Network Topology Inference Based on Timing Meta-Data

Wenbo Du, Member, IEEE, Tao Tan, Fellow, IEEE, Haijun Zhang, Senior Member, IEEE, Xianbin Cao, Fellow, IEEE, Gang Yan, Member, IEEE, and Osvaldo Simeone, Fellow, IEEE

Abstract—A set of low-cost sensors is deployed to infer the network topology of a self-organizing wireless network. The sensors operate in a non-invasive fashion, extracting only the timings of data packets and acknowledgment (ACK) packets from all nodes in a network. The meta-data also reports the source node of each packet, but not the destination nodes or the contents of the packets. A central processor collects the meta-data from the sensors, and the goal is for the processor to infer the network topology based solely on such information. Prior work leveraged causality metrics to identify which links are active. If the data timings and ACK timings of two nodes—say node 1 and node 2, respectively—are causally related, this may be taken as evidence that node 1 is communicating to node 2 (which sends back ACK packets to node 1). This paper starts with the observation that packet losses can weaken the causality relationship between data and ACK timing streams. To obviate this problem, a new Expectation Maximization (EM)-based algorithm is introduced—EM-causality discovery algorithm (EM-CDA)—which treats packet losses as latent variables. EM-CDA iterates between the estimation of packet losses and the evaluation of causality metrics. The method is validated through extensive experiments in wireless sensor networks on the NS-3 simulation platform.

Index Terms—Network topology inference, meta-data, causality metrics, packet loss, expectation maximization.

I. INTRODUCTION

A. Motivation and Overview

INFORMATION about the topology of a device-to-device wireless network, e.g., a sensor network, is essential to implement functionalities such as routing, anomaly detection, and load balance. In recent years, passive monitoring methods that leverage only observations of network traffic have received significant attention, owing to their cost-effectiveness as compared to active methods that probe nodes for information [1], [2]. Passive monitoring methods can be "invasive", implementing packet inspection techniques like demodulation and decryption [3]; or "non-invasive", leveraging only meta-data. Invasive methods can achieve high accuracy, but they require complex sensors and baseband processors. Non-invasive techniques have the advantage of requiring only information about the timings of data packets and acknowledgment (ACK) packets, which are easier to collect and process than invasive methods that require packet decoding. (see Fig. 1). On the flip side, non-invasive methods cannot leverage more advanced tools such as traffic sniffing applications [4]. This paper contributes to the line of work on passive, non-invasive, network topology estimation.

To elaborate, consider, as illustrated in Fig. 1, a central processor that is deployed along with a set of low-cost sensors for the purpose of inferring the network topology. As also assumed in [7], [8], and [9], the sensors operate in a non-invasive fashion, that is, they do not decode any packet. Rather, they are only able to extract meta-data consisting of the timings of data packets and ACK packets from all nodes in a network, as well as the identity of the packets’ sender node. This latter information can be practically obtained by means such as RF-geolocation, energy detection, or direction finding as discussed in [7]. Such meta-data is reported by the sensors to the central node.

The goal of the processor is to infer the network topology based solely on such information. Reference [7] proposed to leverage causality metrics to identify which links are active. The key underlying idea is that, if the data timings and ACK timings of two nodes—say node 1 and node 2, respectively—are causally related, this may be taken as evidence that node 1 is communicating to node 2 (which sends back ACK packets to node 1). The same principle underpins network discovery in fields as diverse as biology and sociology [10], [11], [12], [13].

The causality discovery algorithm (CDA) introduced in [7] was based on Granger causality test (GCT), a mea-
Fig. 1. An example of a wireless device-to-device network with a set of nodes $\mathcal{N} = \{1, 2, 3, 4, 5\}$ and a set of directional links $\mathcal{L} = \{(2, 1), (3, 1), (3, 2), (4, 3), (4, 5)\}$. The central monitor receives meta-data from sensors in the form of data and acknowledgment (ACK) packet timings, based on which it aims to estimate the network topology. Unlike previous work [5], [6], this paper allows for packet losses, making it more challenging to interpret and use meta-data.

sure of causal dependence based on auto-regressive modelling [14]. Asymmetric Granger causality was used in [5], which outperforms GCT at a finer time resolution. Transfer Entropy (TE) was then adopted for CDA in [6]. TE has the advantage of capturing also non-linear causality relationships [15], [16], [17].

This paper starts with the observation that packet losses can weaken the causality relationship between data and ACK timing streams. To obviate this problem, a new Expectation Maximization (EM)-based algorithm is introduced – EM-causality discovery algorithm (EM-CDA) – which treats packet losses as latent variables.

B. Related Work

Active probing is a traditional method used in wireless topology inference, whereby information is collected from neighboring nodes [18], [19]. In such methods, a subset of “privileged” nodes usually performs the probing task [20]. While these schemes can potentially infer accurately the functional network without location information, the energy cost associated with active methods is a critical drawback.

As for passive schemes, references [21] and [22] exploit spectral coherence to infer the network topology, but this approach tends to detect spurious links. In [8] and [9], multivariate Hawkes processes, a parametric formulation of packet arrival statistics, is considered to recover the network topology. These solutions are model-based, and hence operate under strict assumptions on the validity of the model. CDA-based passive topology inference methods currently provide state-of-the-art results for passive topology inference. Apart from the papers reviewed in the previous subsection, the authors of [23] leverage blind source separation to improve the problem caused by interference. The work [24] considers an equidistant missing-data problem based on Granger causality. Nonetheless, the problem of missing observations caused by packet loss is still an open issue, which can result in a significant drop in inference accuracy [5], [6], [7], [25].

C. Main Contributions

Addressing the need for passive topology inference techniques that are robust to packet losses, this paper introduces EM-CDA. The main contributions of this paper can be summarized as follows.

- We formulate the problem of network topology inference as the maximum likelihood problem of estimating existing network links in the presence of latent variables representing packet losses. EM-CDA is derived as a tractable approximation of the resulting EM algorithm. Accordingly, EM-CDA iterates between the estimation of packet losses and the evaluation of causality metrics based on the estimated missing packets.
- EM-CDA is validated through experiments in the wireless network on the NS-3 simulation platform, demonstrating that EM-CDA can improve the detection probability and false alarm probability rate of CDA ranging from 4% to 12% under a variety of practical conditions.

The rest of this paper is organized as follows. The wireless network scenario and system model are described in Section II. The state-of-the-art causality discovery algorithm (CDA) for wireless network topology inference is presented in Section III. EM-CDA scheme is introduced in Section IV. In Section V, numerical results are given to demonstrate the performance of the proposed algorithm. Finally, the paper is concluded in Section VI.

II. SYSTEM MODEL AND PROBLEM SETUP

In this section, we describe the setting under study in which, as illustrated in Fig. 1, a central monitor collects meta-data about packet timings from the nodes of a network in order to infer the network topology. In this paper, unlike [5], [6], we allow packet losses to occur on the communication links. This creates additional challenges in relating the timings of data and control (acknowledgment) packets, motivating the novel estimation algorithm introduced in the next section.

A. Setting

Consider the problem of estimating the topology of a network consisting of a set $\mathcal{N} = \{1, 2, \ldots, N\}$ of $N$ nodes and of a set $\mathcal{L} = \{(i, j)|i, j \in \mathcal{N}\}$ of $M \leq N(N - 1)$ directional links. The presence of a link $(i, j) \in \mathcal{L}$ with $i, j \in \mathcal{N}$ and $i \neq j$ indicates that node $i$ communicates with node $j$.

As in [5] and [6], we assume that a central monitor collects meta-data in the form of transmission timestamps reporting the time instants at which data packets or acknowledgments (ACKs) are sent by each node within a given time window. Only timing meta-data is collected, and hence the monitor is only aware of packet timings, and not of the intended destination of any given packet. Successful transmission of a data packet from one node to another causes the transmission of an ACK from the receiving node to the transmitting one. ACKs are assumed to be much shorter than data packets and not subject to data losses.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
B. Data Transmission and Channel Model

The observation period $T$ is discretized into $K$ equal time slots of duration $T_s = T/K$, which are indexed by integer $k \in \mathcal{K} = \{1, 2, \ldots, K\}$. To describe the timing information recorded by node $i \in \mathcal{N}$, two integer-valued time sequences $Y^D_i[k]$ and $Y^A_i[k]$ are introduced, corresponding to data packets and ACKs, respectively. The data packet timing sample $Y^D_i[k]$ equals the number of data packets sent by node $i$ in time slot $k$. In a similar way, the timing information sample $Y^A_i[k]$ for ACK packets equals the number of ACK packets sent by node $i$ in time slot $k$. We collect the data timing information across all time slots for node $i$ in the $K \times 1$ vector

$$Y^D_i = [Y^D_i[1], Y^D_i[2], \ldots, Y^D_i[K]]^T,$$

(1)

and the ACK timing information in the $K \times 1$ vector

$$Y^A_i = [Y^A_i[1], Y^A_i[2], \ldots, Y^A_i[K]]^T.$$

(2)

The data and ACK timing series for node $i$ can be expressed as the sum of individual contributions corresponding to the distinct communication links stemming from node $i$. To elaborate, we define the per-link binary sequences

$$Y^D_{i,j}[k] = \begin{cases} 1 & \text{if a data packet is sent on link } (i, j) \text{ in time slot } k, \\ 0 & \text{otherwise}, \end{cases}$$

(3)

and

$$Y^A_{i,j}[k] = \begin{cases} 1 & \text{if an ACK is sent on link } (i, j) \text{ in time slot } k, \\ 0 & \text{otherwise}. \end{cases}$$

(4)

Note that the time slot $T_s$ is assumed to be sufficiently small so that no more than one data packet is sent by a node to another node within a single slot. Using the per-link sequences (3)-(4), the per-node observations (1)-(2) can be written as the sums

$$Y^D_i[k] = \sum_{(i,j) \in \mathcal{L}} Y^D_{i,j}[k],$$

(5)

and

$$Y^A_i[k] = \sum_{(i,j) \in \mathcal{L}} Y^A_{i,j}[k].$$

(6)

Importantly, by collecting the sequences (5)-(6), the monitor only has aggregate information regarding the achieving of each node $i$ while not having access to the per-link series $Y^D_{i,j}[k]$ and $Y^A_{i,j}[k]$.

The data packet and ACK timing sequences are related by the ARQ protocol. Let us denote as $\tau_{i,j}[k]$ the delay, measured in the number of time slots, between the transmission of a data packet in time slot $k$ by node $i$ to node $j$ and the transmission of the corresponding ACK packets from node $j$ to node $i$. We also introduce the per-link binary error variable $E_{i,j}[k]$ defined as

$$E_{i,j}[k] = \begin{cases} 1 & \text{if an error occurs on link } (i, j) \text{ in time slot } k, \\ 0 & \text{otherwise}. \end{cases}$$

(7)

With these definitions, we have the equality

$$Y^A_{j,i}[k + \tau_{i,j}[k]] = (1 - E_{i,j}[k]) Y^D_{i,j}[k],$$

(8)

which indicates that an ACK is sent in time slot $k + \tau_{i,j}[k]$ on link $(j, i)$, i.e., $Y^A_{j,i}[k + \tau_{i,j}[k]] = 1$, when a data packet is sent in time slot $k$ on the reverse link $(i, j)$, i.e., $Y^D_{i,j}[k] = 1$ and an error does not occur on link $(i, j)$, i.e., $E_{i,j}[k] = 0$. The $\tau_{i,j}[k]$ may severely vary across links and time slots, and it is unknown to the monitor.

C. Topology Inference

The timing information sequences $\{Y^D_i|\forall i \in \mathcal{N}\}$ and $\{Y^A_i|\forall i \in \mathcal{N}\}$ in (1)-(2) collected from all nodes are used by the monitor to infer the topology, which is defined by set of links $\mathcal{L}$. The links set $\mathcal{L}$ can be equivalently also described by the adjacency matrix $A = [a_{i,j}|\forall i, j \in \mathcal{N}]$ with entries

$$a_{i,j} = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{L}, \\ 0 & \text{otherwise}. \end{cases}$$

(9)

Therefore, the goal of the monitor is to use sequence $\{Y^D_i|\forall i \in \mathcal{N}\}$ and $\{Y^A_i|\forall i \in \mathcal{N}\}$ to produce an estimate $\hat{A}$ of the adjacency matrix $A$, or equivalently an estimate $\hat{\mathcal{L}}$ of the link set $\mathcal{L}$.

III. CAUSALITY-BASED TOPOLOGY ESTIMATION

In this section, we review the Causality Discovery Algorithms (CDAs) introduced in [5], [7], [25], and [6] wherein links are included in the estimated set $\hat{\mathcal{L}}$ based on measures of causal dependence between data and ACK sequences of two nodes.

A. Causality Discovery Algorithm

In CDA schemes, the monitor estimates a measure of causal dependence $\Phi(Y^D_i \rightarrow Y^A_j)$ between sequences $Y^D_i$ and $Y^A_j$ for each pair of nodes $i$ and $j$. The measure $\Phi(Y^D_i \rightarrow Y^A_j)$ quantifies the degree to which the future of sequence $Y^A_j$ can be predicted based on the past of sequence $Y^D_i$. A link $(i, j)$ is added to the estimated set $\hat{\mathcal{L}}$ if the measure $\Phi(Y^D_i \rightarrow Y^A_j)$ is larger than some threshold $\theta_{i,j}$. This condition can be equivalently expressed as

$$\hat{a}_{i,j} = \begin{cases} 1 & \text{if } \Phi(Y^D_i \rightarrow Y^A_j) > \theta_{i,j}, \\ 0 & \text{otherwise}. \end{cases}$$

(10)

The rationale for this decision rule is that, if link $(i, j)$ exists, then by (8) data packets from node $i$ cause ACKs from node $j$, assuming that there are no errors. This, in turn, ideally contributes to increasing the causal dependence measure $\Phi(Y^D_i \rightarrow Y^A_j)$.

We now discuss specific choice for the causal dependence measure $\Phi(Y^D_i \rightarrow Y^A_j)$.

B. Causality Metrics

Granger causality (GC) is a standard measure of causal dependence that is based on linear prediction. Given two time sequences $Y^D_i$ and $Y^A_j$, GC evaluates the extent to which omitting the past of time series $Y^D_i[k]$ increases the prediction...
error for sequence $Y_j^A[k]$ when prediction is based on a linear $R$-order autoregressive (AR) model. Formally, GC uses the available observations $Y_i^D$ and $Y_j^A$ to fit separately two models, namely

$$Y_j^A[k] = \sum_{r=1}^{R} a_1 r Y_j^A[k - r] + \sum_{r=1}^{R} a_2 r Y_i^D[k - r] + \varepsilon_k, \quad (11)$$

and

$$Y_j^A[k] = \sum_{r=1}^{R} b_r Y_j^A[k - r] + \eta_k, \quad (12)$$

by optimising over parameters $\{a_1, a_2, \ldots, a_R\}^R_r$ and $\{b_1, b_2, \ldots, b_R\}^R_r$ via least squares minimization. In (11)-(12), the quantities $\varepsilon_k$ and $\eta_k$ represent the prediction residuals. The prediction residuals $\varepsilon_k$ in (11) account for prediction errors accrued on the ACK sequence $Y_j^A[k]$ when the past of data packet sequence $Y_i^D[k]$ is known; while the residuals $\eta_k$ in (12) are obtained when prediction can only use the past sample for the ACK sequence $Y_j^A[k]$ itself. The GC-based measure is given by [7]

$$\Phi_{GC}(Y_j^D \rightarrow Y_j^A) = \frac{\sum_{k=1}^{H} |\eta_k|^2 - \sum_{k=1}^{H} |\varepsilon_k|^2/R}{\sum_{k=1}^{H} |\varepsilon_k|^2/(K - 3R - 1)}, \quad (13)$$

which is large when the sum-residual $\sum_{k=1}^{H} |\varepsilon_k|^2$ is larger than $\sum_{k=1}^{H} |\eta_k|^2$, where $H = K - R$. GC was used in [7] and [25] for topology estimation.

Transfer entropy (TE) is an information-theoretic causality measure that does not assume a linear relation between sequences $Y_i^D$ and $Y_j^A$ as GC. To introduce it, let us define as $I(A; B|C)$ the conditional mutual information of random variable $A$ and $B$ given $C$, which is defined as

$$I(A; B|C) = \mathbb{E} \left[ \log_2 \frac{p(A, B|C)}{p(A|C)} \right], \quad (14)$$

where the expectation is taken over the joint distribution $p(A, B|C)$ and $p(A|C)$. With these definitions, the TE is defined as [15]

$$\Phi_{TE}(Y_j^D \rightarrow Y_j^A) = I(Y_j^A[k]; Y_i^D[k - 1 : k - s]|Y_j^A[k - 1 : k - r]), \quad (15)$$

where $s$ and $r$ are fixed integers; $Y_j^A[k - 1 : k - r] = \{Y_j^A[k - 1], Y_j^A[k - 2], \ldots, Y_j^A[k - r]\}$ and $Y_i^D[k - 1 : k - s] = \{Y_i^D[k - 1], Y_i^D[k - 2], \ldots, Y_i^D[k - s]\}$ denote windows of past samples for $Y_j^A$ and $Y_i^D$, respectively. In practice, the TE is estimated using available data sequences $Y_j^A$ and $Y_i^D$. The TE was used for topology estimation in [5] and [6].

C. Setting the Threshold

The threshold $\theta_{i,j}$ in (10) can be set via a permutation test [26]. Accordingly, one considers a statistical significance test in which the null hypothesis corresponds to the assumption that the two sequences $Y_i^D$ and $Y_j^A$ are not causally related. To obtain the distribution of the causality metrics $\Phi(Y_j^D \rightarrow Y_j^A)$ under the null hypothesis, $S$ random permutations of the sequences are obtained by considering permutations of the observed sequences. Through random shuffling, the permutations $Y_i^D$ and $Y_j^A$ of sequences $Y_i^D$ and $Y_j^A$ are produced, with $s \in \{1, 2, \ldots, S\}$. Then, the causality metrics $\Phi(Y_i^D \rightarrow Y_j^A)$, are evaluated for each permutation, with $s \in \{1, 2, \ldots, S\}$, and the empirical distribution of the samples $\Phi(Y_i^D \rightarrow Y_j^A)_{s=1}^{S}$ is obtained. Finally, the threshold $\theta_{i,j}$ is set as the $(1-\alpha)$-quantile of the empirical distribution, when $\alpha \in [0,1]$ is a fixed false alarm probability. In other words, one can sort $\Phi(Y_i^D \rightarrow Y_j^A)_{s=1}^{S}$ in descending order, and observe if $\Phi(Y_i^D \rightarrow Y_j^A)$ is contained within the top $(1-\alpha)$ of this list. If not, the presence of the link $(i,j)$ is rejected at the $\alpha$ significance level.

IV. EM-BASED TOPOLOGY ESTIMATION

The CDA schemes reviewed in the previous section were devised under the assumption that there are no packet losses [5], [6], [7]. As we argue in Sec. IV-A, packet losses tend to make the CDA test (10) unreliable, since the causality metrics are decreased in the presence of packet losses due to the missed association between data and ACK sequences erased by lost data packets. To address this challenge, in this section, we introduce the EM-based CDA, which models packet losses using latent random variables.

A. Impact of Packet Losses on CDA

In order to gain insights into the impact of packet losses on the performance of CDA, we now consider an IEEE 802.11 ad-hoc network simulated with NS-3, and evaluate the GC metric (13) for a given link $(i,j)$ in the presence and absence of packet losses. Details of the experimental setting can be found in Sec. V-B. Fig. 2 reports the GC metric evaluated with losses as a function of the corresponding metric evaluated in a lossless scenario under the same conditions. Different points correspond to distinct links in the set $L$. The figure confirms that the GC metric tends to be decreased by packet losses, making CDA methods potentially ineffective.
Table I

Notation Table of EM-CDA

| Indices | \( N \) | number of nodes |
|---------|--------|----------------|
| \( K \) | number of time slots |
| \( N' \) | nodes set (\( N = \{1, \ldots, N'\} \)) |
| \( K' \) | time slots set (\( K = \{1, \ldots, K'\} \)) |

| Parameters | \( a_{i,j} \) | topology of link \((i,j)\) |
|------------|-----------------|--------------------------------|
| \( L_{i,j} \) | packet loss probability on link \((i,j)\) |
| \( R_{i,j} \) | transmission probability on link \((i,j)\) |
| \( \tau_{i,j} \) | delay between ACK and packet transmission on link \((i,j)\) |

| \( A \) | network topology (\( A = \{a_{i,j} \mid vi, j \in N'\} \)) |
| \( L \) | error rate set in the network (\( L = \{L_{i,j} \mid vi, j \in N'\} \)) |
| \( R \) | transmission rate set in the network (\( R = \{R_{i,j} \mid vi, j \in N'\} \)) |
| \( T \) | ACK delay set in the network (\( T = \{\tau_{i,j} \mid vi, j \in N'\} \)) |
| \( \Theta \) | the set of all unknown parameters (\( \Theta = \{A, L, R, T\} \)) |

| Random variables | \( Y^D_i[k] \) | observation of data packet sequences from node \( i \) at time slot \( k \) |
|------------------|-----------------|--------------------------------|
| \( Y^A_i[k] \) | observation of ACK sequences from node \( i \) at time slot \( k \) |
| \( D_{i,j}[k] \) | data packet sent from nodes \( i \) to node \( j \) at time slots \( k \) |
| \( E_{i,j}[k] \) | packet loss event for data packet sent from nodes \( i \) to node \( j \) at time slot \( k \) |
| \( Y \) | observations set (\( Y = \{Y^D_i[k], Y^A_i[k] \mid vi \in N', k \in K'\} \)) |
| \( D \) | the latent data packet set (\( D = \{D_{i,j}[k] \mid vi, j \in N', k \in K'\} \)) |
| \( E \) | the latent packet loss set (\( E = \{E_{i,j}[k] \mid vi, j \in N', k \in K'\} \)) |
| \( Z \) | the set of all latent variables (\( Z = \{D, E\} \)) |

B. Parametric Model With Latent Variables

As we have demonstrated in Sec. IV-A, the direct application of causality metrics is significantly affected by packet losses. The general idea underlying the proposed approach is to estimate the packet losses in order to improve the estimation of causality metrics. However, this estimate is made challenging by the fact that the central monitor has only access to the total number of packet losses, and neither the communication link nor the time slot of each lost packet is known. Furthermore, there are additional unknown parameters, such as the transmission rate and the packet loss rate. We address these challenges in a principled way by formulating the problem in terms of hidden variables. Accordingly, we adopt the Expectation Maximization (EM) algorithm, a method that iterates between an estimate of the hidden variables in the E-step and an estimate of the topology in the M-step. The notations of EM-CDA are listed in Table I.

The EM-CDA scheme is based on the idea of formulating the problem of topology inference as the MLE of the adjacency matrix \( A \) in the presence of latent variables describing packet losses. To elaborate, let \( Y = \{Y^D_i[k], Y^A_i[k] \mid vi \in N'\} \) be the observations. We also introduce two sets of latent variables. The first, \( D = \{D_{i,j}[k] \mid vi, j \in N', k \in K'\} \), contains variables \( D_{i,j}[k] \) for all pairs of nodes \( i \) and \( j \) and time slots \( k \), such that

\[
D_{i,j}[k] = \begin{cases} 
1 & \text{if a data packet is sent on link } (i, j) \text{ in time slot } k, \\
0 & \text{otherwise}.
\end{cases}
\]

The second, \( E = \{E_{i,j}[k] \mid vi, j \in N', k \in K'\} \), contains the packet loss variables defined in (7). Note that the true value of the latent variables \( D_{i,j}[k] \) and \( E_{i,j}[k] \) are undefined for links not in set \( L \). The set \( Z = \{E, D\} \) defines the latent variables. Overall, we have observations \( Y \) and latent variables \( Z \).

We now define a parametric model that specifies the point distribution \( p(Y, Z | \Theta) \) of observations \( Y \) and latent variables \( Z \) as a function of a set of parameters, \( \Theta \). Set \( \Theta \) includes the adjacency matrix \( A \), which is the quantity of interest, as well as some nuisance parameters to be introduced next. We emphasize that the probabilistic model \( p(Y, Z | \Theta) \) does not generally describe the ground-truth data generation mechanism, which is unknown. Rather, it amounts to a set of assumptions made in order to develop the proposed topology estimation algorithm.

The parametric model, \( p(Y, Z | \Theta) = p(Z | \Theta)p(Y | Z, \Theta) \), depends on the set of unknown parameters \( \Theta = \{A, L, R, T\} \), where matrices \( \{L, R, T\} \) are nuisance parameter matrices representing error rate, transmission rate, and ACK delay on each link, respectively. Let us define as \( E(A) \) the set of coordinates of non-zero entries of the adjacency matrix \( A \), that is, the estimated links given matrix \( A \). To start, we assume that variables \( \{E_{i,j}[k], D_{i,j}[k]\} \) corresponding to different link \((i, j)\) are independent, i.e.,

\[
p(Z | \Theta) = \prod_{(i,j) \in E(A)} p(E_{i,j}, D_{i,j} | \Theta), \tag{17}
\]

where we have the sequences \( E_{i,j} = \{E_{i,j}[k]\}_{k=1}^{K} \) and \( D_{i,j} = \{D_{i,j}[k]\}_{k=1}^{K} \). Focusing now on sequences \( E_{i,j} \) and \( D_{i,j} \), we assume the joint distribution

\[
p(E_{i,j}, D_{i,j} | \Theta) = \prod_{k \in K} \left[ p(D_{i,j}[k] | D_{i,j}[1 : k - 1], E_{i,j}[1 : k - 1]; \Theta) \times p(E_{i,j}[k] | D_{i,j}[1 : k], E_{i,j}[1 : k - 1]; \Theta) \right]
\]

\[
= \prod_{k \in K} \left[ p(D_{i,j}[k] | R_{i,j}) p(E_{i,j}[k] | D_{i,j}[k] ; L_{i,j}) \right], \tag{18}
\]

where the first equality follows from the chain rule of probability, and the second is a consequence of the following two assumptions. First, we assume that an error on a link \((i, j)\), indicated by \( E_{i,j}[k] = 1 \), occurs with probability \( L_{i,j} \) if a transmission took place on the same link, i.e., if \( D_{i,j}[k] = 1 \). That is, the Bernoulli distribution of \( E_{i,j}[k] \) takes the value 1 with probability \( L_{i,j} \) if \( D_{i,j}[k] = 1 \), and with a probability 0 if \( D_{i,j}[k] = 0 \). This is expressed with the conditional distribution

\[
p(E_{i,j}[k] | D_{i,j}[k] ; L_{i,j}) = \begin{cases} 
L_{i,j}^{E_{i,j}[k]}(1 - L_{i,j})^{1 - E_{i,j}[k]} & \text{if } D_{i,j}[k] = 1, \\
1 - E_{i,j}[k] & \text{otherwise}.
\end{cases}
\]

Second, transmissions occur independently of previous transmissions and errors with probability \( R_{i,j} \). Similarly, the Bernoulli distribution of \( D_{i,j} \) is formulated as below.

\[
p(D_{i,j}[k] | R_{i,j}) = R_{i,j}^{D_{i,j}[k]}(1 - R_{i,j})^{1 - D_{i,j}[k]}, \tag{19}
\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
We emphasize that the conditional distribution (20) entails a significant approximation, since transmissions in many network scenarios encompass also retransmission of previous, erroneously received, packets. The Bayesian network that describes the assumed model for the latent variables is shown in Fig. 3.

To fully specify the parametric model $p(Y, Z; \Theta) = p(Z; \Theta)p(Y | Z; \Theta)$ we need to describe also the distribution $p(Y | Z; \Theta)$. In this regard, the observations $Y$ are assumed to be a function $f(D, E|A, T)$ of the latent variables $D$ and $E$ that is parameterized by the adjacency matrix $A$ and the matrix of delays $T$. Accordingly, the distribution of $Y$ conditioned on $Z$ is given by

$$p(Y | Z ; \Theta) = p(Y | D, E; A, T) = \delta(Y - f(D, E|A, T)), \quad (21)$$

where $\delta(\cdot)$ is the Kronecker delta function. Function $f(D, E|A, T)$ is defined as follows. Since the number of packets observed from a node $i$ equals the sum of the numbers of packets sent to other nodes $j$ with $(i,j) \in \mathcal{E}(A)$ in the given time slot, we have the equality

$$Y_i^D[k] = \sum_{j: (i,j) \in \mathcal{E}(A)} D_{i,j}[k]. \quad (22)$$

This is reflected by the Bayesian network in Fig. 3. Similarly, the number of ACKs reported by a node $j$ is equal to the sum of the numbers of ACKs sent to other nodes $i$ with $(i,j) \in \mathcal{E}(A)$. Defining the model parameter $\tau_{i,j}$ as the delay between ACK and packet transmission on link $(i,j)$, we thus assume the equality

$$Y_j^A[k] = \sum_{i: (i,j) \in \mathcal{E}(A)} (1 - E_{i,j}[k - \tau_{i,j}]) D_{i,j}[k - \tau_{i,j}]. \quad (23)$$

It is recalled that, while the actual unknown time delays $\tau_{i,j}[k]$ in (8) may depend on the time slot $k$, the parameters $\tau_{i,j}$ in (23), which are collected in matrix $T$, are assumed to be static in order to facilitate estimation. Overall, equalities (22)-(23) define function $f(D, E|A, T)$ and hence distribution (21).

C. EM-Based Causality Discovery Algorithm (EM-CDA)

Given the likelihood $p(Y, Z; \Theta)$ of the complete data $(Y, Z)$, EM-CDA aims to address the MLE problem

$$\max_{\Theta} \left\{ p(Y | \Theta) = E_{p(Z | \Theta)} \left[ p(Y | Z; \Theta) \right] \right\} \quad (24)$$

via EM. Accordingly, EM-CDA updates the current estimate $\Theta$ across a number of iteration, producing a sequence of iterates $\Theta^{(1)}, \Theta^{(2)}, \ldots, \Theta^{(n)}$. At each iteration $n$, EM first performs the expectation step (E-step), which evaluates the expected value

$$Q(\Theta | \Theta^{(n)}) = E_{p(Z | Y; \Theta^{(n)})} \left[ \log p(Y, Z | \Theta) \right] \quad (25)$$

of the complete log-likelihood $\log p(Y, Z | \Theta)$ with respect to the current posterior distribution $p(Z | Y; \Theta^{(n)})$. Then, the maximization step (M-step) is carried out, wherein the next update is obtained as

$$\Theta^{(n+1)} = \arg\max_{\Theta} Q(\Theta | \Theta^{(n)}). \quad (26)$$

A direct application of EM to the model $p(Y, Z; \Theta)$ described in the previous subsection is computationally infeasible. To obtain a scalable solution, the proposed EM-CDA approximates the E-step using Monte Carlo sampling, and the M-step via CDA (see Sec. III). The resulting algorithm can be viewed as an iterative generalization of CDA, wherein estimates of packet losses are accounted for in the estimates of the causality metrics in order to address the issue described in Sec. IV-A. The E-step and M-step are repeated until convergence, and the convergent topology is produced as the final result. We detail both E-step and M-step in the rest of this section, and the overall EM-CDA is described in Algorithm 1.

Algorithm 1 EM-CDA

**Require:** The observations $Y$, learning rate sequences $\{\gamma^{(n)}\}$, number of samples $M$, maximum estimated delay $\tau_{\text{max}}$, and significance level $\alpha$;

**Ensure:** Matrix of inferred communication links $\hat{A}$;

1: **Initialization:** Initialize $L^{(0)}$ and $R^{(0)}$ with 0-1 uniform distribution; the adjacency matrix $A^{(0)}$ to have every entry equal to one; $T^{(0)}$ by (31) using $Y$; $n$ to 0;
2: **while** $\Theta^{(n)}$ has not converged **do**
3: **Generate** samples based on (28)-(30);
4: **Update** $\Theta^{(n+1)}$ based on (31)-(35);
5: $n \leftarrow n + 1$;
6: **end while**
7: **Obtain** $\hat{A} = A^{(n)}$.

D. Expectation Step (E-Step)

At iteration $n$, given the current parameters $\Theta^{(n)}$, the E-step aims at generating $M$ samples $\{Z_1^{(n)}, \ldots, Z_M^{(n)}\}$ from the posterior distribution $p(Z | Y; \Theta^{(n)})$. With such samples, the function $Q(\Theta | \Theta^{(n)})$ in (25) is approximated via the
stochastic estimate [27]

\[
Q\left(\Theta | \Theta^{(n)}\right) = \left(1 - \gamma^{(n)}\right) Q\left(\Theta | \Theta^{(n-1)}\right) + \frac{\gamma^{(n)}}{M} \sum_{m=1}^{M} \log p\left(Y, Z_{m}^{(n)} | \Theta\right),
\]

(27)

where \(\gamma^{(n)} \in [0, 1]\) is a learning rate.

In order to generate the samples \(Z_{m}^{(n)} \sim p\left(Z | Y; \Theta^{(n)}\right)\) for \(m = 1, \ldots, M\), we apply Gibbs sampling. Gibbs sampling generates the samples \(Z_{m}^{(n)}\) sequentially over index \(m = 1, \ldots, M\) by drawing samples from the conditional probabilities of one variable in \(Z\) given all other variables in \(Z\) [28]. Accordingly, each sample \(Z_{m}^{(n)} = \{D_{i,j,m}^{(n)} | k, E_{i,j,m}^{(n)} | k | i,j \in \mathcal{N}, k \in \mathcal{K}\}\) is generated as follows.

Using the notations \(Z_{i,j}[k] = (D_{i,j}[k], E_{i,j}[k])\) and \(Z_{-(i,j)}[-k] = \{D_{i',j',k'} | (i', j', k') \neq (i,j)\}\), for each pair of variables \(Z_{i,j}[k]\), we sample from the posterior \(p\left(Z_{i,j}[k] | Z_{-(i,j)}[-k], Y; \Theta^{(n)}\right)\) given all other variables. This can be evaluated as (28), shown at the bottom of the page, where the first equality follows from d-separation based on the Bayesian network in Fig. 3 (see, e.g., [29]), and \(p\left(Z_{i,j}[k] | \Theta^{(n)}\right)\) is given by the product of (19) and (20) as

\[
p\left(Z_{i,j}[k] | \Theta^{(n)}\right) = p\left(D_{i,j}[k] | R_{i,j}\right) p\left(E_{i,j}[k] | D_{i,j}[k] ; l_{i,j}\right).
\]

(29)

We now leave with the problem evaluating the distribution \(p\left(Y_{i}^{D}[k], Y_{j}^{A}[k + \tau_{i,j}] | Z_{i,j}[k]; \Theta^{(n)}\right)\). According to (22)-(23), it is given by the probability of that \(Y_{i}^{D}[k] - D_{i,j}[k]\) packets are sent by node \(i\) to other nodes except \(j\) at time \(k\), and that \(Y_{j}^{A}[k + \tau_{i,j}] - D_{i,j}[k](1 - E_{i,j}[k])\) ACKs are sent by node \(j\) to other nodes except \(i\) at time \(k + \tau_{i,j}\). Therefore, by (18)-(20) we have

\[
p\left(Y_{i}^{D}[k], Y_{j}^{A}[k + \tau_{i,j}] | Z_{i,j}[k]; \Theta^{(n)}\right) = p\left(Y_{i}^{D}[k] | Z_{i,j}[k]; \Theta^{(n)}\right) p\left(Y_{j}^{A}[k + \tau_{i,j}] | Z_{i,j}[k]; \Theta^{(n)}\right)
\]

\[
= \text{Bin}\left(Y_{i}^{D}[k] - D_{i,j}[k] | \{R_{i,j}\}_{i,j} \in \mathcal{E}(A)\right)
\]

\[
\times \text{Bin}\left(Y_{j}^{A}[k + \tau_{i,j}] - D_{i,j}[k](1 - E_{i,j}[k]) | \{R_{i,j}\}_{i,j} \in \mathcal{E}(A)\right),
\]

(30)

where we denote as Bin\((y; n, p)\) the probability mass function of a sum of \(L\) independent Bernoulli random variables, with each \(i\)th random variables having probability \(p_{i}\) of being equal to 1. The second equality in (30) is derived by the independent assumption stated in (19)-(20), since \(Y_{i}^{D}[k]\) can be regarded as the composition of the independent variables \(Y_{i}^{D}[k]\), and \(Y_{j}^{A}[k]\) can be regarded as the composition of independent random variables \(Y_{j}^{A}[k]\).

**E. Maximization Step (M-Step)**

Given \(Y\) and samples generated \(\{Z_{m}^{(n)}\}_{m=1}^{M}\) in the E-step, the M-step aims at updating parameters \(\Theta\). The discrete parameters \(T\) and \(A\) are updated by generalizing the CDA approach described in Sec. III to include the estimate of delays. The continuous parameters \(R\) and \(L\) are then updated by finding the stationary points of the objective function of \(Q\left(\Theta | \Theta^{(n)}\right)\) in (27).

For each sample \(Z_{m}^{(n)}\), we define as \(Y_{i,m}^{D,(n)} = \{Y_{i}^{D}[k] \forall k \in \mathcal{K}\}\) the estimated data packet sequence for node \(i\); and as \(Y_{j,m}^{A,(n)} = \{Y_{j}^{A}[k] + \sum_{(i,j) \in \mathcal{E}(A)} E_{i,j}[k - \tau_{i,j}] | \forall k \in \mathcal{K}\}\) the estimated ACK sequence for node \(j\). To update the delay matrix for \(T_{m}^{(n)} = \{t_{i,j,m} | \forall i,j \in \mathcal{N}\}\), we obtain the sequences \(Y_{i,m}^{D,(n)} = \{Y_{i,m}^{D,(n)}[k + \tau]\}_{k=1}^{K}\) by shifting backward in time by \(\tau\) steps the sequences \(Y_{i,m}^{D,(n)}\). Then, the causal dependence measure \(\Phi(Y_{i,m}^{D,(n)}, \tau \rightarrow Y_{j,m}^{A,(n)})\) is calculated using (13) or (15) for a range of values \([1, \tau_{max}]\) to obtain the estimate

\[
\tau_{i,j}^{(n)} = \arg \max_{\tau} \Phi\left(Y_{i,m}^{D,(n)}, \tau \rightarrow Y_{j,m}^{A,(n)}\right).
\]

(31)

Furthermore, using (10), the estimated topology entries \(a_{i,j,m}^{(n)}\) of the adjacency matrix \(A_{m}^{(n)}\) are given by

\[
a_{i,j,m}^{(n)} = \begin{cases} 1 & \text{if } \Phi(Y_{i,m}^{D,(n)} \rightarrow Y_{j,m}^{A,(n)}) > \theta_{i,j,m}^{(n)} \\ 0 & \text{otherwise}, \end{cases}
\]

(32)

where \(\theta_{i,j,m}^{(n)}\) is a threshold. Then, we set

\[
a_{i,j}^{(n+1)} = \begin{cases} 1 & \text{if } \sum_{m=1}^{M} a_{i,j,m}^{(n)} \geq \frac{M}{2} \\ 0 & \text{otherwise}, \end{cases}
\]

(33)

that, is an edge \((i,j)\) is included in the set \(\mathcal{E}(A)^{(n)}\) of the majority of tests (32) set \(a_{i,j,m}^{(n)} = 1\).

Finally, setting the partial derivatives of \(Q\left(\Theta | \Theta^{(n)}\right)\) in (27) with respect to \(R\) and \(L\) to zero, respectively, the updated
V. NUMERICAL RESULTS

In this section, numerical results are provided to demonstrate the performance of the proposed EM-CDA scheme as corresponds to the conventional CDA methods reviewed in Sec. III [5], [6], [7]. We first consider a toy example in which we can evaluate the impact of the approximations adopted in the derivation of EM-CDA via an exact implementation of EM. We refer to this scheme as EM-ES. To enable EM-ES over the exponential number of possible choices \( \mathbf{A} \), we consider a small network with \( N = 4 \) nodes that is allowed to follow the same model adopted for the derivation of EM as explained in Sec. IV. In the next subsection, we will consider a more realistic scenario in NS-3.

Half of the links are randomly selected to be active; the ground-truth average transmission rate \( R_{i,j}^{\ast} \) for all active links is set as 0.1; the average packet loss rate \( L_{i,j}^{\ast} \) for all links is set to 0.05 or 0.5; and the ground-truth delay \( \tau_{i,j}^{\ast} \) are set to 1 time slot. We simulate the network for 5000 time slots. In the E-steps of both methods, the number of samples is set as \( M = 30 \). In M-step, GCT or TE is adopted as the causality discovery algorithm in EM-CDA. The significance level \( \alpha \) in (10) is set to 0.05 as in [6]. All the results are generated in 20 trials with different random initial values.

The probability of false alarm, \( p_{FA} \), and the probability of detection, \( p_{D} \), are adopted to measure the performance of topology inference. These metrics are defined as

\[
P_{FA} = \frac{FP}{FP + TN},
\]

and

\[
P_{D} = \frac{TP}{TP + FN},
\]

where TP denotes the number of correctly detected existing links, FN denotes the number of missed existing links, TN denotes the number of correctly detected missing links, and FP denotes the number of incorrectly detected missing links.

Fig. 4. Probability of false alarm \( P_{FA} \) and probability of detection \( p_{D} \) for topology inference versus the number of EM iterations for the ideal EM-ES scheme and for EM-CDA with GCT and TE causality metrics: iterations in the case of (a) \( L_{i,j}^{\ast} = 0.05 \), and (b) \( L_{i,j}^{\ast} = 0.5 \).

Fig. 5. (a) Probability of detection \( p_{D} \) and (b) probability of false alarm \( P_{FA} \) for topology inference versus the length of observation in time slots for CDA and EM-CDA with GCT and TE causality metrics.

### A. Small-Scale Experiments

In this subsection, we compare EM-CDA with an implementation of EM to address the MLE problem (24) that applies the exhaustive search (ES) method in the M-step to maximize the function \( Q(\mathbf{\Theta}|\mathbf{\Theta}^{(n)}) \) over variables \( \mathbf{A} \) and \( \mathbf{T} \). We refer to

\[
R^{(n+1)}_{i,j} = \left( 1 - \gamma^{(n)} \right) R_{i,j}^{(n)} + \frac{\gamma^{(n)}}{MK} \sum_{m=1}^{M} \sum_{k \in K} D_{i,j,m}^{(n)} [k],
\]

\[
\forall (i, j) \in \mathcal{E}(\mathbf{A}^{(n)}),
\]

and

\[
L^{(n+1)}_{i,j} = \left( 1 - \gamma^{(n)} \right) L_{i,j}^{(n)} + \frac{\gamma^{(n)}}{\sum_{m=1}^{M} \sum_{k \in K} E_{i,j,m}^{(n)} [k]} \sum_{m=1}^{M} \sum_{k \in K} D_{i,j,m}^{(n)} [k],
\]

\[
\forall (i, j) \in \mathcal{E}(\mathbf{A}^{(n)}).
\]
Fig. 6. (a) Probability of detection $P_D$ and (b) probability of false alarm $P_{FA}$ for topology inference versus the packet loss rate for CDA and EM-CDA with GCT and TE causality metrics.

**TABLE II**

| Parameter              | Value             |
|------------------------|-------------------|
| Area size              | 100 m²            |
| Carrier frequency $f_0$| 2.412 GHz         |
| Data packet size       | 1024 Bytes        |
| MAC ACK size           | 36 Bytes          |
| Channel packet loss rate | varies     |
| Transmission rate      | varies            |
| Simulation duration $T$| 60 s              |
| Time slot duration $T_s$| 1.5 ms          |

Fig. 4 shows the probabilities $P_{FA}$ and $P_D$ across the EM iterations. EM-ES is seen to obtain the optimal solution, yielding the ideal case $P_{FA} = 0$ and $P_D = 1$, in a single iteration, while EM-CDA with both GCT and TE requires more iterations, but it is able to converge to the optimal solution. Furthermore, the number of required EM iterations for the performance of EM-CDA increases as the ground-truth average loss rate $L^*_{i,j}$ increases, because, as discussed in Sec. IV-A, unreliable observations provide missing and spurious information that needs to be compensated for by refining the estimates of the latent variables.

**B. Simulations on NS-3**

In this subsection, we test EM-CDA in different wireless scenarios simulated on NS-3 using the parameters in Table II. The system consists of $N$ nodes randomly and uniformly distributed within a $10 \text{ m} \times 10 \text{ m}$ area that follow the an IEEE 802.11 ad-hoc protocol operating at carrier frequency $f_0 = 2.412$ GHz. Omnidirectional antennas are used at the nodes, with path-loss, log-normal shadowing, and thermal noise accounted for as in [30]. The simulation lasts $T = 60$ s, and the time slot duration is $T_s = 1.5$ ms. The offered traffic for each link is 1 Mbps, with a data packet size of 1024 Bytes and an ACK size of 36 Bytes. If not stated otherwise, we set $N = 12$ nodes, fraction of active links 0.5, number of observed time slots 5000, and average packet loss rate 0.3.

Table III compares state-of-the-art wireless topology inference methods with EM-CDA under the setting described above. GCT and TE are two state-of-the-art CDA methods as mentioned in Sec. III. The neural network (NN)-based method proposed in [25] extracts time-related features from the observed data, and turns the problem of topology inference into one of binary classification. The results in the table show that NN performs better than GCT and TE in terms of probability of false alarm, $P_{FA}$, but worse than TE in terms of probability of detection, $P_D$. Importantly, EM-CDA outperforms all other methods, including NN, in terms of both $P_D$ and $P_{FA}$, bringing a $5\% - 8\%$ improvement over different CDA benchmarks in this scenario.

**TABLE III**

| Metrics            | GCT   | TE   | NN    | EM-CDA-GCT | EM-CDA-TE |
|--------------------|-------|------|-------|------------|-----------|
| $P_D$              | 0.7910| 0.8105| 0.7933| 0.8223     | 0.8478    |
| $P_{FA}$           | 0.3922| 0.3709| 0.3645| 0.3609     | 0.3396    |

Fig. 7. (a) Probability of detection $P_D$ and (b) probability of false alarm $P_{FA}$ for topology inference versus the number of links for CDA and EM-CDA with GCT and TE causality metrics.
EM-CDA is able to outperform CDA methods in terms of both probabilities, with gains saturating when enough information is collected.

The performance of CDA and EM-CDA is investigated as a function of the ground-truth packet loss rate in Fig. 6. It is shown that the detection probability of the CDA schemes decreases as the packet loss rate increases, while the false alarm probability increases. EM-CDA is seen to be able to compensate for some of this performance loss, especially when using TE.

The relation between inference performance and the active link is investigated in Fig. 7, while the number of nodes is fixed as N = 10, and the number of the active link is changed from 31 to 43. With an increase in the active link, the mutual interference between nodes gets larger, but EM-CDA is able to retain its performance advantage as compared to CDA method. A similar conclusion is reached from Fig. 8, which varies the number of nodes N for a fixed fraction, 0.3, of active links.

VI. CONCLUSION

In this paper, we have introduced EM-CDA, a novel algorithm for passive network topology inference based on the observation of timing meta-data. The approach builds on the state-of-the-art causality discovery algorithm (CDA), and it addresses the important open problem of mitigating the effect of packet losses. Packet losses cause some of the timings of data packets to have no ACK packet counterparts, making CDA schemes potentially ineffective. EM-CDA formulates the topology inference problem as the discrete maximum likelihood (ML) problem of identifying active links in the presence of latent packet losses. It alternates between estimation of packet losses and application of a CDA strategy. Numerical results based on NS-3 simulations of real-world networks show that EM-CDA outperforms CDA in terms of detection probability and false alarm probability by a range of 4% to 12% under a variety of network conditions accounting for different packet loss rates, number of nodes, and active links. Future work may investigate more accurate approximations of the EM algorithm, e.g., in the evaluation of the posterior distribution in the EM step, as well as the adoption of a more detailed model to define the ML problem.

REFERENCES

[1] M. Cociglio, G. Fioccola, G. Marchetto, A. Sapio, and R. Sisto, “Multipoint passive monitoring in packet networks,” IEEE/ACM Trans. Netw., vol. 27, no. 6, pp. 2377–2390, Dec. 2019.
[2] P.-O. Brissaud, J. Françoise, I. Chrisment, T. Cholez, and O. Bettan, “Transparent and service-agnostic monitoring of encrypted web traffic,” IEEE Trans. Netw. Serv. Manage., vol. 16, no. 3, pp. 842–856, Sep. 2019.
[3] Y. Gao et al., “iPath: Path inference in wireless sensor networks,” IEEE/ACM Trans. Netw., vol. 24, no. 1, pp. 517–528, Feb. 2016.
[4] W. Albazrooee, J. Huang, and G. Xing, “A practical Bluetooth traffic sniffing system: Design, implementation, and countermeasure,” IEEE/ACM Trans. Netw., vol. 27, no. 1, pp. 71–84, Feb. 2019.
[5] M. Laghate and D. Cabric, “Learning wireless networks’ topologies using asymmetric Granger causality,” IEEE J. Sel. Topics Signal Process., vol. 12, no. 1, pp. 233–247, Feb. 2018.
[6] P. Sharma, D. J. Bucci, S. K. Brahme, and P. K. Varshney, “Communication network topology inference via transfer entropy,” IEEE Trans. Netw. Sci. Eng., vol. 7, no. 1, pp. 562–575, Jan. 2019.
[7] P. Tilghman and D. Rosenbluth, “Inferring wireless communications links and network topology from externals using Granger causality,” in Proc. MILCOM, Nov. 2013, pp. 1284–1289.
[8] M. G. Moore and M. A. Davenport, “Analysis of wireless networks using Hawkes processes,” in Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Jul. 2016, pp. 1–5.
[9] H. Xu, M. Farajtabar, and H. Zha, “Learning Granger causality for Hawkes processes,” in Proc. ICML, 2016, pp. 1717–1726.
[10] Z. Guo, V. M. McClelland, O. Simeone, K. R. Mills, and Z. Cveticovic, “Multiscale wavelet transfer entropy with application to corticomuscular coupling analysis,” IEEE Trans. Biomed. Eng., vol. 69, no. 2, pp. 771–782, Feb. 2022.
[11] J. D. Finkle, J. J. Wu, and N. Bagheri, “Windowed Granger causal inference strategy improves discovery of gene regulatory networks,” Proc. Nat. Acad. Sci. USA, vol. 115, no. 9, pp. 2252–2257, Feb. 2018.
[12] J. Runge et al., “Inferring causation from time series in earth system sciences,” Nature Commun., vol. 10, no. 1, pp. 1–13, Jun. 2019.
[13] A. Bovet and H. A. Makse, “Influence of fake news in Twitter during the 2016 U.S. presidential election,” Nature Commun., vol. 10, no. 1, pp. 1–14, Jan. 2019.
[14] C. W. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” Econometrica, vol. 37, no. 3, pp. 424–438, 1969.
[15] T. Schreiber, “Measuring information transfer,” Phys. Rev. Lett., vol. 85, no. 2, p. 461, 2000.
[16] M. Staniek and K. Lehnertz, “Symbolic transfer entropy,” Phys. Rev. Lett., vol. 100, no. 15, Apr. 2008, Art. no. 158101.
[17] J. Zhang, O. Simeone, Z. Cveticovic, E. Abele, and M. Richardson, “ITENE: Intrinsic transfer entropy neural estimator,” 2019, arXiv:1912.07277.
[18] W. Liang, S. X. Ng, and L. Hanzo, “Cooperative overlay spectrum access in cognitive radio networks,” IEEE Commun. Surveys Tuts., vol. 19, no. 3, pp. 1924–1944, 3rd Quart., 2017.
[19] Z. Liu, G. Ding, Z. Wang, S. Zheng, J. Sun, and Q. Wu, “Cooperative topology sensing of wireless networks with distributed sensors,” IEEE Trans. Cogn. Commun. Netw., vol. 7, no. 2, pp. 524–540, Jun. 2021.
[20] B. Deb, S. Bhatnagar, and B. Nath, “A topology discovery algorithm for sensor networks with applications to network management,” Dept. Comput. Sci., Rutgers Univ., Tech. Rep. DCS-TR-441, 2001.
Wenbo Du (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is a Professor with the School of Electronic and Information Engineering, Beihang University, Beijing, China. His current research interests include data science and intelligent transportation.

Tao Tan received the B.S. degree from the School of Electronic Information Engineering, Beihang University, Beijing, China, in 2020, where he is currently pursuing the Ph.D. degree in signal and information processing. His current research interests include network inference and machine learning.

Haijun Zhang (Fellow, IEEE) is currently a Full Professor and an Associate Dean with the School of Computer and Communications Engineering, University of Science and Technology of China, Hefei, China. He was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia (UBC), Canada. He received the the IEEE ComSoc Young Author Best Paper Award in 2017, IEEE CSIM Technical Committee Best Journal Paper Award in 2018, and the IEEE ComSoc Asia-Pacific Best Young Researcher Award in 2019. He serves/served as the Track Co-Chair for VTC Fall 2022 and WCNC 2020/2021, the Symposium Chair for Globecom'19, the TPC Co-Chair for INFOCOM 2018 Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Networks, and the General Co-Chair for GameNets’16. He serves/served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. He is a Distinguished Lecturer of IEEE.

Xianbin Cao (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in computer applications and information science from Anhui University, Hefei, China, in 1990 and 1993, respectively, and the Ph.D. degree in information science from the University of Science and Technology of China, Hefei, in 1996. He is currently a Professor with the School of Electronic and Information Engineering, Beihang University, Beijing, China. His current research interests include intelligent transportation systems, air traffic management, and intelligent computation.

Gang Yan (Member, IEEE) is currently a Full Professor with Tongji University, Shanghai, China. He has published more than 30 articles in international journals, including Nature and Nature Physics. His main research interests include network science, especially brain and brain-inspired networks. He was a recipient of the Thousand Youth Talents Plan in 2016. He is an Associate Editor of the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING.

Osvaldo Simeone (Fellow, IEEE) received the M.Sc. (Hons.) and Ph.D. degrees in information engineering from Politecnico di Milano, Milan, Italy, in 2001 and 2005, respectively. He is currently a Professor of information engineering with the Centre for Telecommunications Research, Department of Engineering, King’s College London, London, U.K., where he directs the King’s Communications, Learning, and Information Processing Laboratory. From 2006 to 2017, he was a Faculty Member with the Electrical and Computer Engineering Department, New Jersey Institute of Technology, Newark, NJ, USA, where he was affiliated with the Center for Wireless Information Processing. His research has been also supported by the U.S. National Science Foundation, European Commission, European Research Council, Vienna Science and Technology Fund, European Space Agency, and number of industrial collaborations, including with Intel Labs and InterDigital. He is the author of the textbook Machine Learning for Engineers (Cambridge University Press), four monographs, two edited books, and more than 180 research journal and magazine articles. His research interests include information theory, machine learning, wireless communications, neuromorphic computing, and quantum machine learning. He is a fellow of the IET and EPSRC. He was a co-recipient of the 2022 IEEE Communications Society Outstanding Paper Award, the 2021 IEEE Vehicular Technology Society Jack Neubauer Memorial Award, the 2019 IEEE Communication Society Best Tutorial Paper Award, the 2018 IEEE Signal Processing Best Paper Award, the 2017 IET Best Paper Award, the 2015 IEEE Communication Society Best Tutorial Paper Award, the Best Paper awards of IEEE SPAWC 2007, and the IEEE WIRECOM 2007. He was a recipient of an Open Fellowship by the EPSRC in 2022 and a Consolidator Grant by the European Research Council (ERC) in 2016. He is the Chair of the Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society and the U.K. and Ireland Chapter of the IEEE Information Theory Society. He is currently a Distinguished Lecturer of the IEEE Communications Society. He was a Distinguished Lecturer of the IEEE Information Theory Society in 2017 and 2018.