i)_{i=1}^2 \text{ is an information flow; }$$

(8)

$$3) \text{ CTR} \leq r \text{ a.s.}).$$

In (8), $(X, Y) | (S_i)_{i=1}^2$ is the marginal of $(X, Y)$ in $(X, Y, S_1, S_2)$ specified by the distribution of $(S_1, S_2)$ and the conditional distribution $Q_1(X|S_1), Q_2(Y|S_2)$. □

The undetectability-rate function maps capacity constraints to the minimum level of chaff noise required to make an information flow mimic $H_0$ in distribution under the optimal quantization. Here the conditional distributions $Q_1(X|S_1), Q_2(Y|S_2)$ specify the quantizers of $S_1$ and $S_2$, and (8) calculates the minimum noise level (condition (3)) for an information flow (condition (2)) to appear identical

4Note that $P_1$ is well-defined because $S_i (i = 1, 2)$ have the same distributions under both hypotheses.

5The notation “$\triangleq$” means equal in distribution.

6We can write the conditional distribution $Q(X, Y|S_1, S_2)$ in product form because the quantization of the two processes is independent.
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 noise level over all the 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 $R_i \geq 0$ ($i = 1, 2$), $\alpha(R_1, R_2) \leq \phi(R_1, R_2)$.

**Proof:** By (8), there exists an information flow $(F_1, F_2)$ and chaff noise $(W_1, W_2)$ constituting $\phi(\mathcal{H}_0; X, Y)$ fraction of the total traffic such that the mixture $F_i \oplus W_i$ ($i = 1, 2$) appears statistically identical with traffic under $\mathcal{H}_0$ to both the sensors and the detector. No detector can detect this flow consistently, and thus the consistency level under these quantizers is bounded by $\phi(\mathcal{H}_0; X, Y)$. Therefore, maximizing $\phi(\mathcal{H}_0; X, Y)$ over all 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 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 proved 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 $\lambda$ is the rate of $S_i$ ($i = 1, 2$). Under finite capacities, the results are unknown.

![Fig. 5. The achievable and unachievable regions of distributed information flow detection in the consistency-rate space ($R_1 = R_2 \equiv R$). 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 purpose.](image)

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

a. 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.

b. 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 that $\alpha(R_1, R_2) \leq 1/(1+\lambda\Delta)$ 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 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 $S$, a slotted counter with slot length $T$ is defined as $\gamma(S) = (Z_1, Z_2, \ldots)$, 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 $S$. □

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 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 above definition is equivalent to the point-wise quantizer $\hat{\gamma}(t) = [t/T]$, where $t \in \mathbb{R}^+$. For applications requiring 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 $S$, a one-bit quantizer is a binary quantization of the output of a slotted counter, defined as $\hat{\gamma}(S) = (I_{\{Z_j > 0\}})_{j=1}^{\infty}$, where $Z = \gamma(S)$, 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 $S_1$ and $S_2$ by $X^n = (X_j)_{j=1}^n$ and $Y^n = (Y_j)_{j=1}^n$, 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 $S_2$ in the interval $[s, t]$.

If $S_i$ ($i = 1, 2$) 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

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

(9)
V. DETECTORS

In this section, we will present detectors for each of the quantization schemes 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: I) $S_1$ is under slotted quantization and $S_2$ fully available; II) both $S_1$ and $S_2$ are under slotted quantization; III) $S_1$ is under one-bit quantization and $S_2$ fully available; IV) both $S_1$ and $S_2$ are under one-bit quantization. Since the optimal detection performance in high capacity 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 $S_1$ is under slotted quantization, and $S_2$ is fully available. Then the detector’s observations are realizations $x^n$ and $s_2$.

To compute the minimum chaff noise in these realizations, we want to find realizations of an information flow $(f_i)_{i=1}^2$ and chaff noise $w_i$ $(i = 1, 2)$ such that: i) $x^n = \gamma(f_1 \oplus w_1)$, $s_2 = f_2 \oplus w_2$; ii) the CTR is minimized. If both $s_1$ and $s_2$ 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 $\Delta$, BGM works as follows:

1) sequentially match every epoch $s$ in $s_1$ with the first unmatched epoch in $[s, s + \Delta]$ in $s_2$;
2) the matched epochs form $f_i$ $(i = 1, 2)$ and the unmatched epochs $w_i$.

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

Algorithm 1 Bounded Greedy Match (BGM)

Require: Realizations of point processes $(s_1, s_2)$; maximum delay $\Delta$.

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 $s_1$, $s_2$ do
3: if $s_1(m) > s_2(n)$ then \{If the pair $(s_1(m), s_2(n))$ is noncausal,\}
4: \hspace{1em} $C \leftarrow C + 1$, $n \leftarrow n + 1$ \{then $s_2(n)$ is chaff noise.\}
5: else if $s_1(m) < s_2(n) \Delta$ then \{If the delay exceeds $\Delta$,\}
6: \hspace{1em} $C \leftarrow C + 1$, $m \leftarrow m + 1$ \{then $s_1(m)$ is chaff noise.\}
7: else \{Otherwise, the pair is matched,\}
8: \hspace{1em} $m \leftarrow m + 1$, $n \leftarrow n + 1$ \{and the indices are advanced.\}
9: return $C$ \{$C$ is the total number of chaff packets found by BGM.\}

Now that we only know $x^n$ and $s_2$, the idea is to reconstruct $s_1$ from $x^n$ 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 $(x^n, s_2)$, SF-GM does the following:

1) construct a point process $\hat{s}_1$ as busts of $x_j$ simultaneous epochs at $(j-1)T$ $(j \geq 1)$, as illustrated in Fig. 6;
2) run BGM on $(\hat{s}_1, s_2)$ 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 information flow according to the (quantized) schedule $(x^n, s_2)$ for any $(x^n, s_2)$.

Proof: See Appendix.

Since SF-GM is optimal, we can compute the minimum number of chaff packets to mimic the measurements $(x^n, s_2)$ using SF-GM. This idea leads to the following detector.

Given $(x^n, s_2)$, define a detector $\delta_i$ as

$$\delta_i(x^n, s_2) = \begin{cases} 1 & \text{if } C_i/N \leq \tau_i, \\ 0 & \text{o.w.,} \end{cases}$$

where $\tau_i \in [0, 1]$ is a predetermined threshold, $N = \sum_{j=1}^n x_j + |S_2|$, and $C_i$ is the number of chaff packets found by SF-GM in $(x^n, s_2)$, excluding chaff packets in $10 S_2 \cap [0, \Delta)$, i.e., if $(w_1, w_2)$ is the chaff noise found by SF-GM for input $(x^n, s_2)$, then $C_i = |W_1| + |W_2 \setminus [0, \Delta]|$.

Under $\mathcal{H}_1$, the actual number of chaff packets has to be at least $C_i$. Thus, $\delta_i$ vanishes miss probability for all the information flows with CTR bounded by $\tau_i$ a.s. To guarantee vanishing false alarm probability, on the other hand, $\tau_i$ cannot be too large. The overall performance of $\delta_i$ is characterized by the following theorem.

Theorem 5.2: If $S_1$ and $S_2$ are Poisson processes of rates bounded by $\lambda$, and $T$ is large, then the consistency-rate function (see Definition 3.3) of $\delta_i$ (combined with quantizers in Case I) is lower bounded by $\alpha_i(R^{-1}_i(R))$, where

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

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

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 $\lambda T$. 

Fig. 6. SF-GM: greedy match between a batched process $\hat{s}_1$ and $s_2$. Epochs not matched within delay $T + \Delta$ are considered as chaff noise.
Furthermore, for any \( \tau_i < \alpha_i(T) \), the false alarm probability decays exponentially with \( n \).

**Proof:** See Appendix.

The theorem contains the following results: for slot length \( T \),

1) the consistency level of \( \delta_i \) is \( \tau_i \) for any \( \tau_i < \alpha_i(T) \);
2) the rate \( R_\ell(T) \) suffices to deliver \( X^n \) reliably for large \( n \).

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

**B. Case II: Symmetric Slotted Quantization**

Suppose that both \( S_1 \) and \( S_2 \) are both compressed by slotted counters of slot length\(^{12}T \). 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 \( (x^n, y^n) \), 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 \( \lceil \frac{\Delta}{T} \rceil T \), as illustrated in Fig. 7.

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

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 \( (x^n, y^n) \).

**Proof:** See 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 \( (x^n, y^n) \), define a detector \( \delta_\ell \) as

\[
\delta_\ell(x^n, y^n) = \begin{cases} 
1 & \text{if } C_u/N \leq \tau_\ell, \\
0 & \text{o.w.}
\end{cases}
\]

where \( \tau_\ell \in [0, 1] \) is a predetermined threshold, \( N = \sum_{j=1}^n (x_j + y_j) \), and \( C_u \) is the number of chaff packets found by SS-GM in \( (x^n, y^n) \), except for chaff packets in \(^{13}\) \( \hat{s}_2 \cap [0, \lceil \frac{\Delta}{T} \rceil T] \).

The optimality of SS-GM implies that the actual number of chaff packets under \( H_1 \) is no smaller than \( C_u \). Therefore, \( \delta_\ell \) has vanishing miss probability for all the information flows with CTR bounded by \( \tau_\ell \) a.s. The consistency of \( \delta_\ell \) is guaranteed by the following theorem.

**Theorem 5.4:** If \( S_1 \) and \( S_2 \) are Poisson processes of maximum rate \( \lambda \), and \( T \) is large, then the consistency-rate function of \( \delta_\ell \) is lower bounded by \( \alpha_\ell(R_\ell^{-1}(R)) \), where

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

and \( R_\ell(T) \triangleq R_\ell(T) \). Moreover, its false alarm probability decays exponentially with \( n \) for any \( \tau_\ell < \alpha_\ell(T) \).

**Proof:** See Appendix.

Note that as \( T \) increases, \( \alpha_\ell(T) \) decays exponentially at the rate \( O(e^{-\lambda T/6}) \). Compared with the \( O(1/\sqrt{T}) \) decay of \( \alpha_\ell(T) \), the results suggest that the consistency level decays much faster due to the quantization of \( S_2 \).

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

Consider the scenario when \( S_1 \) is compressed by one-bit quantization, and \( S_2 \) 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 \( s_1 \). To overcome this difficulty, we use a greedy matching starting from epochs in \( s_2 \). Specifically, we develop a chaff-inserting algorithm called “One-bit-Full Greedy Match” (OF-GM) which works as follows.

Given \( (x^n, s_2) \), OF-GM:

1) match every epoch in \( s_2 \) with the earliest unmatched nonempty slot within delay \( \Delta \), as illustrated in Fig. 8;
2) unmatched epochs become chaff; each unmatched nonempty slot contains a chaff packet.

![Fig. 8. OF-GM: Greedy matching starting from \( s_2 \). Each epoch in \( s_2 \) is matched to the first unmatched nonempty slot (i.e., \( x_j = 1 \)) that is no more than \( \Delta \) earlier.](image)

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

\(^{13}\)As in the computation of \( C_u \), this adjustment is needed because packets at the beginning of \( \hat{s}_2 \) may be the relays of packets transmitted before the detector starts.
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 \((x^n, s_2)\).

Proof: See Appendix. ■

Based on OF-GM, we develop the following detector. Given \((x^n, s_2)\), the detector \(\delta_{II}(x^n, s_2)\) is defined as

\[
\delta_{II}(x^n, s_2) = \begin{cases} 
1 & \text{if } C_{II}/(n \hat{N}_1 + |S_2|) \leq \tau_{II}, \\
0 & \text{otherwise},
\end{cases}
\]

where \(\tau_{II} \in [0, 1]\) is a fixed threshold, \(C_{II}\) is the number of chaff packets found by OF-GM in \((x^n, s_2)\), excluding chaff packets in \(S_2 \cap [0, \Delta]\), 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 Likelihood estimate of the mean number of epochs per slot in \(S_1\) based on the assumption that \(S_1\) is Poisson.

Proposition 5.5 guarantees that the actual number of chaff packets under \(T_{II}\) is no smaller than \(C_{II}\). Moreover, under the Poisson assumption, \(\hat{N}_1\) converges to the average traffic size per slot in \(S_1\) a.s. Thus the statistic \(C_{II}/(n \hat{N}_1 + |S_2|)\) is upper bounded by the actual CTR a.s. as \(n \to \infty\), implying that \(\delta_{II}\) has vanishing miss probability for CTR bounded by \(\tau_{II}\) a.s. The overall performance of \(\delta_{II}\) is guaranteed by the following theorem.

Theorem 5.6: If \(S_1\) and \(S_2\) are independent Poisson processes of maximum rate \(\lambda\), and \(T\) is large, then \(\delta_{II}\) has a consistency-rate function no smaller than \(\alpha_{II}(R_{II}^{-1}(R))\), where

\[
\alpha_{II}(T) = \frac{1}{4} e^{-\lambda T}, \quad (13)
\]

\[
R_{II}(T) = \begin{cases} 
\log 2/T & \text{if } \lambda T \geq \log 2, \\
h(e^{-\lambda T})/T & \text{otherwise},
\end{cases}
\]

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

Proof: See Appendix. ■

Note that as \(T\) increases, \(\alpha_{II}(T)\) decays exponentially with exponent \(\lambda\), which is much faster than the \(O(1/\sqrt{T})\) decay of \(\alpha_{I}(T)\), indicating that for the same slot length, one-bit quantization significantly reduces consistency compared with slotted quantization. It, however, 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 both \(S_1\) and \(S_2\). To compute the minimum chaff noise, we observe that if slot \(j\) in \(S_1\) is nonempty, then the corresponding slots within the delay bound (i.e., slots \(j, \ldots, j + \lceil \Delta/T \rceil\)) in \(S_2\) cannot be all empty; otherwise, slot \(j\) in \(S_1\) must contain chaff noise. Similar arguments hold for \(S_2\). Based on this observation, we develop an algorithm called “One-bit-One-bit Greedy Match” (OO-GM) which works as follows. Given \((x^n, y^n)\), OO-GM inserts a chaff packet in slot \(j\) if

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

for \(j = 1, \ldots, n\).

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 \((x^n, y^n)\).

Proof: See Appendix. ■

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

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

where \(\tau_{IV} \in [0, 1]\) is a predetermined threshold, \(C_{IV}\) is the number of chaff packets inserted by OO-GM in \((x^n, y^n)\), excluding chaff packets in \(S_2 \cap [0, \Delta/T]\), and \(\hat{N}_i\) \((i = 1, 2)\) are defined as in \(\delta_{II}\) as functions of \(x^n\) and \(y^n\), respectively. Detector \(\delta_{IV}\) has vanishing miss probability as long as the CTR is bounded by \(\tau_{IV}\) a.s. because of arguments similar to those in Section V-C. Its achievable consistency-rate function is given in the following theorem.

Theorem 5.8: If \(S_1\) and \(S_2\) are independent Poisson processes of maximum rate \(\lambda\), and \(T\) is large, then the consistency-rate function of \(\delta_{IV}\) is at least \(\alpha_{IV}(R_{IV}^{-1}(R))\), where

\[
\alpha_{IV}(T) = \frac{(1 - e^{-\lambda T}) e^{-2\lambda T}}{2\lambda T}, \quad (15)
\]

and \(R_{IV}(T) \geq R_{II}(T)\). Furthermore, its false alarm probability decays exponentially if \(\tau_{IV} < \alpha_{IV}(T)\).

Proof: See Appendix. ■

For the same \(T\), \(\alpha_{IV}(T)\) decays 12 times faster than \(\alpha_{II}(T)\). 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—quantization, data transmission, and 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 of the quantization scenarios in Section V to derive some heuristics.

A. Performance Analysis

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

14Here \(h(p)\) is the binary entropy function defined as \(h(p) = -p \log p - (1 - p) \log (1 - p)\).
Lemma 6.1: The minimum CTR of $j$-GM ($j = SF, SS, OF, OO$) under $\mathcal{H}_0$ converges a.s. to a constant, denoted by $\alpha_i^*(T)$ ($i = I, \ldots, IV$).

Proof: See Appendix.

By Lemma 6.1, it can be shown that the consistency-rate function of detector $\delta_i$ ($i = I, \ldots, IV$) is equal to $\alpha_i^*(R_i^{-1}(R))$ (for $R_i(\cdot$ defined in (11,14)). 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 $\tilde{\alpha}_i(T)$ in (10, 12, 13, 15).

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

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

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

where $X$ and $Y$ are independent Poisson random variables with mean $\lambda T$.

Proof: See Appendix.

The theorem contains both achievability and converse results regarding the performance under the proposed quantizers. Specifically, the proposed detectors may not be optimal because it 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 $\mathbb{E}[(X - Y)^2]/(2\lambda T)$ (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 $T = R_i^{-1}(R)$,

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

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

B. Numerical Comparison

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

1) Consistency-Rate Functions: Given capacity $R$, we compare the consistency-rate functions$^{15}$ $\alpha_i^*(R_i^{-1}(R))$ ($i = I, \ldots, IV$) together with the upper bound$^{16}$ 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 Fig. 9–11.

The plots yield several observations. First, for small $\lambda$ (Fig. 9), the detectors under slotted and one-bit quantization (i.e., $\delta_i$ vs. $\delta_{in}$ and $\delta_{hi}$ vs. $\delta_{oi}$) have similar performance; as $\lambda$ increases (Fig. 10, 11), the detectors under slotted quantization ($\delta_i$ and $\delta_{hi}$) outperform their counterparts. This observation clearly suggests that which 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 decrease with the increase of $\lambda$. 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 $\Delta$ normalized by the average interarrival time $1/\lambda$) is an increasing function of the traffic rate $\lambda$, 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, $\delta_i$ 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 $\mathcal{H}_0$, $S_1$ and $S_2$ are independent Poisson processes of rate $\lambda$. Under $\mathcal{H}_1$, $S_i = F_i \oplus W_i$ ($i = 1, 2$), where $F_2$ is generated by adding i.i.d. uniform delays in $[0, \Delta]$ to a Poisson process $F_1$ of rate $(1 - f_c)\lambda$ ($f_c \in (0, 1)$), and $W_i$’s are independent Poisson processes (also independent of $F_i$) 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 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 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 provide competitive performance. This work combined with the previous work on centralized detection in [10] establishes a mathematical framework for statistical information flow detection.

$^{15}$Since $\alpha_i^*(\cdot)$ 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 $\lambda$.

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

$^{17}$It can be shown that the quantization results of $S_2$ 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.
The consistency-rate functions of $\delta_1, \ldots, \delta_n$ for various traffic rates: $\Delta = 1, 10^4$ slots.
The overall consistency-rate function is bounded between the largest $\alpha^*_f(\cdot)$ and $1/(1 + \lambda \Delta)$ (not shown).

![Fig. 9. $\lambda = 0.1$.](image)

![Fig. 10. $\lambda = 0.5$.](image)

![Fig. 11. $\lambda = 1$.](image)

![Fig. 12. The ROCs of $\delta_1$ and $\delta_{\text{II}}$ 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, 10000 Monte Carlo runs.](image)

### VIII. 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 $x^n$ be the numbers of matched epochs in $s_1$, and $f_2 = (t_1, t_2, \ldots)$ be the sequence of matched epochs in $s_2$. We construct a sequence $f_1 = (s_j)_{j \geq 1}$ as follows. For an epoch $t_1$ in $f_2$ matched to the same slot, we construct an epoch at $t_1$ in $f_1$; for an epoch $t_2$ matched to the previous slot, we construct an epoch in $f_1$ at the end of that slot. Such construction guarantees that slotted quantization of $f_1$ yields $x^n$, and $(f_1, f_2)$ forms a realization of an information flow.

Then we show that SF-GM is optimal. Since BGM is optimal, it remains to show that our choice of $s_1$ and the delay bound $T + \Delta$ minimizes chaff noise. Given $x_j$ $(j = 1, \ldots, n)$, the maximum interval for the $x_j$ packets to be relayed through $s_2$ is $[(j-1)T, jT + \Delta]$, and SF-GM allows all such matching. Therefore, SF-GM inserts the minimum chaff noise for any $(x^n, s_2)$.

#### B. Proof of Theorem 5.2

It is known that for Poisson processes of maximum rate $\lambda$, the rate $R_i(T)$ suffices to deliver $X^n$ (for large $n$). We only need to show that the consistency level of $\delta_i$ can approximate $\alpha(T)$. By our argument on the miss probability before Theorem 5.2, the desired result holds if we can prove that the false alarm probability decays exponentially for any $\tau_i < \alpha(T)$. It suffices to consider the worst case where $S_i$ $(i = 1, 2)$ have equal rate $\lambda$.

First, for $\beta < \frac{\Delta}{\alpha(T)}$, note that a necessary condition for false alarm is that $C_i/n \leq \sqrt{2\lambda T \alpha_i(T)}$ or $N/n \geq (2\lambda T)/\sqrt{\beta}$. By union bound, we have

$$P_F(\delta_i) \leq \Pr\left\{ \frac{C_i}{n} \leq \sqrt{2\lambda T \alpha_i(T)} \right\} + \Pr\left\{ \frac{N}{n} \geq \frac{2\lambda T}{\sqrt{\beta}} \right\}. $$

By Cramer’s Theorem [19], the second term $\Pr\{N/n \geq 2\lambda T/\sqrt{\beta}\}$ decays exponentially with $n$ for all $\tau_i < \alpha(T)$ ($\beta < 1$). It remains to show the exponential decay of the first term.

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 $\delta_i$ on every other slot, then the numbers of chaff packets $\tilde{C}_2, \tilde{C}_4, \ldots$ will be i.i.d. Obviously, $C_i \geq \sum_{i=1}^{n/2} \tilde{C}_2i$. Thus, we have

$$\Pr\left\{ \frac{C_i}{n} \leq \eta \right\} \leq \Pr\left\{ \frac{1}{n} \sum_{i=1}^{n/2} \tilde{C}_2i \leq 2\eta \right\}. $$

Again by Cramer’s Theorem, the RHS of (16) decays exponentially as long as $2\eta < \mathbb{E}[\tilde{C}_2]$. Since $\tilde{C}_2 \overset{\Delta}{=} (Y(0, T) - X_1)_+ + (X_1 - Y(0, T + \Delta))_+$, where $(x)_+ \overset{\Delta}{=} \max(x, 0)$, 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,

$$\mathbb{E}[\tilde{C}_2] \approx 2\sqrt{\frac{\lambda T}{\pi}} - \frac{1}{2} \lambda \Delta = 4\lambda T \alpha_i(T).$$
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 an information flow. Moreover, by the constraints of causality and bounded delay, packets in slot $j$ can only be matched to packets in slots $j, \ldots, j + \lceil \frac{T}{2} \rceil$, 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 $\Pr \{ C_n/n \leq \eta \}$ decays exponentially for any $\eta < 2\lambda T \alpha_a(T)$ for processes of equal rate $\lambda$. As before, we bound this probability by $\Pr \{ (\frac{1}{n} \sum_{i=1}^{n/2} \tilde{C}_{2i} \leq 2\eta \}$, where $\tilde{C}_{2i}$ is the number of chaff packets in slot $2i$ if we only run $\delta_i$ on even slots. Now that

\[
\mathbb{E}[\tilde{C}_2] = \mathbb{E}[\{Y_2 - X_1 - X_2\}_{+} + \{X_2 - Y_2 - Y_3\}_{+}],
\]

and $(Y_2 - X_1 - X_2, X_2 - Y_2 - Y_3) \sim \mathcal{N}(\lambda T + 3\lambda T)$ for large $T$, we have

\[
\frac{1}{2} \mathbb{E}[\tilde{C}_2] = \sqrt{\frac{3\lambda T}{2\pi}} e^{-\lambda T T/6 - \lambda T} Q \left(\frac{\lambda T}{3}\right) \approx c_1 \lambda T e^{-\lambda T T/6} = \lambda T \alpha_a(\lambda T) + \lambda T \alpha_a(\lambda T + \Delta),
\]

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

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 $s_2$, 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 $\lambda$, the rate $R_m(T)$ suffices to deliver the one-bit quantization results. For consistency results, it suffices to show that $\bar{N}_1$ converges to $\lambda T$ exponentially, and $\Pr \{ C_n/n \leq \eta \} \leq \alpha(T)$ decays exponentially for any $\eta < 2\lambda T \alpha_a(T)$. For the former, since $\bar{X}_k$’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 $\bar{N}_1$. For the latter, by Cramer’s Theorem, it is reduced to showing that $\mathbb{E}[\tilde{C}_{2i}] \geq 4\lambda T \alpha_a(\lambda T)$, where $\tilde{C}_{2i}$ is the number of chaff packets per slot if $\delta_m$ is only run on even slots. If $X_{2i} = 0$ (with probability $e^{-\lambda T}$), then all the epochs in $[(2i - 1)T + \Delta, 2iT]$ in $S_2$ 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 $S_1$ in slot $2i$. Thus,

\[
\mathbb{E}[\tilde{C}_{2i}] \geq e^{-\lambda T}(T - \Delta) + (1 - e^{-\lambda T})e^{-\lambda T(T + \Delta)} \approx 4\lambda T \alpha_a(\lambda T).
\]

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 $k \in \{j, \ldots, j + \lceil \Delta/T \rceil \}$ (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 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 $\mathbb{E}[\tilde{T}_i]$ 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

\[
\left\{ C_n/n \leq \tau \right\} \leq \left\{ \frac{1}{n \tau} \sum_{i=1}^{n \tau} T_i \geq \frac{1}{\tau} \right\}.
\]

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 $\mathbb{E}[\tilde{T}_i] = 2Z$, where $Z$ has the geometric distribution $\Pr \{ Z = n \} = (1 - \rho)^n \rho$ $(n \geq 1)$ 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

\[
A_{2i} \Delta \{X_{2i-1} + X_{2i} < Y_{2i}, \text{ or } X_{2i} > Y_{2i} + Y_{2i+1}\},
\]

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

\[
\Pr \{ C_n/n \leq \tau \} \leq \Pr \left\{ \frac{1}{n \tau} \sum_{i=1}^{n \tau} \tilde{T}_i \geq \frac{1}{\tau} \right\},
\]

where $\tilde{n} = n \tau$. Now that $\mathbb{E}[\tilde{T}_i] = 2\mathbb{E}[Z] = 2/\rho = 1/(2\lambda T \alpha_a(\lambda T))$, by Cramer’s Theorem, $\Pr \{ C_n/n \leq \tau \}$ decays exponentially for any $\tau < 2\lambda T \alpha_a(\lambda T)$.

This result coupled with the exponential convergence of $\bar{N}_1$ $(i = 1, 2)$ (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 $S_1$ and $S_2$ have equal rate $\lambda$.

1) Case I: Let the $k$th interarrival time in $S_2$ be $V_k \equiv S_2(k) - S_2(k-1)$ ($k \geq 1$). Let $Z_j$ ($j \geq 0$) be the starting time for finding matches in the $(j+1)$th slot of $S_2$ ($Z_0 = 0$ by definition). For each $X_j$ ($j \geq 1$), SF-GM matches the $X_j$ reconstructed epochs with epochs in $S_2$. Let $K_j$ ($j \geq 1$) be the index of the last epoch in $S_2$ 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

$$Z_j = \min(\max(Z_{j-1} + V_{K_{j-1}+1}, V_{K_{j-1}+1} - T, 0), \Delta),$$

where $V_{K_{j-1}+1} = S_2(K_{j-1} + 1) - (j-1)T - Z_{j-1}$ is a truncated interarrival time. Since $V_k$’s are i.i.d. exponential random variables, by the memoryless property of exponential distribution, we have that $\{Z_j\}_{j=0}^{\infty}$ is a random walk with reflecting barriers at $0$ and $\Delta$. The steps have the same distribution as $\sum_{k=1}^{X_1} V_k - T$, where $X_1$ is a Poisson variable with mean $\mu T$, $V_k$ ($k \geq 1$) are i.i.d. exponential random 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

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

It is clear that $C_j$’s only depend on each other through $Z_j$. Now that $\{Z_j\}_{j=0}^{\infty}$ is ergodic, $\left(\sum_{j=1}^{n} C_j\right)/n$ 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

$$\alpha^*_{\infty}(T) \equiv \frac{\mathbb{E}[\max(Y(Z, T) - X_1, X_1 - Y(Z, T + \Delta), 0)]}{2\lambda T} \tag{19}$$

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

2) Case II: Consider $T \geq \Delta$ 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 $S_2$ that are matched to slots before $j$ ($Z_0 = 0$ 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 $\left(\sum_{j=1}^{n} C_j\right)/n$ converges a.s., which implies that the CTR converges a.s. The limiting CTR can be computed by

$$\alpha^*_{\infty}(T) \equiv \frac{\mathbb{E}[\max(Y - Z - X_1, X_1 - Y + Z - Y_1, 0)]}{2\lambda T} \tag{20}$$

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 $\lambda T$, and $(Z, Y)$ is independent of $(X_1, Y_1)$.

3) Case III: Let $U_j$ ($j \geq 0$) be the starting time for finding matches in the $(j+1)$th slot in $S_2$ 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

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

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

where $K \equiv \inf\{k : S_2(k) \geq L_j\}$. 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

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

By the ergodicity of $\{(U_j - jT, L_j - jT)\}_{j=0}^{\infty}$ and the homogeneity of 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

$$\alpha^*_{\infty}(T) \equiv \frac{\mathbb{E}[\max(Y(U, T))]e^{-\lambda T} + \Pr\{Y(U, T + \Delta) = 0\}(1 - e^{-\lambda T})}{2\lambda T} \tag{21}$$

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 $\{(X_{j-1}, X_j, Y_{j+1})\}_{j=1}^{\infty}$ (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

$$\alpha^*_{\infty}(T) \equiv \frac{\mathbb{E}[\max(Y - X - X_1, X - Y_1, 0)]}{2\lambda T} \tag{22}$$

where $(X_1, X, Y, Y_1)$ has the limiting distribution of $\{(X_{j-1}, X_j, Y_j, Y_{j+1})\}_{j=1}^{\infty}$.

Proof of Theorem 6.2

The lower bound is by definition. For the upper bound, it suffices to consider Case I because the observations in the other cases are its further compression. We give the following method to mimic distributions under $\mathcal{H}_{0}$. Given realizations
(x^n, s_2) where x^n is the result of slotted quantization, we construct a sequence \( \hat{s}_1 \) as follows. For \( j = 1, \ldots, n \), if \( x_j \leq y((j-1)T, jT) \), then randomly select \( x_j \) epochs from the \( j \)th slot of \( s_2 \); otherwise, select all the epochs in the \( j \)th slot of \( s_2 \) 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 \( \lambda T \) and an independent Poisson process \( S_2 \) of rate \( \lambda ; \)
2) construct a process \( \hat{S}_1 \) as described above;
3) use BGM with delay bound \( \Delta \) to decompose \( (\hat{S}_1, S_2) \) into an information flow and chaff noise.

Traffic generated by this method is equal to traffic under \( H_0 \) in both marginal distributions and the joint distribution after quantization. Moreover, there are at least min(\( X_1, Y((j-1)T, jT) \)) pairs of matched epochs in slot \( j \), i.e., the average number of chaff packets per slot is upper bounded by \( E[|X - Y|] \), where \( X \) and \( Y \) are independent Poisson variables with mean \( \lambda T \). Therefore, the CTR \( E[|X - Y|]/(2\lambda T) \) suffices to mimic \( H_0 \) under the quantization in Case I and thus is an upper bound on the corresponding highest consistency level.

References