Signal Processing 92 (2012) 2577–2593

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Unsupervised and nonparametric detection of information flows

Jinsub Kim*, Lang Tong

School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, United States

ARTICLE INFO

Article history:
Received 22 September 2011
Received in revised form
27 February 2012
Accepted 25 March 2012
Available online 2 April 2012

Keywords:
Nonparametric detection
Point process
Information flow
Traffic analysis
Network surveillance

ABSTRACT

The problem of detecting the presence of possibly bidirectional and time-varying information flows through two nodes in a network is considered. Only the transmission timing measurements are used in the detection. The proposed technique assumes no parametric flow model and requires no training data. The consistency of the detector is established for a class of non-homogeneous Poisson traffic. The proposed detector is tested in a simulation using LBL TCP traces (Paxson and Floyd, 1995 [24]) and an experiment involving MSN VoIP sessions.

1. Introduction

We consider the problem of detecting information flows through a pair of monitored nodes as illustrated in Fig. 1. In particular, given the measurements of transmission timings from the monitored nodes, we are interested in determining whether the two monitored nodes are engaged in relaying packets of certain information flows (the alternative hypothesis), or they are merely transmitting independently (the null hypothesis). The network of our interest can be either wireless or wired as long as transmission timings can be measured.

The generic problem of flow detection arises from a number of practical applications, especially in the context of information forensics, network surveillance, and anonymous networking. For example, in the so-called stepping-stone attack [1] in a network, an adversary may attack a node by compromising a sequence of nodes that serve as stepping stones. When the attacker is involved in an interactive session (*e.g.*, SSH), a flow of packets travel through a chain of stepping stones. By detecting the presence of unexpected flows through monitored nodes, the network owner can alert the possibility of an attack. Other applications include the detection of wormhole attack [2] in which a set of colluding nodes divert a valid network flow through a "wormhole tunnel." Understanding the problem of flow detection is also valuable for the design and assessment of anonymous networks [3,4].

In this paper, we restrict ourselves to the use of timing measurements only. Such a restriction is of course unnecessary because there are often other information available such as source–destination addresses, packet statistics, etc.; a detector should incorporate such side information. We choose to focus exclusively on the use of timing information for two reasons. First, timing can only be distorted but cannot be hidden by the transmitter, and its measurements can be obtained by simple devices. In contrast, source–destination addresses and packet characteristics can be masked using standard techniques in anonymous networking [4]. Second, timing is a fundamental traffic characteristic. It is therefore useful to understand the extent that timing reveals the presence of information flows. Furthermore, any

Abbreviation: BiBGM, bidirectional-bounded-greedy-match; BFD, bidirectional flow detector; ITA, independent traffic approximation; ITAd, ITA-double; NBFD, nonparametric bidirectional flow detector; DAC, detect-attack-chaff

* Corresponding author. Tel.: +1 607 220 3738.
E-mail addresses: jk752@cornell.edu (J. Kim), ltong@ece.cornell.edu (L. Tong).

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.03.015

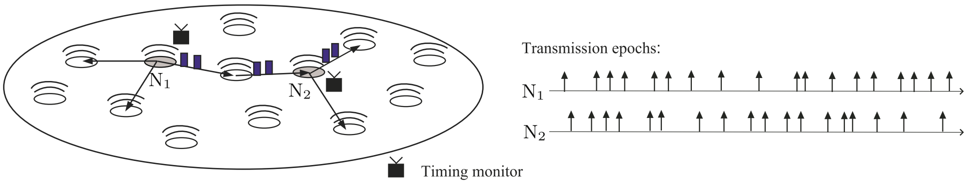

Fig. 1. In the above wireless network, the transmission timings of two nodes, N$_1$ and N$_2$, are recorded. The horizontal axis is the time axis, and arrows represent packet transmissions at different time points. As illustrated, packets of certain information flows may travel through N$_1$ and N$_2$.

side information, when incorporated properly, will enhance the performance of techniques based solely on timing information.

Even though transmission timings of nodes can be easily monitored, detecting information flows based on timing is non-trivial, partly because of non-stationary traffic characteristics: transmission timings of nodes often have time-varying intensities, and they may be bursty when interactive users are involved. Moreover, in general, it is difficult to obtain an accurate parametric model for the monitored traffic, especially when there is no prior knowledge about the nature of the traffic and no training data available. The presence of noise-like epochs is another source of difficulty. When an information flow travels through two nodes, the two nodes may have transmissions that do not belong to the flow. They may multiplex transmissions of other flows that go through only one of the two nodes, or intentionally superpose dummy transmissions to avoid detection. We refer to the epochs of such transmissions as *chaff* epochs.

It is easy to see that, if a node can arbitrarily delay packets in a flow, timing information is insufficient for detection. For latency-sensitive applications such as VoIP, multimedia streaming, etc., however, packets must satisfy certain end-to-end delay constraints, which make the presence of such flows detectable. For instance, VoIP applications require end-to-end delays to be bounded above by 150 ms [5]. This paper will consider the constraint that flow packets should satisfy the end-to-end delay constraint of Δ seconds.

1.1. Related works

Detection of information flows has been studied in the context of intrusion detection, especially in the detection of interactive stepping-stone attacks [1]. The use of timing only measurement for detection is motivated by the fact that packets involved in an attack can be easily encrypted. Donoho et al. [1] were among the first to consider the flow model with a uniform delay bound. Following their model, many algorithms have been proposed to detect a flow with a delay constraint. As an *active* detection scheme, Wang et al. [6] proposed a watermark-based detector which embeds watermarks by slightly adjusting transmission timings of a node; if the same watermarks are detected in another node, two nodes are claimed to have flows between them. Their work was followed by a large number of watermark-based detectors [7–14]. The insertion of watermarks, however, requires the ability of the detector to modify traffic at different locations of the network, which may not be possible in practical situations.

If the network traffic cannot be modified to facilitate detection, the problem is referred to as *passive* flow detection. Zhang et al. [15,16] proposed matching-based algorithms. However, they assumed that only one of two nodes can insert chaff transmissions, and their algorithms are vulnerable to chaff insertion at both nodes. Donoho et al. [1] proposed a wavelet analysis with a claim that it can detect a flow in chaff if the chaff part is independent of the flow part and the sample size is sufficiently large. Blum et al. [17] presented a counting-based method which was shown to be able to detect a flow in chaff if the fraction of chaff is small enough. Under the Poisson traffic assumption, they characterized the sufficient sample size for satisfying a given false alarm probability constraint. However, their method may result in high miss detection probability if chaff transmissions are bursty. He and Tong [18] proposed a matching-based detector with better chaff tolerance and characterized the maximum tolerable fraction of chaff under the homogeneous Poisson traffic assumption. Their approach requires choosing a detection threshold which is a function of the parameter of the underlying Poisson traffic. When the traffic deviates from the Poisson model, the detection algorithm is not always robust. The approach in [18] can be applied to the general traffic if a training data with a sufficiently long time span is available. Coskun and Memon [19,20] presented detectors based on random projection of transmission processes. Similar to [18], their methods also require choosing an appropriate detection threshold, which can be successful only if a large volume of training data or an accurate parametric model is available.

Statistical inference on timing measurements has been studied in various other fields. In communication via timing channels [21,22], a transmitter embeds a message into its packet transmission timings, and a receiver infers the message based on its packet arrival timings. In neuroscience, spike train observations of neurons form sequences of timings, and they are analyzed to infer causal relations among neurons [23].

1.2. Summary of results and organization

The main results of this paper include three parts: a nonparametric flow detection algorithm for unidirectional or bidirectional flows, the related performance analysis, and experiments with synthetic and real data. In developing an algorithm, our main contribution is a new nonparametric technique that does not rely on knowledge of traffic distribution; nor does it require a training data for either hypothesis. The key idea lies in a particular transformation of the measurements that leads to distinct statistical behaviors under two different hypotheses. The

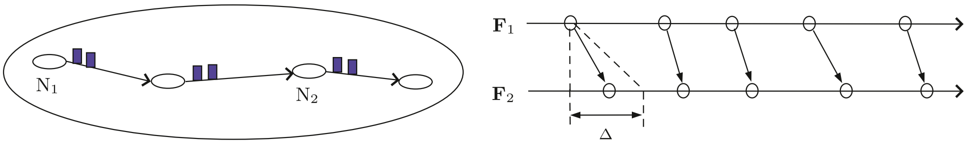

Fig. 2. Every packet transmission of a unidirectional flow is assumed to satisfy packet conservation, causality, and the delay constraint Δ.

proposed detector does not assume stationarity of traffic and hence is applicable in time-varying traffic conditions. Furthermore, it is memory-efficient and has linear computational complexity with respect to the sample size thereby making real-time inference feasible.

In algorithm analysis, we aim to give theoretical justifications for the proposed approach. To this end, we establish the consistency property of the proposed detector for a class of non-homogeneous Poisson traffic. Even though the detector is analyzed only for non-homogeneous Poisson traffic, the intuition behind it suggests that it may perform well on the traffic with more general distribution.

The performance of our detector is evaluated using synthetic Poisson traffic, LBL TCP traces [24], and real-world measurements from MSN VoIP sessions, and comparison with other passive detectors is provided. The use of synthetic data allows us to examine the trade-offs between miss detection and false alarm probabilities using Monte Carlo simulations. LBL TCP traces and MSN VoIP traces are of course not guaranteed to satisfy the assumptions made in our algorithm analysis, and our results indicate a level of robustness.

The rest of the paper is organized as follows. Section 2 gives the notations and definitions employed throughout the paper and formulates flow detection as a binary composite hypothesis testing problem. In Section 3, we consider the simpler case where the parametric model of the traffic is available. Then, Section 4 presents a nonparametric flow detection algorithm and its consistency property. Theorems are stated with proofs presented in the Appendix. In Section 5, the proposed detector is evaluated using synthetic Poisson traffic, LBL TCP traces, and MSN VoIP traffic. Finally, Section 6 concludes the paper with remarks.

2. Mathematical formulation

This section introduces notations and definitions and formulates flow detection as one of binary composite hypothesis testing.

2.1. Notations and flow models

Transmission timings of each node are modeled as a point process on $[0,\infty)$, and detectors begin recording the timings at time 0. Bold upper-case letters (*e.g.*, $\mathbf{S}$) denote point processes, and bold lower-case letters (*e.g.*, $\mathbf{s}$) denote their realizations. $S(i)$ represents the ith epoch (*i.e.*, the time of the ith transmission) of $\mathbf{S}$, and $s(i)$ is its realization. The upper-case script letter $\mathcal{S}$ denotes the set of epochs in the realization $\mathbf{s}$: $\mathcal{S} \triangleq \{s(i),\ i \geq 1\}$. In addition, we define a *superposition* operator $\oplus$: given two increasing sequences $(a_i)_{i=1}^{\infty}$ and $(b_i)_{i=1}^{\infty}$, $(a_i)_{i=1}^{\infty} \oplus (b_i)_{i=1}^{\infty} = (c_i)_{i=1}^{\infty}$, where c_i is the ith element of the sequence of all the elements of $(a_i)_{i=1}^{\infty}$ and $(b_i)_{i=1}^{\infty}$ ordered in the increasing order.

First, we define a *unidirectional flow* as follows.

Definition 2.1. An ordered pair of point processes $(\mathbf{F}_1,\mathbf{F}_2)$ forms a *unidirectional flow*, if for any realization $(\mathbf{f}_1,\mathbf{f}_2)$ there exists a bijection $g : \mathcal{F}_1 \to \mathcal{F}_2$ satisfying $g(s)-s \in [0,\Delta]$ for all $s \in \mathcal{F}_1$.

As illustrated in Fig. 2, when packets of an information flow travel through node N_1 and node N_2, $\mathbf{F}_1$ and $\mathbf{F}_2$ can be interpreted as the transmission timings of the flow packets at N_1 and N_2 respectively. The bijection condition of g means packet conservation; every flow packet sent by N_1 is received and forwarded by N_2. The condition $g(s)-s \in [0,\Delta]$ means that every flow packet transmission satisfies causality and the delay constraint Δ. Based on the above definition, we define a *bidirectional flow* as a superposition of two unidirectional flows with opposite directions.

Definition 2.2. A pair of point processes $(\mathbf{F}_1,\mathbf{F}_2)$ forms a *bidirectional flow*, if $\mathbf{F}_i$ can be decomposed into $\mathbf{F}_i^{12}$ and $\mathbf{F}_i^{21}$ (*i.e.*, $\mathbf{F}_i = \mathbf{F}_i^{12} \oplus \mathbf{F}_i^{21}$) such that $(\mathbf{F}_1^{12},\mathbf{F}_2^{12})$ and $(\mathbf{F}_2^{21},\mathbf{F}_1^{21})$ are unidirectional flows.

We allow $(\mathbf{F}_1^{12},\mathbf{F}_2^{12})$ and $(\mathbf{F}_2^{21},\mathbf{F}_1^{21})$ to have zero rate, so that a unidirectional flow is a special case of a bidirectional flow.

2.2. Problem statement

We formulate detection of bidirectional flow as a binary composite hypothesis testing problem. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ denote the transmission processes of N_1 and N_2, respectively. Given the measurements $(\mathbf{s}_i)_{i=1}^{2}$ in $[0,t]$, we test the following hypotheses:

$\mathcal{H}_0$: $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent;

$\mathcal{H}_1$: $\mathbf{S}_i = \mathbf{F}_i \oplus \mathbf{W}_i$, $i=1,2$, and
$(\mathbf{F}_1,\mathbf{F}_2)$ forms a bidirectional flow.

We further assume that, under $\mathcal{H}_1$,

1. $\mathbf{F}_1$ and $\mathbf{F}_2$ are point processes with non-zero rates.[1]
2. $\mathbf{F}_1$ and $\mathbf{F}_2$ are not independent.
3. $(\mathbf{F}_1,\mathbf{F}_2)$, $\mathbf{W}_1$, and $\mathbf{W}_2$ are independent.

$\mathcal{H}_0$ corresponds to the scenario that N_1 and N_2 have independent transmissions. $\mathcal{H}_1$ corresponds to the scenario that N_1 and N_2 relay packets of information flows in either or both directions: $(\mathbf{F}_i)_{i=1}^{2}$ and $(\mathbf{W}_i)_{i=1}^{2}$ represent the flow part and the chaff part, respectively. Note that

[1] In other words, if $N_{f_i}(t)$ denotes the number of epochs of $\mathbf{F}_i$ in $[0,t]$, there exists $\delta > 0$ such that $\lim\inf_{t\to\infty} N_{f_i}(t)/t \geq \delta$ almost surely, $i=1,2$.

under both hypotheses, no restriction is imposed on the marginal distributions of $\mathbf{S}_i$, $\mathbf{F}_i$, and $\mathbf{W}_i$.

The assumptions under $\mathcal{H}_1$ are imposed to make two hypotheses disjoint. The first assumption implies that the bidirectional flow should have positive rate. The second assumption means that the flow parts of N_1 and N_2 should not be independent, and this assumption is expected to hold in general due to the delay constraint Δ. The third assumption implies that the chaff parts of N_1 and N_2 are independent, and they are also independent of the flow part. We note here that the third assumption is more restrictive than that used in earlier works [17,18].

We employ the notion of Chernoff consistency [25] to evaluate the asymptotic performance of detectors.

Definition 2.3. For $j=0,1$, $\mathcal{P}_j$ denotes the set of all possible distributions of $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_j$. A detector $\delta((\mathbf{s}_i)_{i=1}^2,t)$ is a function of the epochs of $(\mathbf{s}_i)_{i=1}^2$ in $[0,t]$, which is equal to j if the decision is $\mathcal{H}_j$. $\delta((\mathbf{s}_i)_{i=1}^2,t)$ is said to be *consistent* if

1. $\forall\ Q_0 \in \mathcal{P}_0$, $\lim_{t\to\infty} Q_0(\delta((\mathbf{S}_i)_{i=1}^2,t)=1)=0$, and
2. $\forall Q_1 \in \mathcal{P}_1$, $\lim_{t\to\infty} Q_1(\delta((\mathbf{S}_i)_{i=1}^2,t)=0)=0$.

In other words, a detector is consistent if its false alarm and miss detection probabilities vanish as t grows under all possible distributions in $\mathcal{P}_0$ and $\mathcal{P}_1$. In the following sections, we will reduce $\mathcal{P}_0$ and $\mathcal{P}_1$ to the sets of distributions satisfying certain additional conditions, and prove the consistency of our detection algorithms.

Intuitively, the greater the amount of chaff epochs, the harder the flow detection becomes. To measure the relative strength of the flow part with respect to the chaff part, we introduce the following definition of *flow fraction*.

Definition 2.4. Under $\mathcal{H}_1$, suppose that $(\mathbf{S}_i)_{i=1}^2$ consists of the bidirectional flow $(\mathbf{F}_i)_{i=1}^2$ and the chaff part $(\mathbf{W}_i)_{i=1}^2$. Given a realization $(\mathbf{s}_i)_{i=1}^2$, where $\mathbf{s}_i=\mathbf{f}_i\oplus\mathbf{w}_i$, $i=1,2$, the *flow fraction* of $(\mathbf{s}_i)_{i=1}^2$ is defined as

$$R(t) \triangleq \frac{\sum_{i=1}^2 |\mathcal{F}_i\cap[0,t]|}{\sum_{i=1}^2 |\mathcal{S}_i\cap[0,t]|},\quad R \triangleq \liminf_{t\to\infty} R(t) \tag{1}$$

where $|\mathcal{F}_i\cap[0,t]|$ is the number of flow packet transmissions at N_i in $[0,t]$, and $|\mathcal{S}_i\cap[0,t]|$ is the number of total transmissions at N_i in $[0,t]$.

In other words, $R(t)$ is the fraction of the flow epochs in the measurements up to time t, and R is its limiting value.

3. Parametric flow detection

We begin with an easier case where an accurate parametric model for traffic is available. The main result in this section is a simple algorithm that computes, for measurements $(\mathbf{s}_i)_{i=1}^2$, the maximum schedulable flow fraction ($\overline{R}$) as our decision statistic. The flow detection algorithm is a threshold decision rule based on $\overline{R}$. The computation of the threshold, however, requires the knowledge of the traffic distribution under $\mathcal{H}_0$, which we assume available at the moment; this assumption is removed in Section 4.

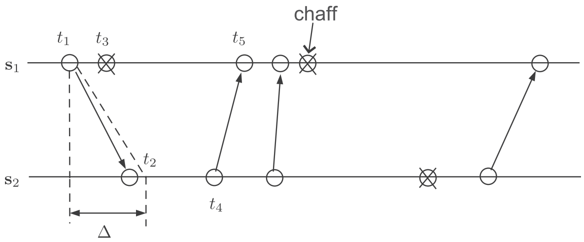

Fig. 3. Bidirectional-bounded-greedy-match.

Table 1
Bidirectional-bounded-greedy-match.

```
BiBGM ($\mathbf{s}_1$,$\mathbf{s}_2$,$\Delta$):
    $m=n=1$;
    while $m\le|\mathcal{S}_1|$ and $n\le|\mathcal{S}_2|$
     if $s_2(n)<s_1(m)-\Delta$
       $s_2(n)$ is chaff; $n\leftarrow n+1$;
     else if $s_2(n)>s_1(m)+\Delta$
       $s_1(m)$ is chaff; $m\leftarrow m+1$;
    else
       match $s_1(m)$ with $s_2(n)$; $m\leftarrow m+1$; $n\leftarrow n+1$;
     end
    end
    return $\frac{|\{\text{Matched epochs}\}|}{|\mathcal{S}_1|+|\mathcal{S}_2|}$
```

3.1. Decision statistic: maximum schedulable flow fraction

Under both hypotheses, given a realization $(\mathbf{s}_i)_{i=1}^2$ in $[0,t]$, its *maximum schedulable flow fraction* $\overline{R}(t)$ is defined as

$$\overline{R}(t) \triangleq \max_{\substack{(\mathbf{f}_i,\mathbf{w}_i)_{i=1}^2:\\ \mathbf{s}_i=\mathbf{f}_i\oplus\mathbf{w}_i\sim\mathcal{H}_1}} \frac{\sum_{i=1}^2 |\mathcal{F}_i\cap[0,t]|}{\sum_{i=1}^2 |\mathcal{S}_i\cap[0,t]|}$$

where $\mathbf{s}_i=\mathbf{f}_i\oplus\mathbf{w}_i\sim\mathcal{H}_1$ denotes the constraint that $\mathbf{s}_i=\mathbf{f}_i\oplus\mathbf{w}_i$, $i=1,2$, and $(\mathbf{f}_1,\mathbf{f}_2)$ is a realization[2] of a bidirectional flow. In other words, we schedule a maximum number of bidirectional flow transmissions between $\mathbf{s}_1$ and $\mathbf{s}_2$ in $[0,t]$, and denote the fraction of the flow part by $\overline{R}(t)$.

To effectively evaluate $\overline{R}(t)$, we propose a matching algorithm called Bidirectional-Bounded-Greedy-Match (BiBGM). To achieve its goal, BiBGM starts with the first epoch in $\mathcal{S}_1\cup\mathcal{S}_2$, and subsequently finds the earliest one-to-one matches satisfying causality and the delay constraint. We explain below the operation of BiBGM using an example in Fig. 3 accompanied by a pseudocode implementation in Table 1:

1. At the beginning, all the epochs in $\mathcal{S}_1\cup\mathcal{S}_2$ are unmatched. Start with the earliest epoch in $\mathcal{S}_1\cup\mathcal{S}_2$, and go to **MATCH** to find its match.
2. **MATCH**: Let t denote the epoch for which we want to find a match. For $i=1,2$, if $t\in\mathcal{S}_i$, search for the earliest unmatched epoch in $[t,t+\Delta]\cap\mathcal{S}_{(3-i)}$ and match it with

[2] In other words, $\mathbf{f}_i$ can be partitioned into two subsequences $\mathbf{f}_i^{12}$ and $\mathbf{f}_i^{21}$ such that there exist bijections $g_1:\mathcal{F}_1^{12}\to\mathcal{F}_2^{12}$ and $g_2:\mathcal{F}_2^{21}\to\mathcal{F}_1^{21}$ satisfying $g_1(s)-s\in[0,\Delta]$, $\forall s\in\mathcal{F}_1^{12}$ and $g_2(s)-s\in[0,\Delta]$, $\forall s\in\mathcal{F}_2^{21}$.

t; if there is no unmatched epoch in the interval, label t as chaff (an epoch is said to be *checked* if it is either matched with another epoch or labeled as chaff). Go to **MOVE**.

3. **MOVE**: If every epoch in $\mathcal{S}_1 \cup \mathcal{S}_2$ is checked, terminate. Otherwise, move to the next unchecked epoch in $\mathcal{S}_1 \cup \mathcal{S}_2$ and go to **MATCH** to find its match.

For the example in Fig. 3, BiBGM starts with t_1. Since $t_1 \in \mathcal{S}_1$, we search for the earliest unmatched epoch in $[t_1, t_1+\Delta] \cap \mathcal{S}_2$, which is t_2. Hence, t_1 is matched with t_2. Then, we move to the next unchecked epoch, t_3 of $\mathcal{S}_1$. Because t_2 is the only epoch in $[t_3, t_3+\Delta] \cap \mathcal{S}_2$ and it is already matched with t_1, we label t_3 as chaff. Next, we move to the next unchecked epoch (t_4 of $\mathcal{S}_2$) and searches for the earliest unmatched epoch in $[t_4, t_4+\Delta] \cap \mathcal{S}_1$. BiBGM continues until the last epoch of $\mathcal{S}_1 \cup \mathcal{S}_2$ is checked.

From Table 1, it can be easily seen that BiBGM has linear computational complexity with respect to the sample size (*i.e.*, the total number of observed epochs). The following theorem states that BiBGM indeed achieves the optimal scheduling such that the flow part is maximized.

Theorem 3.1. *Suppose we run BiBGM on* $(\mathbf{s}_i)_{i=1}^2$ *in* $[0,t]$. *Then, the fraction of the matched epochs is equal to* $\overline{R}(t)$.

Proof. See Appendix.

3.2. Parametric flow detection under Poisson models

In this section, we assume the knowledge of the underlying parametric model for transmission processes and propose a detection algorithm called Bidirectional Flow Detector (BFD). BFD is a threshold decision rule based on $\overline{R}(t)$. Specifically, BFD with a threshold τ takes the following form:

$$\begin{cases} \text{If } \overline{R}(t) \geq \tau & \text{declare } \mathcal{H}_1 \\ \text{otherwise} & \text{declare } \mathcal{H}_0 \end{cases}$$

If $(\mathbf{S}_i)_{i=1}^2$ contains a bidirectional flow, $\overline{R}(t)$ is, by definition, an upper bound on $R(t)$ and will tend to be greater compared to the case that $(\mathbf{S}_i)_{i=1}^2$ is an independent pair; this is the intuition behind declaring $\mathcal{H}_1$ when $\overline{R}(t)$ is greater than τ.

Under $\mathcal{H}_1$, since $\overline{R}(t) \geq R(t)$, BFD with τ can detect any flow with $R(t) \geq \tau$, and a smaller τ makes BFD capable of detecting a larger set of flows. However, a smaller τ results in a higher false alarm probability. Hence, there exists a trade-off between the detectability of BFD and its false alarm probability, and we need to consult the parametric model for $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_0$ to find out how small τ should be. Specifically, if under $\mathcal{H}_0$, as t increases $\overline{R}(t)$ converges to or stays close to a certain constant τ_0 with high probability, we can set τ slightly greater than τ_0 and make the false alarm probability become negligible as t grows. For homogeneous Poisson traffic, the following convergence result gives a guidance for setting τ.

Theorem 3.2. *Under* $\mathcal{H}_0$, *if* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are homogeneous Poisson processes with rates* λ_1 *and* λ_2 *respectively, then as t grows to infinity,* $\overline{R}(t)$ *converges almost surely* (*a.s.*) *to*

$$\phi_{(\lambda_1,\lambda_2)} = \begin{cases} \dfrac{2\lambda_1\lambda_2(1-e^{2\Delta(\lambda_1-\lambda_2)})}{(\lambda_1+\lambda_2)(\lambda_2-\lambda_1 e^{2\Delta(\lambda_1-\lambda_2)})} & \textit{if } \lambda_1 \neq \lambda_2 \\ \dfrac{2\lambda\Delta}{1+2\lambda\Delta} & \textit{if } \lambda_1 = \lambda_2 = \lambda \end{cases}$$

Proof. See Appendix.

Especially, if $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_0$ and $(\mathbf{W}_i)_{i=1}^2$ under $\mathcal{H}_1$ are homogeneous Poisson processes, the following theorem states that any bidirectional flow with a positive rate is detectable regardless of the amount of chaff epochs.

Theorem 3.3. *Suppose that* (i) *under* $\mathcal{H}_0$, $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are homogeneous Poisson processes,* (ii) *under* $\mathcal{H}_1$, $\mathbf{W}_1$ *and* $\mathbf{W}_2$ *are homogeneous Poisson processes, and* (iii) *under both hypotheses, the rates*[3] *of* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are* λ_1 *and* λ_2, *respectively. Then, for any* $\eta \in (0,1)$, *there exists a proper threshold* τ, *such that BFD with* τ *can consistently detect*[4] *any bidirectional flow with* $R \geq \eta$ *a.s., with the false alarm probability decaying exponentially fast as the sample size grows. Especially, for* $\eta \in (0, 2\min\{\lambda_1,\lambda_2\}/(\lambda_1+\lambda_2))$, *the following* τ *can be used*:

$$\begin{cases} \dfrac{2\lambda_1 - 2\lambda_2 \frac{\lambda_1(4-\eta)-\lambda_2\eta}{\lambda_2(4-\eta)-\lambda_1\eta} e^{2\Delta(\lambda_1-\lambda_2)}}{(\lambda_2+\lambda_1)\left(1-\frac{\lambda_1(4-\eta)-\lambda_2\eta}{\lambda_2(4-\eta)-\lambda_1\eta} e^{2\Delta(\lambda_1-\lambda_2)}\right)} & \textit{if } \lambda_1 \neq \lambda_2, \\ \dfrac{\eta+2\lambda(2-\eta)\Delta}{2+2\lambda(2-\eta)\Delta} & \textit{if } \lambda_1 = \lambda_2 = \lambda. \end{cases}$$

Proof. See Appendix.

It can be shown that the suggested τ in Theorem 3.3 is a strictly increasing function of η, and as η decreases to 0, it decreases to $\phi_{(\lambda_1,\lambda_2)}$ in Theorem 3.2. This means that to detect a larger set of flows, τ should be closer to $\lim_{t\to\infty}\overline{R}(t)$ under $\mathcal{H}_0$.

Instead of the knowledge of the parametric model, training data can also be used to set τ. If a large set of different realizations of $\mathcal{H}_0$ traffic is available, we can run BiBGM over each realization in the training data set, estimate the statistical behavior of $\overline{R}(t)$ under $\mathcal{H}_0$, and set τ such that the probability that $\overline{R}(t) \geq \tau$ under $\mathcal{H}_0$ (*i.e.*, false alarm probability) becomes reasonably small as t grows. However, if neither a parametric model nor training data is available, it is non-trivial how to determine an appropriate τ; this is the case in many practical applications.

4. Nonparametric flow detection

In this section, we assume that neither a parametric model nor a training data set is available, and present a novel nonparametric flow detector.

[3] By rates, we mean that $\lim_{t\to\infty} N_i(t)/t = \lambda_i$ a.s., where $N_i(t)$ denotes the number of epochs of $\mathbf{S}_i$ in $[0,t]$.

[4] In other words, if the distributions of $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ satisfy (i)–(iii), and $R \geq \eta$ a.s., then under all those distributions, the false alarm and miss detection probability vanish as t increases (as in Definition 2.3).

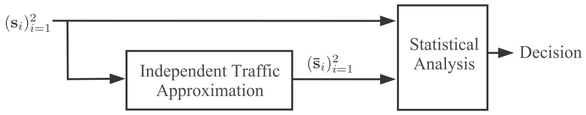
$(\mathbf{s}_i)_{i=1}^2$
Independent Traffic Approximation
$(\bar{\mathbf{s}}_i)_{i=1}^2$
Statistical Analysis
Decision

Fig. 4. The structure of our nonparametric detection algorithm.

4.1. Algorithm structure

We begin by introducing the structure and the main intuition of our detection algorithm. Fig. 4 is describing its structure. A key component of our algorithm is a transformation of measurements $(\mathbf{s}_i)_{i=1}^2$, which we refer to as Independent Traffic Approximation (ITA). As the name suggests, ITA produces an approximately independent pair of transmission processes $(\bar{\mathbf{s}}_i)_{i=1}^2$ such that $\bar{\mathbf{s}}_i$ has similar traffic characteristics (*e.g.*, normalized intensity function,[5] interarrival distribution) with $\mathbf{s}_i$. After $(\bar{\mathbf{s}}_i)_{i=1}^2$ is generated, we compare the statistical characteristics of $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$. If the true hypothesis is $\mathcal{H}_0$, both $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$ are independent pairs with similar traffic characteristics. On the other hand, if $\mathcal{H}_1$ is true, $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$ have similar traffic characteristics, but $(\mathbf{s}_i)_{i=1}^2$ is a correlated pair containing a flow while $(\bar{\mathbf{s}}_i)_{i=1}^2$ approximates an independent pair. Thus, we attempt to infer the true hypothesis by exploiting the gap between the statistical characteristics of $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$: the larger the gap, the more the probable $\mathcal{H}_1$ is.

4.2. Nonparametric bidirectional flow detector

This section presents our detection algorithm, referred to as Nonparametric Bidirectional Flow Detector (NBFD). Here, we simply assume that ITA generates an output $(\bar{\mathbf{s}}_i)_{i=1}^2$ with desired properties: (i) $(\bar{\mathbf{s}}_i)_{i=1}^2$ approximates an independent pair of transmission processes, and (ii) its normalized intensity function and interarrival distribution resemble that of $(\mathbf{s}_i)_{i=1}^2$. The detail about ITA is delayed to the next section, and here we focus on the operation of NBFD.

As described in Fig. 4, NBFD first runs ITA on $(\mathbf{s}_i)_{i=1}^2$ to generate $(\bar{\mathbf{s}}_i)_{i=1}^2$. The next step is to compare the statistical characteristics of $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$. It was shown in Theorem 3.3, although stated under the homogeneous Poisson traffic assumption, that the maximum schedulable flow fraction $\overline{R}(t)$ can be effectively used to distinguish whether the measurements are from a flow-containing pair or an independent pair. Moreover, $\overline{R}(t)$ can be easily evaluated by running BiBGM; hence, NBFD employs $\overline{R}(t)$. NBFD runs BiBGM separately on $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$ and compares the fractions of the matched epochs in the two cases, denoted by $\overline{R}(t)$ and $\overline{\tau}(t)$ respectively. If the true hypothesis is $\mathcal{H}_0$, both $(\mathbf{S}_i)_{i=1}^2$ and $(\overline{\mathbf{S}}_i)_{i=1}^2$ are independent pairs, and they have similar normalized intensity functions and interarrival distributions; this implies that $\overline{R}(t)$ and $\overline{\tau}(t)$ are expected to be close under $\mathcal{H}_0$. On the other hand, when $\mathcal{H}_1$ is true, $(\mathbf{S}_i)_{i=1}^2$ and $(\overline{\mathbf{S}}_i)_{i=1}^2$ have similar normalized intensity functions and interarrival distributions, but $(\mathbf{S}_i)_{i=1}^2$ contains a flow while $(\overline{\mathbf{S}}_i)_{i=1}^2$ approximates an independent pair; hence, $\overline{R}(t)$ is expected to be greater than $\overline{\tau}(t)$. Based on the above intuition, given $(\mathbf{s}_i)_{i=1}^2$ in $[0,t]$, NBFD with ϵ works as follows:

1. Run ITA on $(\mathbf{s}_i)_{i=1}^2$ in $[0,t]$ to generate $(\bar{\mathbf{s}}_i)_{i=1}^2$.
2. Run BiBGM on $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$: $\overline{R}(t)$ and $\overline{\tau}(t)$ denote the fractions of the matched epochs for $(\mathbf{s}_i)_{i=1}^2$ and $(\bar{\mathbf{s}}_i)_{i=1}^2$ respectively.
3. If $\overline{R}(t) \geq \overline{\tau}(t)+\epsilon$, declare $\mathcal{H}_1$; otherwise, declare $\mathcal{H}_0$.

where ϵ is a positive number added to $\overline{\tau}(t)$ to allow small difference between $\overline{R}(t)$ and $\overline{\tau}(t)$ under $\mathcal{H}_0$. $\overline{\tau}(t)$ can also be seen as an estimate of what $\overline{R}(t)$ would be under $\mathcal{H}_0$. Therefore, recalling the discussion of setting τ of BFD in Section 3.2, NBFD can be alternatively interpreted as BFD with a measurement-dependent threshold $\overline{\tau}(t)+\epsilon$.

It is evident from the form of NBFD that a smaller ϵ will lead to the decrease in the miss detection probability. However, the decrease in ϵ will increase the false alarm probability. Because of the trade-off regarding the choice of ϵ and the nonparametric characteristic of our problem, it is difficult to claim that certain ϵ value is the best choice. The experimental results in Section 5 suggest that setting $\epsilon \approx 0.05$ generally results in satisfactory performance.

4.3. Independent traffic approximation

In this section, we present how ITA approximates an independent pair of transmission processes that has the similar normalized intensity function and interarrival distribution with $(\mathbf{S}_i)_{i=1}^2$.

Fig. 5 is illustrating the operation of ITA. ITA has two parameters: the sampling window width w and the gap α $(\alpha \geq \Delta)$ between neighboring sampling windows. As described in Fig. 5, ITA samples the epochs in the w-second windows separated by α-second gaps, shifts them properly, and assembles them to approximate independent traffic. The intuition behind ITA is that if the gap α between two sampling windows is sufficiently large, the epochs in different windows will tend to be approximately uncorrelated. Note that when $(\mathbf{S}_i)_{i=1}^2$ contains a bidirectional flow, ITA disassembles the flow part and significantly reduces the flow-induced correlation. In addition, since we use a sequence of sampled intervals of $\mathbf{S}_i$ for generating $\overline{\mathbf{S}}_i$, $\overline{\mathbf{S}}_i$ and $\mathbf{S}_i$ are expected to share some common characteristics.

To illustrate how the normalized intensities of $\mathbf{S}_i$ and $\overline{\mathbf{S}}_i$ are related, Fig. 6 describes the intensity change of $(\mathbf{S}_i)_{i=1}^2$ and $(\overline{\mathbf{S}}_i)_{i=1}^2$ for the case that $\mathbf{S}_1$ and $\mathbf{S}_2$ are non-homogeneous Poisson processes with two possible intensity levels. As observed in Fig. 6, if the average time that the intensity of $\mathbf{S}_i$ stays in one level is much longer than $2(w+\alpha)$ seconds, the normalized intensity function of $\overline{\mathbf{S}}_i$ is similar to that of $\mathbf{S}_i$. About interarrival distribution, if w is sufficiently large so that a w-second sampling window is

[5] The normalized intensity function of $\mathbf{S}_i$ represents the overall trend of its intensity change in the whole observation interval $[0,t]$. More specifically, assuming the existence of $\lambda(x) \triangleq \lim_{\delta \to 0+} \mathbb{E}\{N_i[x,x+\delta)\}/\delta$ for all $x \in [0,t]$, where $N_i[a,b)$ denotes the number of $\mathbf{S}_i$ epochs in $[a,b)$, the normalized intensity function of $\mathbf{S}_i$ is defined as the time-scaled intensity function $\overline{\lambda}(x) \triangleq \lambda(tx), x \in [0,1]$.

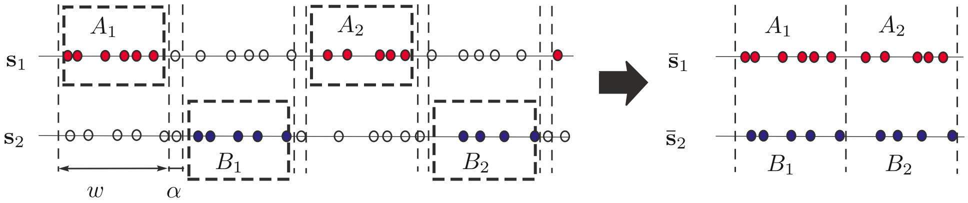

Fig. 5. ITA samples w-second intervals $\{A_1,A_2,\ldots\}$ and $\{B_1,B_2,\ldots\}$ from $(\mathbf{s}_i)_{i=1}^2$ and assemble them to generate $(\bar{\mathbf{s}}_i)_{i=1}^2$.

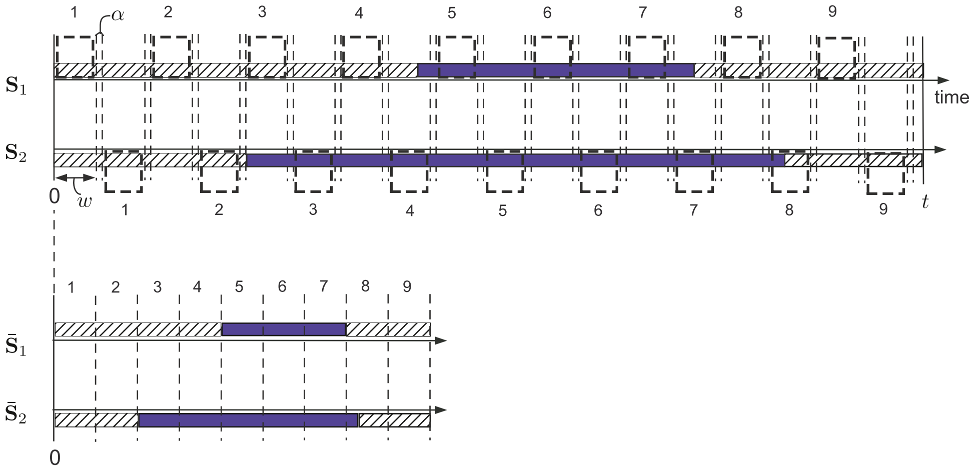

Fig. 6. $\mathbf{S}_1$ and $\mathbf{S}_2$ are non-homogeneous Poisson processes, and $\lambda_1(x)$ and $\lambda_2(x)$ denote their local intensities at time x respectively. $\lambda_1(x)$ and $\lambda_2(x)$ can only take values from $\{\mu_1,\mu_2\}$ $(\mu_1 \neq \mu_2)$. The figure describes the intensity change of $\mathbf{S}_1$, $\mathbf{S}_2$, $\bar{\mathbf{S}}_1$, and $\bar{\mathbf{S}}_2$ using two types of bars. The bars filled with slant lines represent the intervals in which $\lambda_i(x)=\mu_1$, and the blue bars represent the intervals in which $\lambda_i(x)=\mu_2$. The numbers above or below the intervals describe the correspondence between the sampled intervals in $\mathbf{S}_i$ and the intervals in $\bar{\mathbf{S}}_i$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

likely to contain a large number of points, the interarrival distribution of $\bar{\mathbf{S}}_i$ will resemble that of $\mathbf{S}_i$. Moreover, if the interarrival distribution of $\mathbf{S}_i$ varies slowly over time, as in the example of Fig. 6, the interarrival distribution of $\bar{\mathbf{S}}_i$ will also change over time with the similar trend, even though the time scale is different due to the sampling procedure of ITA. Note that resampling from the empirical interarrival distributions (*i.e.*, generating i.i.d. interarrival times of $\bar{\mathbf{S}}_i$ from the empirical interarrival distribution of $\mathbf{S}_i$, for $i=1,2$) can also produce an independent pair of point processes. However, unlike $(\bar{\mathbf{S}}_i)_{i=1}^2$ of ITA, when $(\mathbf{S}_i)_{i=1}^2$ is non-stationary, the results of such resampling approaches may have a totally different dynamics from $(\mathbf{S}_i)_{i=1}^2$; they may not capture the patterns of intensity change or interarrival distribution change in $(\mathbf{S}_i)_{i=1}^2$.

Now, we will check whether $(\bar{\mathbf{S}}_i)_{i=1}^2$ can approximate an independent pair. When $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent, it directly follows that $\bar{\mathbf{S}}_1$ and $\bar{\mathbf{S}}_2$ are independent. On the other hand, if $(\mathbf{S}_i)_{i=1}^2$ contains a bidirectional flow, $\bar{\mathbf{S}}_1$ and $\bar{\mathbf{S}}_2$ are not necessarily independent. However, assuming that correlation across time is weak and the gap α is much larger than Δ, the epochs in different windows are expected to be approximately uncorrelated: *i.e.*, in Fig. 5, the epochs in each A_i will be approximately uncorrelated with the epochs in $\bigcup_{j\geq 1} B_j$. This implies that when temporal correlation is weak, $(\bar{\mathbf{S}}_1, \bar{\mathbf{S}}_2)$ is expected to approximate an independent pair. The following example illustrates a case where $(\mathbf{S}_i)_{i=1}^2$ has weak temporal correlation. Suppose $\mathbf{S}_1$ is a Poisson process and $\mathbf{S}_2$ is such that $S_2(i)=S_1(i)+D_i$, $\forall i$, where D_i's are independent random delays bounded by Δ a.s.: *i.e.*, $(\mathbf{S}_i)_{i=1}^2$ is a unidirectional flow with a delay constraint Δ. The memoryless property of Poisson processes implies that epochs in an interval are correlated with epochs in another disjoint interval only if the gap between the two intervals is less than Δ seconds. Hence, if $\alpha \geq \Delta$, epochs in different sampling windows of ITA are independent, implying that $(\bar{\mathbf{S}}_i)_{i=1}^2$ is an independent pair.

Under $\mathcal{H}_0$, NBFD requires $(\bar{\mathbf{S}}_i)_{i=1}^2$ to be an independent pair having the similar traffic characteristics with $(\mathbf{S}_i)_{i=1}^2$, because $\bar{R}(t)$ and $\bar{\tau}(t)$ have to be close under $\mathcal{H}_0$. However, under $\mathcal{H}_1$, NBFD does not necessitate the independence of $\bar{\mathbf{S}}_1$ and $\bar{\mathbf{S}}_2$, even though the independent case is ideal. Under $\mathcal{H}_1$, NBFD wants $\bar{\tau}(t)$ to be less than $\bar{R}(t)$, and this can be achieved by making $(\bar{\mathbf{S}}_i)_{i=1}^2$ very unlikely to contain a flow. Because, as can be inferred from the discussion in Section 3.2, the maximum schedulable flow fraction (*e.g.*, $\bar{R}(t)$ and $\bar{\tau}(t)$ of NBFD) tends to be higher when the measurements come from a flow-containing pair. Note that ITA does make $(\bar{\mathbf{S}}_i)_{i=1}^2$ unlikely to contain a flow by tearing apart the flow part of $(\mathbf{S}_i)_{i=1}^2$ in its sampling procedure.

Given the measurements $(\mathbf{s}_i)_{i=1}^2$ in $[0,t]$, ITA with (w,α) generates $(\overline{\mathbf{s}}_i)_{i=1}^2$ as follows:

1. Initially, $\overline{\mathbf{s}}_1$ and $\overline{\mathbf{s}}_2$ contain no epoch.
2. For $i=0, 1, \dots, \lfloor t/2(w+\alpha)\rfloor-1$:
   (a) Take the epochs of $\mathbf{s}_1$ in $[2i(w+\alpha), 2i(w+\alpha)+w]$, subtract $i(w+2\alpha)$ from the epochs, and add them to $\overline{\mathbf{s}}_1$.
   (b) Take the epochs of $\mathbf{s}_2$ in $[(2i+1)(w+\alpha), (2i+1)(w+\alpha)+w]$, subtract $i(w+2\alpha)+(w+\alpha)$ from the epochs, and add them to $\overline{\mathbf{s}}_2$.

The implementation of ITA is given in Table 2. As can be seen from Table 2, ITA has linear computational complexity with respect to the sample size.

One drawback of ITA is that it throws away more than a half of the measurements during the sampling procedure, thereby restricting the sample size of $(\overline{\mathbf{s}}_i)_{i=1}^2$ to be at most a half of that of $(\mathbf{s}_i)_{i=1}^2$. $(\overline{\mathbf{s}}_i)_{i=1}^2$, together with $(\mathbf{s}_i)_{i=1}^2$, is used to calculate the decision statistic of NBFD, so a large sample size is desirable. Therefore, we suggest a modification of ITA, referred to as ITA-double (ITAd), to double the sample size of $(\overline{\mathbf{s}}_i)_{i=1}^2$. The operation of ITAd is illustrated in Fig. 7. In ITAd, when $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent, so are $\overline{\mathbf{S}}_1$ and $\overline{\mathbf{S}}_2$. However, if $(\mathbf{S}_i)_{i=1}^2$ contains a flow, $(\overline{\mathbf{S}}_i)_{i=1}^2$ is not an independent pair, because the epochs in A_{i+1} and those in B_i are correlated due to the presence of the flow. However, $(\overline{\mathbf{S}}_i)_{i=1}^2$ is a concatenation of w-second intervals, in each of which the epochs of $\overline{\mathbf{S}}_1$ and $\overline{\mathbf{S}}_2$ are approximately uncorrelated. We believe that this property is enough for NBFD to sense the difference in statistical characteristics between $(\mathbf{S}_i)_{i=1}^2$ and $(\overline{\mathbf{S}}_i)_{i=1}^2$ under $\mathcal{H}_1$, especially when w is large. Although we have no analytical proof for the superiority of ITAd over ITA, the use of ITAd in NBFD instead of ITA consistently resulted in a better performance in all our simulations and experiments in Section 5.

4.4. Performance analysis

This section provides the analysis of algorithmic efficiency and consistency of NBFD.

Table 2
Independent traffic approximation.

```
ITA ($\mathbf{s}_1,\mathbf{s}_2,t,w,\alpha$):
  $\overline{\mathbf{s}}_1 \leftarrow ()$; $\overline{\mathbf{s}}_2 \leftarrow ()$; $a_1 \leftarrow ()$; $a_2 \leftarrow ()$; $j=1$; $k=1$;
  for $i=0:1:\left\lfloor \frac{t}{2(w+\alpha)} \right\rfloor -1$
    while $s_1(j) < 2i(w+\alpha)$, $j \leftarrow j+1$; end
    while $s_1(j) \le 2i(w+\alpha)+w$
      $a_1 \leftarrow a_1 \oplus s_1(j)$; $j \leftarrow j+1$;
    end
    while $s_2(k) < (2i+1)(w+\alpha)$, $k \leftarrow k+1$; end
    while $s_2(k) \le (2i+1)(w+\alpha)+w$
      $a_2 \leftarrow a_2 \oplus s_2(k)$; $k \leftarrow k+1$;
    end
    $a_1 \leftarrow a_1 - i(w+2\alpha)$; $\overline{\mathbf{s}}_1 \leftarrow \overline{\mathbf{s}}_1 \oplus a_1$;
    $a_2 \leftarrow a_2 - (i(w+2\alpha)+w+\alpha)$; $\overline{\mathbf{s}}_2 \leftarrow \overline{\mathbf{s}}_2 \oplus a_2$;
    $a_1 \leftarrow ()$; $a_2 \leftarrow ()$;
  end
  return $(\overline{\mathbf{s}}_i)_{i=1}^2$.
```

∗ For a sequence $(x_i)_{i\ge 1}$ and a real number r, $(x_i)_{i\ge 1} - r \triangleq (y_i)_{i\ge 1}$ where $y_i = x_i - r$, $\forall i$.

NBFD is efficient in terms of computation and memory requirement. Because its main components, ITA and BiBGM, have linear complexity, NBFD also has linear computational complexity with respect to the sample size. In addition, assuming that NBFD with (w,α,ϵ) is executed in real-time over transmission processes of two nodes, it only requires to save the most recent BiBGM matches of $(\mathbf{s}_i)_{i=1}^2$ and $(\overline{\mathbf{s}}_i)_{i=1}^2$ and the timing measurements in the most recent $2(w+\alpha)$-second interval; they are all the information needed to continue running ITA and BiBGM over the future timing measurements.

For a class of non-homogeneous Poisson traffic, NBFD has a consistency property as stated in the following theorem.

Theorem 4.1. *Assume that w and α are any positive numbers with $\alpha \ge \Delta$. For any $\eta \in (0,1)$, there exists an $\overline{\epsilon} \in (0,1)$ such that, for any $\epsilon \in (0,\overline{\epsilon}]$, NBFD with (w,α,ϵ) consistently detects any bidirectional flow with $R \ge \eta$ a.s., if the distributions of $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ satisfy the following assumptions*[6]:

- *Under both hypotheses, $\mathbf{S}_1$ and $\mathbf{S}_2$ are non-homogeneous Poisson processes. In addition, under $\mathcal{H}_1$, $\mathbf{S}_i = (\mathbf{F}_i^{12} \oplus \mathbf{F}_i^{21}) \oplus \mathbf{W}_i$. $\mathbf{F}_1^{12}$, $\mathbf{F}_2^{21}$, $\mathbf{W}_1$, and $\mathbf{W}_2$ are independent non-homogeneous Poisson processes, $\mathbf{F}_2^{12}$ is*[7] ***sort***$\{\mathbf{F}_1^{12}(i)+\alpha_i, i \ge 1\}$, *and $\mathbf{F}_1^{21}$ is* ***sort***$\{\mathbf{F}_2^{21}(i)+\beta_i, i \ge 1\}$ *where $\{\alpha_i, i\ge 1\}$ and $\{\beta_i, i \ge 1\}$ are random variables satisfying $\alpha_i, \beta_i \in [0,\Delta]$ a.s. Furthermore, $\{\alpha_i, i \ge 1\} \perp \mathbf{W}_1$, $\{\beta_i, i \ge 1\} \perp \mathbf{W}_2$, and*[8] *$\perp \{\alpha_i, i\ge 1\}, \{\beta_i, i \ge 1\}, \mathbf{F}_1^{12}, \mathbf{F}_2^{21}$.*
- *Let $\lambda_1(t)$, $\lambda_2(t)$, $\lambda_{f1}(t)$, and $\lambda_{f2}(t)$ denote the local intensities of $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{F}_1^{12}$, and $\mathbf{F}_2^{21}$ respectively. There exist two finite sets $\Lambda_0 \triangleq \{\overrightarrow{\mu}^{(j)} \triangleq (\mu_1^{(j)}, \mu_2^{(j)}), 1 \le j \le M_0\}$ and $\Lambda_1 \triangleq \{\overrightarrow{\lambda}^{(k)} \triangleq (\lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_{f1}^{(k)}, \lambda_{f2}^{(k)}), 1 \le k \le M_1\}$ with $\mu_i^{(j)} > 0, \lambda_i^{(k)} > 0$, $i=1,2$, $\forall j,k$. Under $\mathcal{H}_0$, $(\lambda_1(t),\lambda_2(t))$ can only take values in Λ_0. Under $\mathcal{H}_1$, $\overrightarrow{\lambda}(t) \triangleq (\lambda_1(t),\lambda_2(t),\lambda_{f1}(t),\lambda_{f2}(t))$ can only take values in Λ_1.*
- *Under $\mathcal{H}_0$, if $c(t)$ denotes the number of times that $(\lambda_1(t),\lambda_2(t))$ changes its value in $[0,t]$, then $\lim_{t\to\infty} c(t)/t = 0$. Similarly, under $\mathcal{H}_1$, if $c(t)$ denotes the number of times that $\overrightarrow{\lambda}(t)$ changes its value in $[0,t]$, then $\lim_{t\to\infty} c(t)/t = 0$.*
- *Under $\mathcal{H}_0$, if $\rho_k(t)$ $(1 \le k \le M_0)$ denotes the fraction of the time in $[0,t]$ that $(\lambda_1(t),\lambda_2(t)) = \overrightarrow{\mu}^{(k)}$, then as t increases, each $\rho_k(t)$ converges. Similarly, under $\mathcal{H}_1$, if $\rho_k(t)$ $(1 \le k \le M_1)$ denotes the fraction of the time in $[0,t]$ that $\overrightarrow{\lambda}(t) = \overrightarrow{\lambda}^{(k)}$, then as t increases, each $\rho_k(t)$ converges.*

Proof. See Appendix.

The first assumption means that under $\mathcal{H}_1$, $(\mathbf{S}_i)_{i=1}^2$ is a superposition of three independent parts: the unidirectional

[6] In other words, if the distributions of $(\mathbf{S}_i)_{i=1}^2$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ satisfy the listed assumptions (including $R \ge \eta$ a.s. under $\mathcal{H}_1$), then under all those distributions, the false alarm and miss detection probabilities of NBFD with (w,α,ϵ) vanish as t grows (as in Definition 2.3).

[7] For a countable set A of real numbers, **sort**$\{A\}$ is the sequence of the elements of A ordered in the increasing order.

[8] For random processes $\mathbf{A}_i$'s, $\mathbf{A}_1 \perp \mathbf{A}_2$ means $\mathbf{A}_1$ and $\mathbf{A}_2$ are independent, and $\perp \mathbf{A}_1, \dots, \mathbf{A}_n$ means $\mathbf{A}_1, \dots, \mathbf{A}_n$ are independent.

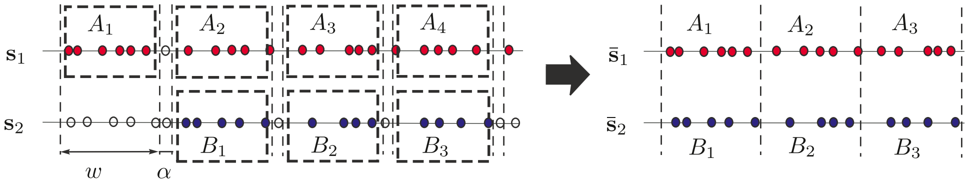

Fig. 7. ITA-double (ITAd): The sample size of $(\bar{\mathbf{s}}_i)_{i=1}^2$ doubles compared to ITA. Unlike ITA, ITAd does not throw away $\{A_2,A_4,\ldots\}$ or $\{B_2,B_4,\ldots\}$; it assembles all of $\{A_1,A_2,\ldots\}$ and $\{B_1,B_2,\ldots\}$ to generate $(\bar{\mathbf{s}}_i)_{i=1}^2$.

flow from $\mathbf{S}_1$ to $\mathbf{S}_2$, the unidirectional flow from $\mathbf{S}_2$ to $\mathbf{S}_1$, and the chaff parts. α_i and β_i represent packet delays of the two unidirectional flows, and they satisfy certain independence relationships involving the flow parts and the chaff parts. The first assumption is sufficient to guarantee that the output of ITA, $(\bar{\mathbf{S}}_i)_{i=1}^2$, is an independent pair under $\mathcal{H}_1$. The second assumption implies that the local intensities of the total traffic and flows can only take a finite number of different values. The third assumption implies that the number of intensity changes in $[0,t]$ grows as $o(t)$. Finally, the last assumption means that the fraction of the time that the intensity vector assumes a specific value converges as the observation time increases. Under these assumptions, Theorem 4.1 states that a bidirectional flow with any positive rate can be consistently detected by NBFD if ϵ is properly set. Note that the assumptions do not restrict traffic to be stationary.

As pointed out by Paxson and Floyd [24], a Poisson process is not always a good model for network arrival processes. Several network traces (*e.g.*, Ethernet and World Wide Web traffic) have been experimentally proved to display self-similarity [26–28], which Poisson processes do not show. To test the performance of NBFD over non-Poisson traffic, we will evaluate NBFD in the following section using LBL TCP traces, which were used in [24] to invalidate Poisson modeling, and real-world measurements from MSN VoIP sessions.

5. Numerical results

NBFD was tested using the synthetic Poisson traffic, LBL TCP traces, and the real-world measurements from MSN VoIP sessions. Comparison with other passive flow detectors is also provided: the wavelet analysis in [1], Detect-Attack-Chaff (DAC) in [17], and the random projection method in [20].

The wavelet analysis [1] calculates the wavelet coefficients of $N_1(t)$ and $N_2(t)$ using the mother Haar wavelet with a sufficiently large scale, where $N_i(t)$ is the number of epochs of $\mathbf{S}_i$ in $[0,t]$. Then, it calculates Pearson's correlation coefficient between the wavelet coefficients of $N_1(t)$ and that of $N_2(t)$, and declares $\mathcal{H}_1$ if the correlation coefficient is greater than a predetermined threshold κ; otherwise, it declares $\mathcal{H}_0$. The intuition of the algorithm is based on their analysis under the Poisson traffic assumption: the correlation coefficient converges to a positive constant as the scale[9] grows to infinity if $(\mathbf{S}_i)_{i=1}^2$ contains a flow.

DAC [17] is based on the intuition that as t increases $|N_1(t)-N_2(t)|$ tends to grow large when $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent, whereas it tends to stay small if $(\mathbf{S}_i)_{i=1}^2$ contains a flow with a much higher rate than the chaff part. DAC with a parameter[10] p_Δ monitors $|N_1(t)-N_2(t)|$. At every $8(p_\Delta+1)^2$ packet transmissions, both N_1 and N_2 are set to be zero and new counting begins. It declares $\mathcal{H}_0$ if $|N_1(t)-N_2(t)|$ grows larger than a threshold $2p_\Delta$. If $|N_1(t)-N_2(t)|$ stays less than $2p_\Delta$ during the whole observation duration, DAC declares $\mathcal{H}_1$. Note that under $\mathcal{H}_1$, if bursty chaff transmissions occur in either node, $|N_1(t)-N_2(t)|$ may suddenly grow larger than $2p_\Delta$ thereby resulting in a miss detection. Hence, DAC is vulnerable to bursty chaff insertion.

The random projection method in [20], which we denote by RP, is based on the idea of measuring the distance between $\mathbf{S}_1$ and $\mathbf{S}_2$ after random projection. It first partitions the observation interval into the time slots with length L_{TS}, and counts the number of epochs in each time slot. The number of epochs of $\mathbf{S}_i$ in the jth time slot is denoted by $V_i(j)$, $i=1,2$, $1\le j\le T$. Then, RP generates a set of K random basis vectors $\{B_k\in\{-1,1\}^T,\ 1\le k\le K\}$, where each $B_k(j)$ $(1\le j\le T)$ is either 1 or -1 with an equal probability.[11] After that, V_i is projected on $\{B_k, 1\le k\le K\}$: $C_i(k)\triangleq\sum_j V_i(j)B_k(j)$, $i=1,2$, $1\le k\le K$. Finally, RP obtains a K-dimensional binary vector $\overline{C}_i$, where $\overline{C}_i(k)\triangleq 1_{\{C_i(k)>0\}}$, referred to as the *binary sketch* of $\mathbf{S}_i$. The decision statistic of RP is the Hamming distance between $\overline{C}_1$ and $\overline{C}_2$. If the distance is less than a threshold *th*, RP declares $\mathcal{H}_1$; otherwise, $\mathcal{H}_0$ is declared.

5.1. Simulation results: Poisson traffic and LBL traces

We first performed Monte Carlo simulations using the synthetic non-homogeneous Poisson traffic. In the simulations, $\mathbf{S}_1$ and $\mathbf{S}_2$ are Poisson processes with intensity

[9] Since the wavelet analysis relies on the convergence of the correlation coefficient as the scale grows, a large scale is desired.

(*footnote continued*)
However, given a fixed observation duration, using too large scale can cause the sample size of the correlation coefficient estimation (*i.e.*, the number of wavelet coefficients) to be very small. To prevent this, in our experiments, the sample size is fixed to be 100, and the scale is set to be (the observation duration)/100.

[10] In [17], p_Δ is defined to be a uniform upper bound on the number of epochs of a node ($\mathbf{S}_1$ or $\mathbf{S}_2$) in any Δ-second interval. However, none of our test traces guarantees such a uniform upper bound. Hence, we tried DAC with various p_Δ values, which include large enough numbers to bound the number of epochs in any Δ-second interval with high probability.

[11] About the parameters of RP, we used $L_{TS}=0.5$ s, as recommended in [20]. As explained in [20], large K is desired since it will allow us to extract more information from $\mathbf{S}_i$. We used K=4096, which we believe is sufficiently large (four times the maximum K used in [20]).

functions $\lambda_1(t)$ and $\lambda_2(t)$ respectively. Under $\mathcal{H}_1$, $(\mathbf{S}_i)_{i=1}^2$ is a superposition of two independent parts, the unidirectional flow $(\mathbf{F}_i)_{i=1}^2$ and the chaff part $(\mathbf{W}_i)_{i=1}^2$. $\mathbf{F}_1$ is a Poisson process with intensity function $\lambda_f(t)$, and $\mathbf{F}_2$ is generated by adding a random delay to each epoch of $\mathbf{F}_1$. Random delays are independent and identically distributed (i.i.d.) and uniformly distributed in $[0,\Delta]$, where $\Delta = 0.1$ s. $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent Poisson processes with intensity functions $\lambda_1(t) - \lambda_f(t)$ and $\lambda_2(t) - \lambda_f(t)$, respectively. In each run of the simulation, $(\lambda_1(t), \lambda_2(t), \lambda_f(t))$ is piecewise constant, and it takes different values in the first third, the second third, and the last third of the observation duration. Specifically, it follows one of the below change scenarios with equal probability:

1. $(15, 15, 5) \rightarrow (15, 15, 12) \rightarrow (15, 15, 7)$.
2. $(25, 10, 8) \rightarrow (10, 10, 8) \rightarrow (10, 25, 8)$.
3. $(25, 25, 20) \rightarrow (12, 12, 7) \rightarrow (8, 8, 3)$.
4. $(21, 15, 14) \rightarrow (12, 6, 5) \rightarrow (12, 24, 5)$.

Under $\mathcal{H}_0$, $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent Poisson processes, and in each run of the simulation, $(\lambda_1(t), \lambda_2(t))$ follows one of the above change scenarios (with no λ_f part) with equal probability. In real world, such changes in intensity may correspond to the beginning of new sessions, the end of old sessions, the rate change of existing sessions, and so on. All change scenarios have the same average rates, but each scenario displays a different dynamics. By this simulation setting, we aimed at testing the performance of detectors over the non-stationary traffic displaying possibly a different dynamics at each observation interval.

Fig. 8 shows the ROC curves of NBFD (with ITAd), NBFD (with ITA), the wavelet analysis, DAC, and RP. To obtain the ROC curves, we increased ϵ of NBFD and κ of the wavelet analysis from 0 to 1 with an increment of 0.01, p_Δ of DAC from 4 to 100 with an increment of 2, and th of RP from 0 to K (K=4096) with an increment of 1 while plotting $(P_F, 1-P_M)$ of each case, where P_F and P_M denote the false alarm probability and miss detection probability respectively. When we further increased the sample size, the ROC curves of NBFD (ITAd), NBFD (ITA), and the wavelet analysis approached the upper left corner implying that perfect detection is possible if the thresholds are properly set. On the other hand, DAC and RP resulted in non-negligible error probabilities in every case, and their ROC curves did not improve much from the curves in Fig. 8, even when we further increased the observation duration to 160 s. By comparing the ROC curves of NBFD (ITAd) and NBFD (ITA), we can observe that ITAd, the heuristic to double the sample size of $(\overline{\mathbf{s}}_i)_{i=1}^2$, resulted in a better detection performance than ITA. In all our simulations and experiments, ITAd consistently showed better results than ITA. In the rest of this section, NBFD is assumed to employ ITAd and will be compared with other detectors.

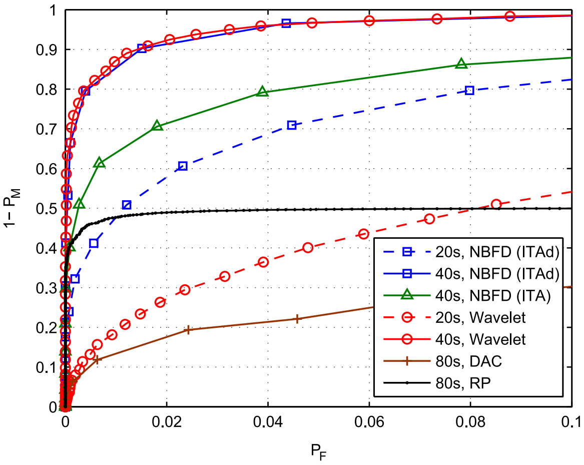

Fig. 8. ROC curves of NBFD (ITAd), NBFD (ITA), the wavelet analysis, DAC, and RP for different observation durations: NBFD parameters are $w = 2$ s and $\alpha = \Delta = 0.1$ s, and the number of Monte Carlo runs is 10 000.

To test the performance of detectors over non-Poisson traffic, we generated synthetic traffic based on the TCP packet timestamps in LBL-PKT-3 (2 hour), LBL-PKT-4 (1 hour), and LBL-PKT-5 (1 hour) in [24]. These traces were measured at the Lawrence Berkeley Laboratory's wide-area Internet gateway, and each trace was gathered at a different date in January 1994. For the detail, refer to [24]. From each data set, we extracted timestamps of TCP packets that originated from specific users, and used them for traffic generation. For the flow part of $\mathcal{H}_1$ traffic, timestamps of one user in LBL-PKT-3 were used as $\mathbf{F}_1$, and $\mathbf{F}_2$ was generated by adding a delay to each epoch in $\mathbf{F}_1$. The delays are i.i.d. and uniformly distributed in $[0,\Delta]$, where $\Delta = 0.1$ s. For the chaff part, timestamps of one user in LBL-PKT-4 were used as $\mathbf{W}_1$, and those of one user in LBL-PKT-5 were used as $\mathbf{W}_2$. For $\mathcal{H}_0$ traffic, $\mathbf{S}_1$ is generated by superposing traces of two users in LBL-PKT-4, and $\mathbf{S}_2$ is similarly generated with two users in LBL-PKT-5. Using different sets of users for the traffic generation, we were able to create the 4-hour long test traffic.

We tested DAC with various p_Δ ranging from 10 to 400, but its miss detection probability was higher than 0.38 in every case. This is not surprising because DAC is vulnerable to bursty chaff transmissions and LBL TCP traces were shown to be bursty in [24]. Table 3 shows the error probabilities of NBFD, the wavelet analysis, and RP. For NBFD, we used $\epsilon = 0.05$. For the wavelet analysis and RP, assuming the absence of a parametric model and training data, we have no clear standard to set their thresholds. Hence, we tried all values from 0 to 1 with an increment of 0.01 for κ of the wavelet analysis and all values from 0 to 4096 with an increment of 1 for th of RP, and found

Table 3
Performance over LBL TCP traces: NBFD parameters are $w = 2$ s, $\alpha = \Delta = 0.1$ s, and $\epsilon = 0.05$. The numbers of experiments are 180, 90, and 45 for observation duration 80 s, 160 s, and 320 s, respectively. Under $\mathcal{H}_0$, the average traffic rate is $(\lambda_1, \lambda_2) = (36.4, 36.1)$. Under $\mathcal{H}_1$, $(\lambda_1, \lambda_2) = (36.1, 36.8)$. The fraction of chaff in $\mathcal{H}_1$ traffic is 0.37.

Time (s)	NBFD		Wavelet			RP		
	P_F	P_M	κ	P_F	P_M	th	P_F	P_M
80	0	0.100	0.19	0.034	0.056	762	0.101	0.144
160	0	0.057	0.20	0.034	0.056	793	0.112	0.133
320	0	0.022	0.19	0.023	0.067	774	0	0.089

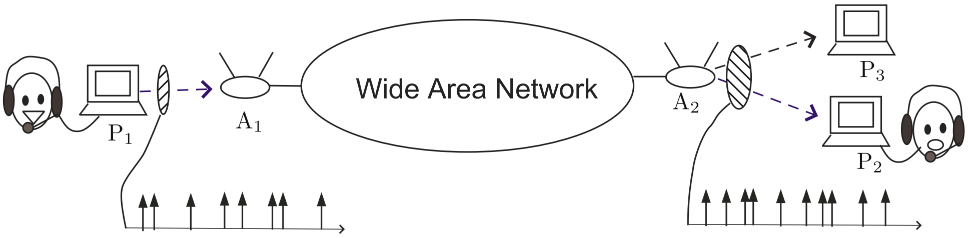

Fig. 9. If P$_1$ has a VoIP conversation with either P$_2$ or P$_3$, the VoIP packets should depart from P$_1$ and travel through A$_2$.

their crossover error rates and the corresponding thresholds, which are listed in Table 3. NBFD and the wavelet analysis outperformed RP, and for long observation durations (160 s and 320 s), NBFD performed better than the wavelet analysis.

5.2. Experimental results: MSN VoIP traffic

We tested the detectors using three-and-a-half-hour long real-world traffic involving the MSN VoIP application,[12] which is a representative example of latency-sensitive applications. Fig. 9 is illustrating the experimental setup. The laptop P$_1$ is located in the place covered by the wireless access point A$_1$, and two other laptops, P$_2$ and P$_3$, are located in the different places covered by the wireless access point A$_2$, which is controlled to serve only P$_2$ and P$_3$. Suppose it is known that P$_1$ is engaged in a VoIP conversation. By measuring the wireless transmission epochs of P$_1$ and A$_2$, our objective is to detect whether P$_1$ is having a VoIP conversation with any device served by the access point A$_2$. In practice, there may be additional information available: packet sizes, protocol types (TCP or UDP), destination addresses, and so on. However, here we assume that we have no access to such information due to encryption or other countermeasures employed by the network administrator, and only the timing measurements are available.

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ denote the transmission processes of P$_1$ and A$_2$ respectively. Under $\mathcal{H}_1$, P$_1$ has a VoIP conversation with P$_2$, and P$_3$ downloads a file from a distant FTP server with 20 kB/s rate. Since A$_2$ transmits packets for both P$_2$ and P$_3$, its transmission timings of FTP packets, destined for P$_3$, form the chaff part of $\mathbf{S}_2$. Under $\mathcal{H}_0$, P$_1$ and P$_2$ engage in independent VoIP conversations while P$_3$ does the same job as in $\mathcal{H}_1$. Hence, VoIP packet timings in $\mathbf{S}_1$ and those in $\mathbf{S}_2$ are independent under $\mathcal{H}_0$. Under both hypotheses, the timings of network control/management packets from P$_1$ and A$_2$ (except beacon frames of A$_2$) are also included in $\mathbf{S}_1$ and $\mathbf{S}_2$.

We assumed that Δ is 150 ms, which is the upper bound of acceptable end-to-end delays of VoIP packets recommended by ITU-T recommendation G.114 [5]. We first tested DAC with various p_Δ ranging from 10 to 400. Similar to the result on LBL TCP traces, the miss detection probability was higher than 0.55 in every case due to the bursty chaff transmissions (*i.e.*, bursty FTP transmissions from A$_2$ to P$_3$).

Table 4
The MSN VoIP experiment: NBFD parameters are $w = 2$ s, $\alpha = \Delta = 0.15$ s, and $\epsilon = 0.05$. The numbers of experiments are 162, 81, and 40 for observation duration 80 s, 160 s, and 320 s, respectively. Under $\mathcal{H}_0$ and $\mathcal{H}_1$, the average rate is $(\lambda_1,\lambda_2) = (26.8,34.9)$. The fraction of chaff in $\mathcal{H}_1$ traffic is 0.18.

Time (s)	NBFD		Wavelet			RP		
	P_F	P_M	κ	P_F	P_M	th	P_F	P_M
80	0.086	0.056	0.14	0.093	0.093	949	0.086	0.099
160	0	0.049	0.17	0.012	0.012	989	0.049	0.074
320	0	0	0.23	0	0	1005	0.075	0.050

Table 4 shows the error probabilities of NBFD, the wavelet analysis, and RP. As in the test using LBL traces, we used $\epsilon = 0.05$ for NBFD; for the wavelet analysis and RP, the crossover error rates and the corresponding thresholds are listed in the table. NBFD and the wavelet analysis outperformed RP, and they displayed vanishing error probabilities as the observation duration increases.

In all the tests we executed, NBFD and the wavelet analysis consistently outperformed DAC and RP. Even though the wavelet analysis performed well over most traces, we need to recall that the results in Tables 3 and 4 were possible because its threshold κ was set *a posteriori* to minimize its error probabilities. If neither a training data set nor a parametric model is available, we have no clear standard to set κ. For the further comparison of NBFD and the wavelet analysis, Fig. 10 shows P_F and P_M of NBFD and the wavelet analysis with various thresholds. We can observe that the optimal κ of the wavelet analysis varies significantly for different observation durations and different test traces. For instance, in the test result for synthetic Poisson traffic, $\kappa \approx 0.25$ gave the best performance when the observation duration is 80 s, but it resulted in $P_M \approx 0.85$ for the 20 s case. In addition, for the fixed observation duration of 80 s, the optimal κ for the Poisson traffic (≈ 0.25) and that for the VoIP traffic (≈ 0.15) are quite different. In contrast, for NBFD, it can be observed that $\epsilon = 0.05$ results in almost optimal performance in every case. Especially, in every test, its false alarm probability vanished as the observation duration increases. This suggests that under $\mathcal{H}_0$, the difference between $\overline{R}(t)$ and $\overline{\tau}(t)$ of NBFD is well bounded by $\epsilon = 0.05$.

6. Conclusion

In this paper, we studied timing-based detection of information flows in a network. We formulated flow

[12] Windows Live Messenger 2009 (14.0.8089.726) was used for MSN VoIP calls, and Wireshark network protocol analyzer (ver 1.2.6.) with the AirPcap classic adaptor was used to record the timings of wireless transmissions.

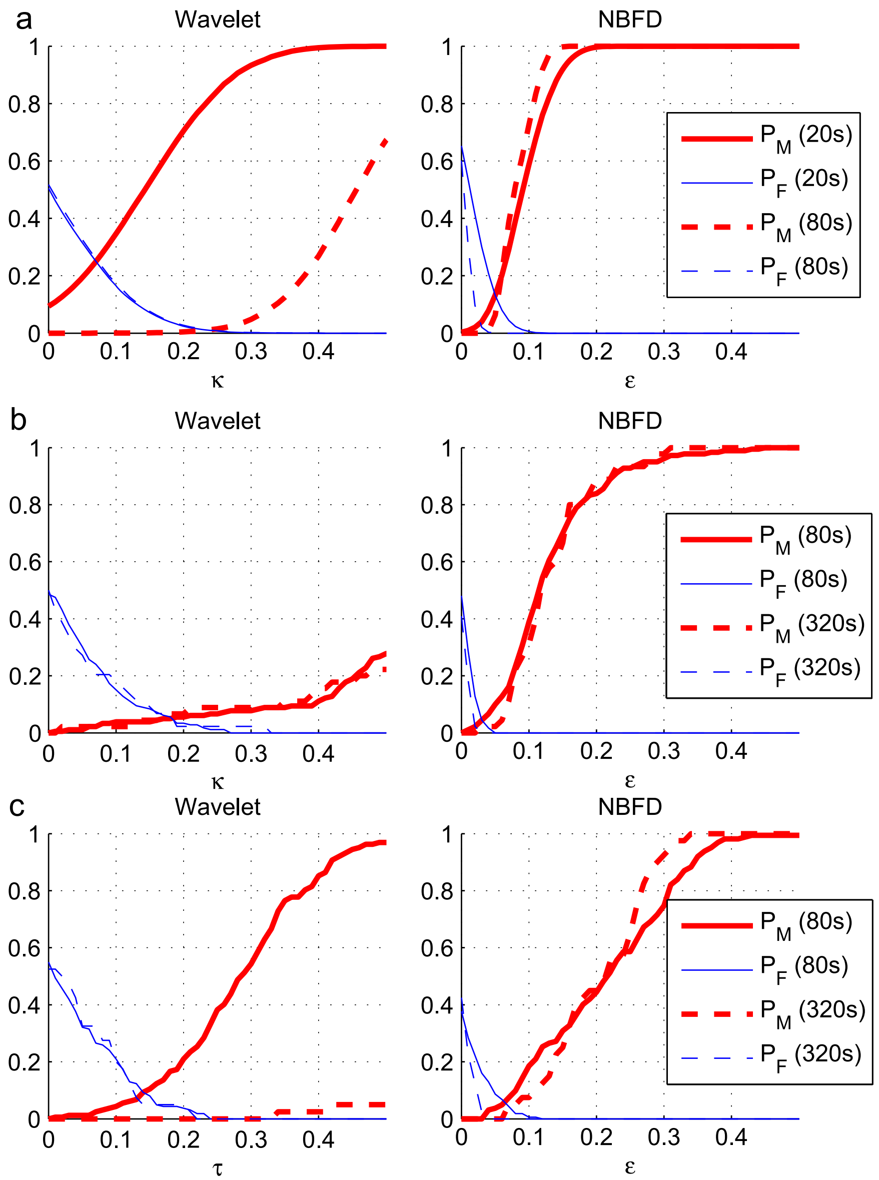

Fig. 10. False alarm and miss detection probabilities of the wavelet analysis and NBFD with various thresholds. (a) Synthetic Poisson traffic, (b) LBL TCP traces, (c) MSN VoIP experiment.

detection as a binary composite hypothesis testing problem and presented a detection algorithm that requires neither a parametric model nor a training data set. The detection algorithm is memory-efficient, and it has linear computational complexity with respect to the sample size. We proved the consistency of the algorithm for a class of non-homogeneous Poisson processes. In addition, the algorithm was tested using the synthetic Poisson traffic, LBL TCP traces, and the real-world measurements from the MSN VoIP experiment. Our detector was superior to other passive detection schemes in terms of detection performance and suitability for the nonparametric and unsupervised setting. Notably, the test results for LBL TCP traces and the MSN VoIP traffic suggest that our detector may perform well over the traffic with more general distribution than a Poisson process.

Acknowledgments

This work was sponsored by National Science Foundation under Contract CCF-0728872 and Army Research Office MURI Program under award W911NF-08-1-0238. The first author was partially supported by Samsung Scholarship. Part of this work was presented at 2010 11th International Workshop on Signal Processing Advances in Wireless Communications, 2010 Military Communications Conference, and 2010 48th Annual Allerton Conference on Communication, Control, and Computing.

Appendix A. Proof of Theorem 3.1

We use the following lemma about the relation between BiBGM with Δ and Bounded-Greedy-Match

(BGM) [17] with $2\varDelta$ (for the detail of BGM, refer to Section 4.A of [18]).

Lemma A.1. *Running BiBGM on* $(\mathbf{s}_i)_{i=1}^2$ *with* $\varDelta$ *is equivalent to the following*:

1. *Increase all the epochs of* $\mathbf{s}_2$ *by* $\varDelta$.
2. *Apply BGM with the delay constraint* $2\varDelta$ *to the modified measurements.*

Proof. Let $\hat{\mathbf{s}}_2$ be a sequence generated by increasing every epoch in $\mathbf{s}_2$ by $\varDelta$ (*i.e.*, $\hat{s}_2(i) = s_2(i)+\varDelta, 1 \le i \le |\mathcal{S}_2|$). Then, replacing $s_2(n)$ with $\hat{s}_2(n) - \varDelta$ in Table 1 results in exactly the same pseudocode with BGM with $2\varDelta$ on $(\mathbf{s}_1,\hat{\mathbf{s}}_2)$ (see Table 3 in [18] for the pseudocode). □

Note that $(a,b) \in \mathcal{S}_1 \times \mathcal{S}_2$ and $|a-b| < \varDelta$ if and only if $(a,b+\varDelta) \in \mathcal{S}_1 \times \hat{\mathcal{S}}_2$ and $b+\varDelta \in [a,a+2\varDelta]$. Hence, the optimal partitioning of $(\mathbf{s}_i)_{i=1}^2$ is equivalent to partitioning $(\mathbf{s}_1,\hat{\mathbf{s}}_2)$ into the unidirectional flow part (with the delay constraint $2\varDelta$) and the chaff part such that the flow part is maximized; BGM with $2\varDelta$ was proved in [17] to achieve the optimal partitioning of $(\mathbf{s}_1,\hat{\mathbf{s}}_2)$. Thus, Lemma A.1 implies the result.

Appendix B. Proof of Theorem 3.2

Let $\hat{\mathbf{S}}_2$ denote the point process with $\hat{S}_2(i) = S_2(i)+\varDelta$, $i \ge 1$. Theorem 4.2 in [18] showed that if we run BGM with $2\varDelta$ on $(\mathbf{S}_1,\hat{\mathbf{S}}_2)$, the fraction of the matched epochs in total epochs converges a.s. to the following:

$$\begin{cases} \dfrac{2\lambda_1\lambda_2(1-e^{2\varDelta(\lambda_1-\lambda_2)})}{(\lambda_1+\lambda_2)(\lambda_2-\lambda_1 e^{2\varDelta(\lambda_1-\lambda_2)})} & \text{if } \lambda_1 \ne \lambda_2 \\ \dfrac{2\lambda\varDelta}{1+2\lambda\varDelta} & \text{if } \lambda_1=\lambda_2=\lambda \end{cases}$$

Therefore, Lemma A.1 implies the result.

Appendix C. Proof of Theorem 3.3

We first introduce the following lemma about the statistical behavior of $\overline{R}(t)$ under $\mathcal{H}_1$.

Lemma C.1. *Suppose that the distributions of* $(\mathbf{S}_i)_{i=1}^2$ *under* $\mathcal{H}_1$ *satisfy the conditions that* (i) $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *have rates* λ_1 *and* λ_2 *respectively*, (ii) $(\mathbf{F}_1,\mathbf{F}_2)$ *is a bidirectional flow with rate*[13] λ_f, *and* (iii) $\mathbf{W}_1$ *and* $\mathbf{W}_2$ *are homogeneous Poisson processes. Then, under every distribution in* $\mathcal{H}_1$, $\liminf_{t\to\infty}\overline{R}(t) \ge \theta_{(\lambda_1,\lambda_2,\lambda_f)}$ *a.s., where* $\theta_{(\lambda_1,\lambda_2,\lambda_f)}$ *is defined as*

$$\begin{cases} \dfrac{2\lambda_1 - 2\lambda_2\left(\frac{\lambda_1-\lambda_f}{\lambda_2-\lambda_f}\right)e^{2\varDelta(\lambda_1-\lambda_2)}}{(\lambda_2+\lambda_1)\left(1-\left(\frac{\lambda_1-\lambda_f}{\lambda_2-\lambda_f}\right)e^{2\varDelta(\lambda_1-\lambda_2)}\right)} & \text{if } \lambda_1 \ne \lambda_2 \\ \dfrac{\lambda_f+2\lambda(\lambda-\lambda_f)\varDelta}{\lambda(1+2(\lambda-\lambda_f)\varDelta)} & \text{if } \lambda_1=\lambda_2=\lambda \end{cases}$$

Proof. Let $N(t)$, $N_f(t)$, and $N_c(t)$ denote the number of epochs of $(\mathbf{S}_i)_{i=1}^2$, $(\mathbf{F}_i)_{i=1}^2$, and $(\mathbf{W}_i)_{i=1}^2$ in $[0,t]$, respectively. $M(t)$ denotes the number of the matched epochs found by running BiBGM over $(\mathbf{S}_i)_{i=1}^2$ in $[0,t]$.

Consider running BiBGM on $(\mathbf{F}_i)_{i=1}^2$ and $(\mathbf{W}_i)_{i=1}^2$ separately in $[0,t]$: $\hat{M}(t)$ denotes the sum of the number of the matched epochs in $(\mathbf{F}_i)_{i=1}^2$ and that in $(\mathbf{W}_i)_{i=1}^2$, and $\overline{R}_w(t)$ denotes the fraction of the matched epochs in $(\mathbf{W}_i)_{i=1}^2$. Theorem 3.1 implies that running BiBGM on $(\mathbf{S}_i)_{i=1}^2$ results in a greater or an equal number of matched epochs than running BiBGM on $(\mathbf{F}_i)_{i=1}^2$ and $(\mathbf{W}_i)_{i=1}^2$ separately. Therefore,

$$M(t) \ge \hat{M}(t) = N_f(t) + N_c(t)\overline{R}_w(t)$$

and

$$\frac{M(t)}{N(t)} \ge \frac{N_f(t)}{N(t)} + \frac{N_c(t)}{N(t)}\overline{R}_w(t) = \frac{N_f(t)/t}{N(t)/t} + \frac{N_c(t)/t}{N(t)/t}\overline{R}_w(t)$$

We have $M(t)/N(t) = \overline{R}(t)$, $\lim_{t\to\infty}(N_f(t)/t)/(N(t)/t) = 2\lambda_f/(\lambda_1+\lambda_2)$ a.s., $\lim_{t\to\infty}(N_c(t)/t)/(N(t)/t) = (\lambda_1+\lambda_2-2\lambda_f)/(\lambda_1+\lambda_2)$ a.s., and $\lim_{t\to\infty}\overline{R}_w(t) = \phi_{(\lambda_1-\lambda_f,\lambda_2-\lambda_f)}$ a.s., where ϕ is defined as in Theorem 3.2. Thus,

$$\liminf_{t\to\infty}\overline{R}(t) \ge \frac{2\lambda_f}{\lambda_1+\lambda_2} + \frac{\lambda_1+\lambda_2-2\lambda_f}{\lambda_1+\lambda_2}\phi_{(\lambda_1-\lambda_f,\lambda_2-\lambda_f)} \quad \text{a.s.}$$

It can be shown that the right hand side is $\theta_{(\lambda_1,\lambda_2,\lambda_f)}$. □

Let η be any fixed number in $(0, 2\min\{\lambda_1,\lambda_2\}/(\lambda_1+\lambda_2))$ and τ be the suggested threshold for η. Then, there exists a positive $\hat{\lambda}_f$ such that $\hat{\lambda}_f/(\lambda_1+\lambda_2) = \eta/4$. Let $h(x) \triangleq \theta_{(\lambda_1,\lambda_2,x)}$. It can be checked that $h(x)$ is strictly increasing in $[0,\min\{\lambda_1,\lambda_2\}]$, and $h(\hat{\lambda}_f)$ is equal to τ.

(i) Miss detection probability: Suppose $\mathcal{H}_1$ is true and $R \ge \eta$ a.s. Then, $R = 2\lambda_f/(\lambda_1+\lambda_2)$ and $\lambda_f = ((\lambda_1+\lambda_2)/2)R > \overline{\lambda}_f \triangleq 3(\lambda_1+\lambda_2)\eta/8 > \hat{\lambda}_f$, because $R \ge \eta > 3\eta/4 > \eta/2$. Lemma C.1 and the monotonicity of h give

$$\liminf_{t\to\infty}\overline{R}(t) \ge \theta_{(\lambda_1,\lambda_2,\lambda_f)} = h(\lambda_f) > h(\overline{\lambda}_f) > h(\hat{\lambda}_f) = \tau \quad \text{a.s.}$$

Hence, $\lim_{t\to\infty}\Pr(\overline{R}(t) < \tau) = 0$.

(ii) False alarm probability: Under $\mathcal{H}_0$,

$$\lim_{t\to\infty}\overline{R}(t) = \phi_{(\lambda_1,\lambda_2)} = h(0) < h(\hat{\lambda}_f) = \tau \quad \text{a.s.}$$

and thus $\lim_{t\to\infty}\Pr(\overline{R}(t) \ge \tau) = 0$. Furthermore, Lemma A.1 and Theorem 6.4 in [18] imply the exponential decay of the false alarm probability.

Appendix D. Proof of Theorem 4.1

We first introduce a useful lemma.

Lemma D.1. *Suppose that* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are non-homogeneous Poisson processes, and their local intensities always stay in* $[\lambda_{min},\lambda_{max}]$, *where* $\lambda_{min} > 0$. *As illustrated in Fig. 11, we partition* $[0, \infty)$ *into a countable number of subintervals*: I_i *denotes the ith subinterval,* T_i *is the length of* I_i, *and* $d(t)$ *denotes the number of* I_i*'s with* $I_i \subset [0,t]$. *Suppose* $d(t)/t$ *decreases to 0 as t grows.*

[13] If $N_1(t)$, $N_2(t)$, and $N_F(t)$ denote the number of epochs of $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{F}_1$ in $[0,t]$, respectively, then $\lim_{t\to\infty}N_i(t)/t = \lambda_i$ a.s. for $i=1, 2$, and $\lim_{t\to\infty}N_F(t)/t = \lambda_f$ a.s.

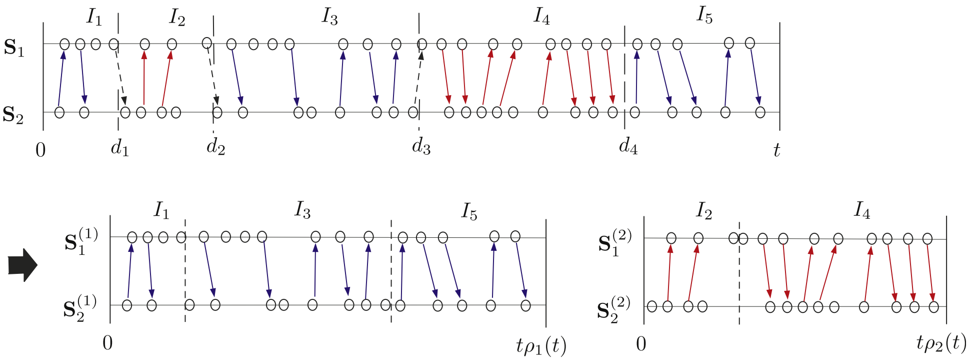

Fig. 11. In this example, $M{=}2$, $a_{1;1}=1$, $a_{1;2}=3$, $a_{1;3}=5$, $a_{2;1}=2$, and $a_{2;2}=4$. We ran BiBGM on $(\mathbf{S}_i)_{i=1}^2$ and marked the matches by the arrows. Some matches consist of epochs in two different partitions, and they are marked by the dashed arrows. The matches consisting of epochs in a single partition are marked by the solid arrows. One can observe that each solid arrow in $(\mathbf{S}_i)_{i=1}^2$ can be found either in $(\mathbf{S}_i^{(1)})_{i=1}^2$ or $(\mathbf{S}_i^{(2)})_{i=1}^2$.

Let M be a finite natural number and suppose we partition $\{I_i, i\ge 1\}$ into M subsets $\{I_{a_{k;i}}, i\ge 1\}$, $1\le k\le M$, where $(a_{k;i})_{i\ge 1}$, $1\le k\le M$, are subsequences of $(1,2,3,\ldots)$. For $1\le k\le M$, we use the epochs of $(\mathbf{S}_i)_{i=1}^2$ in $(I_{a_{k;i}})_{i\ge 1}$ to generate point processes $(\mathbf{S}_i^{(k)})_{i=1}^2$, as described in Fig. 11:

1. *Initially, $\mathbf{S}_1^{(k)}$ and $\mathbf{S}_2^{(k)}$ have no epoch.*
2. *For $n\ge 1$, for $i=1,2$, subtract $\sum_{j=1}^{a_{k;n}-1} T_j$ from all the epochs of $\mathbf{S}_i$ in the interval $I_{a_{k;n}}$, add $\sum_{j=1}^{n-1} T_{a_{k;j}}$ to them, and add these epochs to $\mathbf{S}_i^{(k)}$.*

Let $N(t)$ denote the number of epochs of $(\mathbf{S}_i)_{i=1}^2$ in $[0,t]$ and $N^{(k)}(t)$ denote the number of epochs of $(\mathbf{S}_i^{(k)})_{i=1}^2$, whose original epoch in $(\mathbf{S}_i)_{i=1}^2$ is in $[0,t]$; by definition, $N(t)=\sum_{k=1}^M N^{(k)}(t)$. We run BiBGM on $(\mathbf{S}_i)_{i=1}^2$ and let $\overline{R}(t)$ denote the fraction of the matched epochs in the total epochs in $[0,t]$. In addition, we run BiBGM on $(\mathbf{S}_i^{(k)})_{i=1}^2$ separately for each k, and $N_f^{(k)}(t)$ denotes the number of the matched epochs among the earliest $N^{(k)}(t)$ epochs of $(\mathbf{S}_i^{(k)})_{i=1}^2$. And, we define $\hat{R}(t)$ as $\sum_{k=1}^M N_f^{(k)}(t)/N(t)$.

Then, $\lim_{t\to\infty}\overline{R}(t)-\hat{R}(t)=0$ almost surely.

Proof. Let $N_f(t)$ denote the number of BiBGM-matched epochs of $(\mathbf{S}_i)_{i=1}^2$ in $[0,t]$. Then, by definition, $\overline{R}(t)=N_f(t)/N(t)$. Let d_i denote the time that the ith division occurs; in other words, d_i is the time that the ith jump of $d(t)$ occurs. Formally, we say that a BiBGM match (t_1,t_2), where t_i is an epoch of $\mathbf{S}_i$, is *broken* if $t_1\in I_a$, $t_2\in I_b$, and $a\neq b$. Let $\tilde{N}_f(t)$ denote the number of epochs of the unbroken BiBGM matches in $[0,t]$. As described in Fig. 11, if an unbroken BiBGM match (t_1,t_2) in $[0,t]$ is such that t_1 and t_2 are included in a single partition $I_{a_{k;i}}$, then its shifted version can be found in the $[0,t\rho_k(t)]$ interval of $(\mathbf{S}_i^{(k)})_{i=1}^2$, where $\rho_k(t)$ is the fraction of $(\bigcup_{i\ge 1} I_{a_{k;i}})\cap[0,t]$ in $[0,t]$. In addition, Theorem 3.1 implies that $N_f^{(k)}(t)$ is no less than the number of epochs belonging to the shifted unbroken matches in $[0,t\rho_k(t)]$ of $(\mathbf{S}_i^{(k)})_{i=1}^2$ (*i.e.*, solid arrows in Fig. 11). Therefore, $\sum_{k=1}^M N_f^{(k)}(t)\ge\tilde{N}_f(t)$.

For $j=1,2$, let $X_j(i)$ denote the number of epochs of $\mathbf{S}_j$ in $[\max\{(d_{i-1}+d_i)/2, d_i-\Delta\}, \min\{(d_i+d_{i+1})/2, d_i+\Delta\})$, where $d_0\triangleq -d_1$. The number of epochs of the broken matches in $[0,t]$ is bounded above by $\sum_{i=1}^{d(t)} X_1(i)+\sum_{i=1}^{d(t)} X_2(i)$. Hence,

$$\tilde{N}_f(t)\ge N_f(t)-\sum_{i=1}^{d(t)} X_1(i)-\sum_{i=1}^{d(t)} X_2(i)$$

There exist sequences of i.i.d. Poisson random variables $(\overline{X}_1(i))_{i\ge 1}$ and $(\overline{X}_2(i))_{i\ge 1}$ with mean $2\lambda_{max}\Delta$ such that $X_j(i)\le\overline{X}_j(i)$ a.s. for $i\ge 1$, $j=1,2$. Hence,

$$\sum_{k=1}^M N_f^{(k)}(t)\ge N_f(t)-\sum_{i=1}^{d(t)}\overline{X}_1(i)-\sum_{i=1}^{d(t)}\overline{X}_2(i),$$

$$\overline{R}(t)-\hat{R}(t)\le\frac{\sum_{i=1}^{d(t)}\overline{X}_1(i)}{N(t)}+\frac{\sum_{i=1}^{d(t)}\overline{X}_2(i)}{N(t)}.$$

For $j=1,2$, we have

$$\limsup_{t\to\infty}\frac{d(t)/t}{N(t)/t}\frac{\sum_{i=1}^{d(t)}\overline{X}_j(i)}{d(t)}=0 \quad \text{a.s.}$$

Hence,

$$\limsup_{t\to\infty}(\overline{R}(t)-\hat{R}(t))\le 0 \quad \text{a.s.}$$

Similarly, we can partition $(\mathbf{S}_i^{(k)})_{i=1}^2$ at time points $(d_{k;i})_{i\ge 1}$, where $d_{k;i}\triangleq\sum_{j=1}^i T_{a_{k;j}}$, and use the number of unbroken BiBGM matches of $(\mathbf{S}_i^{(k)})_{i=1}^2$ in $[0,t\rho_k(t)]$, $1\le k\le M$, to obtain a lower bound on the number of BiBGM matches of $(\mathbf{S}_i)_{i=1}^2$ in $[0,t]$. Then, based on the similar argument, we can derive $\liminf_{t\to\infty}(\overline{R}(t)-\hat{R}(t))\ge 0$ a.s. Hence, we have $\lim_{t\to\infty}(\overline{R}(t)-\hat{R}(t))=0$ a.s. □

The proof consists of two parts: one for proving the vanishing false alarm probability under $\mathcal{H}_0$, and the other for proving the vanishing miss detection probability under $\mathcal{H}_1$.

D.1. False alarm probability

Suppose that $\mathcal{H}_0$ is true and the distribution of $(\mathbf{S}_i)_{i=1}^2$ satisfies the assumptions of the theorem. $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent non-homogeneous Poisson processes, and so are the output of ITA, $\bar{\mathbf{S}}_1$ and $\bar{\mathbf{S}}_2$. Suppose we run BiBGM on $(\mathbf{S}_i)_{i=1}^2$ and let $\bar{R}(t)$ denote the fraction of the matched epochs in the total epochs in $[0,t]$. We also run BiBGM on $(\bar{\mathbf{S}}_i)_{i=1}^2$ and let $\bar{T}(t)$ denote the fraction of the matched epochs in the total epochs in $[0,\lfloor t/2(w+\alpha)\rfloor w]$. In the following, we will show that $\bar{R}(t)-\bar{T}(t)$ converges a.s. to 0.

Because $\lim_{t\to\infty}c(t)/t=0$, there are at most a countable number of intensity changes. Let $(c_i)_{i\geq 1}$ denote the increasing sequence of the time points at which $(\lambda_1(t),\lambda_2(t))$ changes. We partition $[0,\infty)$ into a countable number of subintervals $\{I_i \triangleq [c_{i-1},c_i), i\geq 1\}$. For $1\leq k\leq M_0$, $(a_{k;i})_{i\geq 1}$ denotes the increasing sequence of all the indices of I_i's in which $(\lambda_1(t),\lambda_2(t))=\vec{\mu}^{(k)}$. For each k, we use the epochs of $(\mathbf{S}_i)_{i=1}^2$ in $(I_{a_{k;i}})_{i\geq 1}$ to generate a pair of point processes $(\mathbf{S}_i^{(k)})_{i=1}^2$, as described in Lemma D.1.

Let $N(t)$ denote the number of epochs of $(\mathbf{S}_i)_{i=1}^2$ in $[0,t]$. Suppose we run BiBGM on $(\mathbf{S}_i^{(k)})_{i=1}^2$ separately for $1\leq k \leq M_0$. $N^{(k)}(t)$ denotes the number of epochs of $(\mathbf{S}_i^{(k)})_{i=1}^2$ in $[0,t\rho_k(t)]$, and $N_f^{(k)}(t)$ denotes the number of BiBGM-matched epochs among those $N^{(k)}(t)$ epochs. Then, Lemma D.1 implies $\lim_{t\to\infty}(\bar{R}(t)-\sum_{k=1}^{M} N_f^{(k)}(t)/N(t))=0$ a.s. And,

$$\frac{\sum_{k=1}^{M_0} N_f^{(k)}(t)}{N(t)} = \frac{t}{N(t)}\sum_{k=1}^{M_0}\rho_k(t)\frac{N^{(k)}(t)}{t\rho_k(t)}\frac{N_f^{(k)}(t)}{N^{(k)}(t)}.$$

By analyzing the limiting behaviors of $t/N(t)$, $\rho_k(t)N^{(k)}(t)/(t\rho_k(t))$, and $N_f^{(k)}(t)/N^{(k)}(t)$ (use Theorem 3.2), we have

$$\lim_{t\to\infty}\frac{t}{N(t)}\sum_{k=1}^{M_0}\rho_k(t)\frac{N^{(k)}(t)}{t\rho_k(t)}\frac{N_f^{(k)}(t)}{N^{(k)}(t)} = \frac{\sum_{k=1}^{M_0}\rho_k(\mu_1^{(k)}+\mu_2^{(k)})\phi_{(\mu_1^{(k)},\mu_2^{(k)})}}{\sum_{k=1}^{M_0}\rho_k(\mu_1^{(k)}+\mu_2^{(k)})} \quad \text{a.s.}$$

where $\rho_k \triangleq \lim_{t\to\infty}\rho_k(t)$ and ϕ is as defined in Theorem 3.2. Then, by Lemma D.1,

$$\lim_{t\to\infty}\bar{R}(t) = \frac{\sum_{k=1}^{M_0}\rho_k(\mu_1^{(k)}+\mu_2^{(k)})\phi_{(\mu_1^{(k)},\mu_2^{(k)})}}{\sum_{k=1}^{M_0}\rho_k(\mu_1^{(k)}+\mu_2^{(k)})} \quad \text{a.s.} \tag{D.1}$$

Now, we will prove that $\bar{T}(t)$ also converges almost surely to the same constant. Let $\bar{c}_i \triangleq (w/2(w+\alpha))c_i$, $\forall i$. As depicted in Fig. 12, the local intensities of $\bar{\mathbf{S}}_1$ and $\bar{\mathbf{S}}_2$, denoted by $(\bar{\lambda}_1(t),\bar{\lambda}_2(t))$, may be equal to $(\mu_1^{(j)},\mu_2^{(k)})$ with $j\neq k$, and it happens only if any $\bar{c}_i$ is in $[w\lfloor t/w\rfloor, w(\lfloor t/w\rfloor+1))$. Define C as a set

$$\{[w(k-1),wk) : k\in\mathbb{N},\ \exists\, i \text{ s.t. } \bar{c}_i\in[w(k-1),wk)\}$$

As illustrated in Fig. 12, we partition $[0,\infty)$ of $(\bar{\mathbf{S}}_i)_{i=1}^2$ into the intervals in C and the gap intervals between two adjacent intervals in C, and $(J_i)_{i\geq 1}$ denotes the sequence of these intervals arranged in a time order. $\{\bar{a}_{0;i},\ i\geq 1\}$ denotes the increasing sequence of the indices of J_i's satisfying $J_i\in C$. For $1\leq k\leq M_0$, $\{\bar{a}_{k;\,i},\ i\geq 1\}$ denotes the increasing sequence of the indices of J_i's satisfying $(\lambda_1(t),\ \lambda_2(t))=\vec{\mu}^{(k)}$, $\forall t\in J_i$. Then, $\{J_i,\ i\geq 1\}$ can be partitioned into the (M_0+1) sets, $\{J_{\bar{a}_{k;i}},\ i\geq 1\}$, $0\leq k\leq M_0$. For $0\leq k\leq M_0$, we use the epochs of $(\bar{\mathbf{S}}_i)_{i=1}^2$ in $(J_{\bar{a}_{k;i}})_{i\geq 1}$ to generate $(\bar{\mathbf{S}}_i^{(k)})_{i=1}^2$, in the same manner as we generate $(\mathbf{S}_i^{(k)})_{i=1}^2$ based on $(I_{a_{k;i}})_{i\geq 1}$ in Lemma D.1. Then, based on Lemma D.1 and $(\bar{\mathbf{S}}_i^{(k)})_{i=1}^2$ $(0\leq k\leq M_0)$, we can use the similar argument as in obtaining $\lim_{t\to\infty}\bar{R}(t)$ and show

$$\lim_{t\to\infty}\bar{T}(t) = \frac{\sum_{k=1}^{M_0}\rho_k(\mu_1^{(k)}+\mu_2^{(k)})\phi_{(\mu_1^{(k)},\mu_2^{(k)})}}{\sum_{k=1}^{M_0}\rho_k(\mu_1^{(k)}+\mu_2^{(k)})} \quad \text{a.s.} \tag{D.2}$$

From (D.1) and (D.2), we can see that $\bar{R}(t)-\bar{T}(t)$ converges almost surely to 0 as t grows. Hence, for any positive

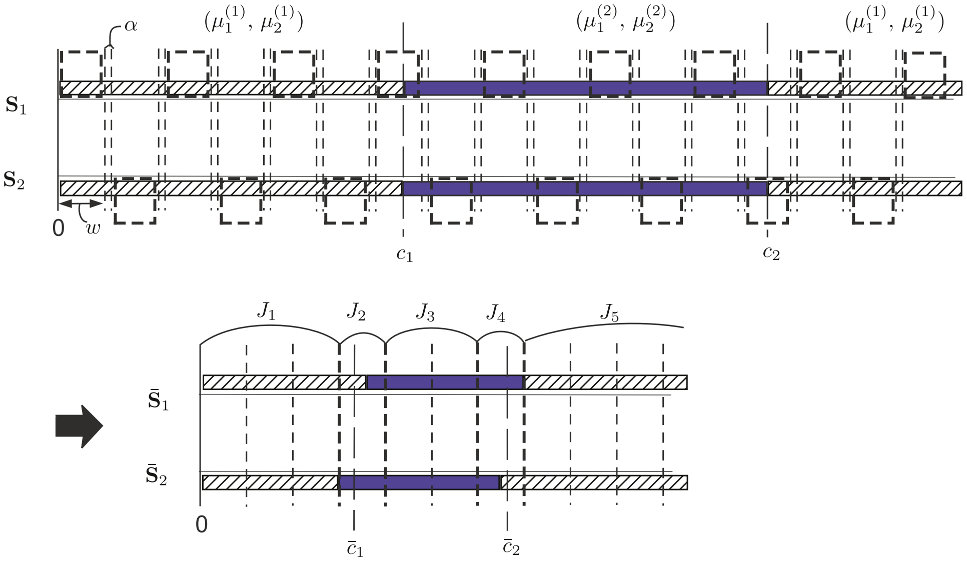

Fig. 12. This figure illustrates a simple case that $(\lambda_1(t),\lambda_2(t))$ can only take either $(\mu_1^{(1)},\mu_2^{(1)})$ or $(\mu_1^{(2)},\mu_2^{(2)})$. The bars filled with slant lines represent the intervals in which $\lambda_i(t)=\mu_i^{(1)}$, and the blue bars represent the intervals in which $\lambda_i(t)=\mu_i^{(2)}$. In this example, J_2 and J_4 are in C. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ϵ, the false alarm probability vanishes as t grows:

$$\lim_{t\to\infty} P_F(t) = \lim_{t\to\infty} \Pr(\overline{R}(t) - \overline{T}(t) > \epsilon) = 0$$

D.2. Miss detection probability

Suppose that $\mathcal{H}_1$ is true and the distribution of $(\mathbf{S}_i)_{i=1}^2$ satisfies the assumptions of the theorem including $R \geq \eta$ a.s. Due to the a.s. convergence of $R(t)$, $R = \lim\inf_{t\to\infty} R(t) \geq \eta$ a.s. is equivalent to

$$\frac{\sum_{k=1}^{M_1} \rho_k(\lambda_{f1}^{(k)} + \lambda_{f2}^{(k)})}{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})} \geq \eta \qquad \text{(D.3)}$$

where $\rho_k \triangleq \lim_{t\to\infty} \rho_k(t)$. In addition, the first assumption of the theorem guarantees that $\overline{\mathbf{S}}_1$ and $\overline{\mathbf{S}}_2$ are independent non-homogeneous Poisson processes. We run BiBGM on $(\mathbf{S}_i)_{i=1}^2$ and let $\overline{R}(t)$ denote the fraction of the matched epochs in the total epochs in $[0,t]$. We also run BiBGM on $(\overline{\mathbf{S}}_i)_{i=1}^2$ and let $\overline{T}(t)$ denote the fraction of the matched epochs in the total epochs in $[0,\lfloor t/2(w+\alpha)\rfloor w]$. First of all, by following exactly the same steps as in the proof of vanishing P_F, we can derive

$$\lim_{t\to\infty} \overline{T}(t) = \frac{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})\phi_{(\lambda_1^{(k)},\lambda_2^{(k)})}}{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})} \quad \text{a.s.} \qquad \text{(D.4)}$$

Let $(c_i)_{i\geq 1}$ denote the increasing sequence of the time points at which $\overrightarrow{\lambda}(t)$ changes, and we partition $[0,\infty)$ into a countable number of subintervals $\{I_i \triangleq [c_{i-1},c_i),\ i \geq 1\}$. For $1 \leq k \leq M_1$, let $(a_{k;i})_{i\geq 1}$ denote the increasing sequence of all the indices of I_i's satisfying $\overrightarrow{\lambda}(t) = \overrightarrow{\lambda}^{(k)}$, $\forall t \in I_i$. We use the epochs of $(\mathbf{S}_i)_{i=1}^2$ in $(I_{a_{k;i}})_{i\geq 1}$ to generate a pair of point processes $(\mathbf{S}_i^{(k)})_{i=1}^2$, as in Lemma D.1. Then, based on Lemmas C.1, D.1, and $(\mathbf{S}_i^{(k)})_{i=1}^2 (1 \leq k \leq M_1)$, we can use the similar argument as in obtaining $\lim_{t\to\infty} \overline{R}(t)$ in the proof of vanishing P_F and derive

$$\liminf_{t\to\infty} \overline{R}(t) \geq \frac{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})\theta_{(\lambda_1^{(k)},\lambda_2^{(k)},\lambda_{f1}^{(k)}+\lambda_{f2}^{(k)})}}{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})} \quad \text{a.s.}$$

where θ is defined in Lemma C.1. For fixed λ_1 and λ_2, $\theta_{(\lambda_1,\lambda_2,\lambda_f)}$ is a strictly increasing function of λ_f, and it decreases to $\phi_{(\lambda_1,\lambda_2)}$ as λ_f decays to 0. Hence, if we define γ as

$$\min_{(\rho_k)_{k=1}^{M_1}} \frac{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})\theta_{(\lambda_1^{(k)},\lambda_2^{(k)},\lambda_{f1}^{(k)}+\lambda_{f2}^{(k)})}}{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})} - \frac{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})\phi_{(\lambda_1^{(k)},\lambda_2^{(k)})}}{\sum_{k=1}^{M_1} \rho_k(\lambda_1^{(k)} + \lambda_2^{(k)})},$$

where the minimization is over $\{(\rho_k)_{k=1}^{M_1} : (D.3) \text{ holds}\}$, then it can be easily seen that γ is strictly greater than 0. Set $\overline{\epsilon} = \frac{1}{3}\gamma$, and let ϵ be an arbitrary number in $(0,\overline{\epsilon}]$. Then, if the condition (D.3) holds,

$$\lim\inf_{t\to\infty}(\overline{R}(t) - \overline{T}(t)) \geq \gamma > 2\epsilon \quad \text{a.s.}$$

Therefore, as long as the condition (D.3) holds, the miss detection probability vanishes as t grows:

$$\lim_{t\to\infty} P_M(t) = \lim_{t\to\infty} \Pr(\overline{R}(t) - \overline{T}(t) < \epsilon) = 0.$$

References

[1] D. Donoho, A. Flesia, U. Shankar, V. Paxson, J. Coit, S. Staniford, Multiscale stepping-stone detection: detecting pairs of jittered interactive streams by exploiting maximum tolerable delay. in: 5th International Symposium on Recent Advances in Intrusion Detection, Lecture Notes in Computer Science, vol. 2516, Zurich, Switzerland, 2002.

[2] P. Kruus, D. Sterne, R. Gopaul, M. Heyman, B. Rivera, P. Budulas, B. Luu, T. Johnson, N. Ivanic, G. Lawler, In-band wormholes and countermeasures in OLSR networks, in: 2nd International Conference on Security and Privacy in Communication Networks (SecureComm 2006), Baltimore, MD, 2006.

[3] T. Chothia, K. Chatzikokolakis, A survey of anonymous peer-to-peer file-sharing, in: Embedded and Ubiquitous Computing Workshops, Lecture Notes in Computer Science, vol. 3823, 2005, pp. 744–755.

[4] J. Ren, J. Wu, Survey on anonymous communications in computer networks, Computer Communications 33 (4) (2010) 420–431.

[5] ITU-T Recommendation G.114, One way transmission time, 2003.

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

[7] P. Peng, P. Ning, D. Reeves, X. Wang, Active timing-based correlation of perturbed traffic flows with chaff packets, in: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops, Columbus, OH, 2005, pp. 107–113.

[8] X. Wang, S. Chen, S. Jajodia, Tracking anonymous peer-to-peer VoIP calls on the internet, in: Proceedings of the 2005 ACM Conference on Computer and Communications Security, Alexandra, VA, 2005.

[9] Y.H. Park, D.S. Reeves, Adaptive watermarking against deliberate random delay for attack attribution through stepping stones, in: Proceedings of the Ninth International Conference on Information and Communications Security (ICICS 2007), 2007.

[10] Y.J. Pyun, Y.H. Park, X. Wang, D. Reeves, P. Ning, Tracing traffic through intermediate hosts that repacketize flows, in: INFOCOM 2007. 26th IEEE International Conference on Computer Communications, IEEE, 2007.

[11] D. Ramsbrock, X. Wang, X. Jiang, A first step towards live botmaster traceback, in: Proceedings of the 11th International Symposium on Recent Advances in Intrusion Detection, Cambridge, MA, 2008.

[12] A. Houmansadr, N. Kiyavash, N. Borisov, RAINBOW: a robust and invisible non-blind watermark for network flows, in: Proceedings of the 16th Annual Network and Distributed System Security Symposium, San Diego, CA, 2009.

[13] A. Houmansadr, N. Borisov, SWIRL: a scalable watermark to detect correlated network flows, in: Proceedings of the 18th Annual Network and Distributed System Security Symposium, San Diego, CA, 2011.

[14] X. Wang, D.S. Reeves, Robust correlation of encrypted attack traffic through stepping stones by flow watermarking, IEEE Transactions on Dependable and Secure Computing 8 (3) (2011) 434–449.

[15] L. Zhang, A. Persaud, A. Johnson, Y. Guan, Stepping Stone Attack Attribution in Non-cooperative IP Networks, Technical Report TR-2005-02-1, Iowa State University, February 2005.

[16] L. Zhang, A. Persaud, A. Johnson, Y. Guan, Detection of stepping stone attack under delay and chaff perturbations, in: Proceedings of The 25th IEEE International Performance Computing and Communications Conference, Phoenix, AZ, 2006.

[17] A. Blum, D. Song, S. Venkataraman, Detection of interactive stepping stones: algorithms and confidence bounds, in: 7th International Symposium on Recent Advance in Intrusion Detection (RAID), Sophia Antipolis, French Riviera, France, 2004.

[18] T. He, L. Tong, Detection of information flows, IEEE Transactions on Information Theory 54 (2008) 4925–4945.

[19] B. Coskun, N. Memon, Efficient detection of delay-constrained relay nodes, in: Proceedings of the 2007 Annual Computer Security Applications Conference, 2007, pp. 353–362.

[20] B. Coskun, N. Memon, Online sketching of network flows for real-time stepping-stone detection, in: Proceedings of the 2009 Annual Computer Security Applications Conference, Washington, DC, 2009, pp. 473–483.

[21] V. Anantharam, S. Verdu, Estimating the directed information to infer causal relationships in ensemble neural spike train recordings, IEEE Transactions on Information Theory 42 (1) (1996) 4–18.

[22] S.H. Sellke, C.-C. Wang, S. Bagchi, N. Shroff, TCP/IP timing channels: theory to implementation, in: The 28th Conference on Computer Communications (INFOCOM 2009), Rio de Janeiro, Brazil, 2009.

[23] C.J. Quinn, T.P. Coleman, N. Kiyavash, N.G. Hatsopoulos, Estimating the directed information to infer causal relationships in ensemble neural spike train recordings, Journal of Computational Neuroscience 30 (1) (2011) 17–44.

[24] V. Paxson, S. Floyd, Wide-area traffic: the failure of Poisson modeling, IEEE/ACM Transactions on Networking 3 (3) (1995) 226–244.

[25] J. Shao, Mathematical Statistics, Springer, 2003.

[26] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, On the self-similar nature of ethernet traffic (extended version), IEEE/ACM Transactions on Networking 2 (1) (1994) 1–15.

[27] M.E. Crovella, A. Bestavros, Self-similarity in World Wide Web traffic: evidence and possible causes, IEEE/ACM Transactions on Networking 5 (6) (1997) 835–846.

[28] Z. Sahinoglu, S. Tekinay, On multimedia networks: self-similar traffic and network performance, IEEE Communications Magazine 37 (1) (1999) 48–52.