2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Partialization Analysis for Estimating HUB Network Topology

Kaori Kuroda, Takuya Matsumoto and Mikio Hasegawa

†Department of Electrical Engineering, Faculty of Engineering, Tokyo University of Science
6-3-1 Niijuku, Katsushika-ku, Tokyo, 125-8585, Japan
Email: kuroda@ee.kagu.tus.ac.jp, matsumoto@haselab.ee.kagu.tus.ac.jp, hasegawa@ee.kagu.tus.ac.jp

Abstract—In layer2 networks, if there is a loop, or a wire is broken on a HUB network, connected terminals will not be able to communicate. To resolve this issue, we estimate a hub network topology from packet data. We transformed the packet data to continuous time series by using an extended kernel density estimation method, and applied the partialization analysis to the transformed continuous time series. To check the validity of the proposed method, we conducted experiments using a network simulator and implemented system composed multiple HUB devices and terminals. As a result, we could estimate the network topology with high estimation accuracy.

1. Introduction

In layer2 networks, network failure occurs when network has a loop structure, or a cable is disconnected. If the network is constructed from many HUB devices, it will take a lot of time to detect the cause of the trouble. However, if we can estimate the network topology, it is easy to detect the cause of the trouble. In this paper, we propose a method to estimate a network topology from packet data which are observed in HUB devices. If observed time series are continuous and smooth, the network structure the network structure can be estimated through statistical measures [1, 2, 3, 4] applied to the continuous time series. However, the packet data is observed as event sequences. It is difficult for us to directly apply the conventional statistical measures [1, 2, 3, 4] to such event sequences. Then, it is an important issue to develop a method to estimate network structures in case that event sequences are observed. We transformed a packet data to a continuous time series. After transforming the packet data to continuous time series, we applied the partialization analysis to the transformed continuous time series, and estimated network topology.

2. Method

2.1. Transforming method

In this paper, we use packet data observed from all HUB devices simultaneously. The packet data is an event time series which has packet arrival time and packet size. Such a time series having an event time and its additional information is called a marked point process data. To estimate a network topology from observed time series, the partial correlation analysis[1, 2, 3, 4] is effective. However, we cannot apply the partial correlation analysis directly to the marked point process. Then, we transform the marked point process to continuous time series. To transform packet data to continuous time series, a kernel density estimator is used [5]. To apply this method to marked point process data, we expanded the kernel density estimator by using the additional information. Let us define the the lth event timing of the ith packet data as $t_i^l (l = 1, 2, \dots, n)$, and packet size at t_i^l as $s(t_i^l)$. The continuous time series $x_i(t)$ which transformed from packet data is defined as follows,

$$x_i(t) = \frac{1}{n}\sum_{j=1}^{n} K(t - t_i^j)s(t_i^j), \tag{1}$$

where $K(t)$ is a Kernel function. In this paper, we used Hanning window function as the Kernel function

$$K(t) = \frac{1}{2}(1 + \cos\frac{2\pi t}{T}), \tag{2}$$

where T is a bandwidth. We show an example of continuous time series transformed from packet data in Fig. 1.

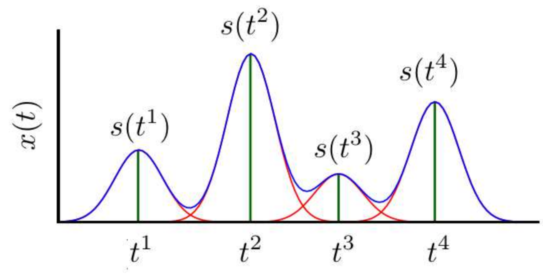


Figure 1: An example of continuous time series transformed from packet data.

2.2. Partialization analysis

If we apply only the correlation analysis to the transformed time series $x_i(t)$ $(i = 1, 2, \dots, N)$, we cannot estimate the network topology correctly. If two nodes are indirectly connected through other nodes or two nodes are driven by a common input (Fig.2), their continuous time

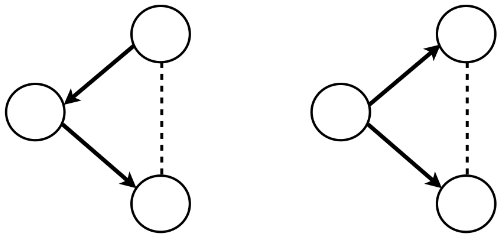


Figure 2: The examples of topology when the spurious correlation occurs.

series can have spurious correlation. To remove such spurious correlations, the partialization analysis is effective. Then, we apply the partialization analysis to the transformed time series $x_i(t)$ $(i = 1, 2, \ldots, N)$. The partial correlation coefficient can be derive from linear regression models,

$$x_i(t) = a_0 + \sum_{k=1, k\neq i\neq j}^{n} a_k x_k(t) + e_i(t), \tag{3}$$

$$x_j(t) = b_0 + \sum_{k=1, k\neq i\neq j}^{n} b_k x_k(t) + e_j(t), \tag{4}$$

where a_i and b_i are regression coefficients, and $e_i(t)$ is a residual. The partial correlation coefficient is a correlation coefficient between residuals,

$$p_{ij} = \frac{\sum_t (e_i(t) - \bar{e}_i)(e_j(t) - \bar{e}_j)}{\sqrt{\sum_t (e_i(t) - \bar{e}_i)^2 \sum_t (e_j(t) - \bar{e}_j)^2}}. \tag{5}$$

The partial correlation coefficient can be calculated as follows,

$$p_{ij} = -\frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}}, \tag{6}$$

where S_{ij} is the (i, j)th element in an inverse correlation matrix.

Finally, we reconstruct a network topology by using partial correlation coefficients. If two HUB devices are connected, the partial correlation coefficient might increase. On the other hand, if these HUB devices are not connected, the partial correlation coefficient might decrease. Thus, we extracted higher values of the partial correlation coefficient by discriminating the coupled and uncoupled pairs by calculating a threshold. To exclude any subjective discrimination, the threshold is decided by the Otsu thresholding [6], which is based on a linear discriminant analysis.

To confirm the estimation accuracy, we compared the topology of an estimated network with that of the original network. We used an index defined by

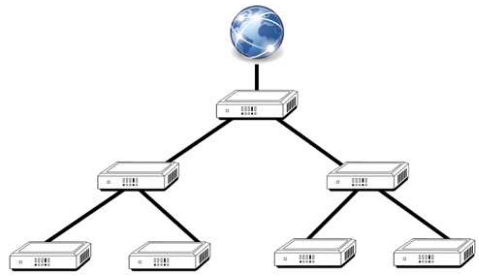


Figure 3: Network topology used in the experiments using simulator.

$$E = \frac{\sum_{i,j=1}^{N} (\alpha_{ij}\tilde{\alpha}_{ij} + (1 - \alpha_{ij})(1 - \tilde{\alpha}_{ij}))}{N(N-1)} \times 100, \tag{7}$$

where α_{ij} and $\tilde{\alpha}_{ij}$ are the (i, j)th element of the adjacency matrix of the original and the estimated network topology, respectively. If the ith and jth neurons are coupled, α_{ij} and $\tilde{\alpha}_{ij}$ take unity. If they are not coupled, α_{ij} and $\tilde{\alpha}_{ij}$ take zero. If E is close to 100, our method estimates the original network topology well.

3. Experimental setting and results

To evaluate the proposed method, we experiment using a simulator and a real machine. We used a network topology as a tree topology in the experiments using simulator (Fig. 3). We set the HUB devices connects to two terminals. Figure 4 shows the estimation accuracy when the number of HUB devices is changed. From the results, when the network size N is 3, the estimation accuracy E in both correlation coefficient and partial correlation coefficient is 100, because a spurious correlation does not occur in this network topology. When the network size is increased, the spurious correlation occurs. Then, the estimation accuracy in the correlation coefficient becomes worse. However, the estimation accuracy in the partial correlation coefficient is 100 even though the network size is increased.

We also show the histogram of the correlation coefficient and the partial correlation coefficient in Fig.5. In Fig.5(a), when the correlation coefficient is used, the distribution of coupled and uncoupled pairs widely overlap. As shown Fig.5(b), when the partial correlation coefficient is used, coupled and uncoupled pairs are effectively discriminated.

To observe the packet data from HUB devices, we implemented the HUB devices by using a raspberry pi and USB Ethernet adapters. The network topology used in this experiment is shown in Fig.6. We compared the estimation

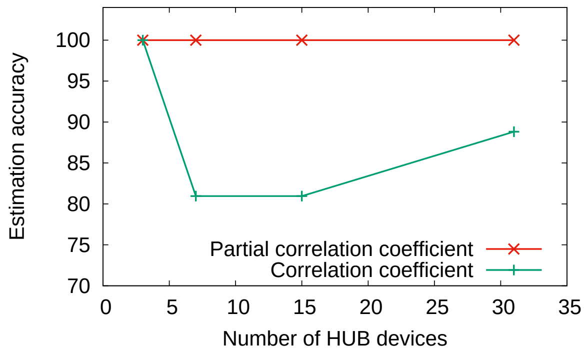
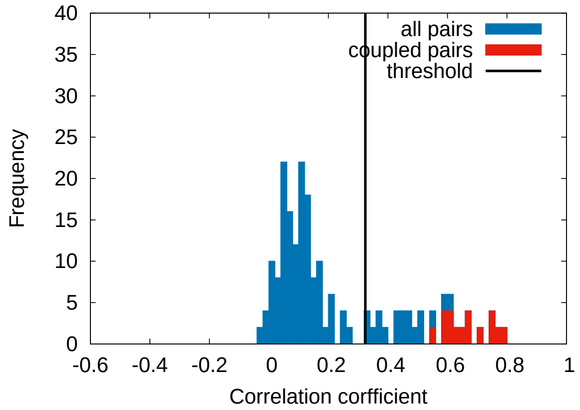
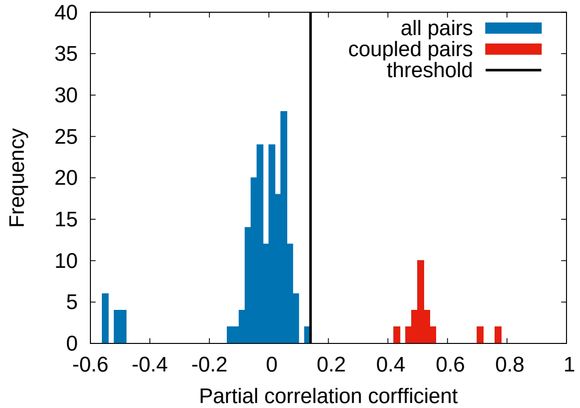


Figure 4: Estimation accuracy of the network topology.



(a) Correlation coefficient



(b) Partial correlation coefficient

Figure 5: Histogram of the correlation coefficient and partial correlation coefficient when the network size N is 15.

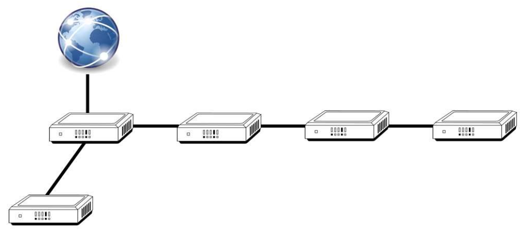


Figure 6: Network topology used in the experiments using real systems.

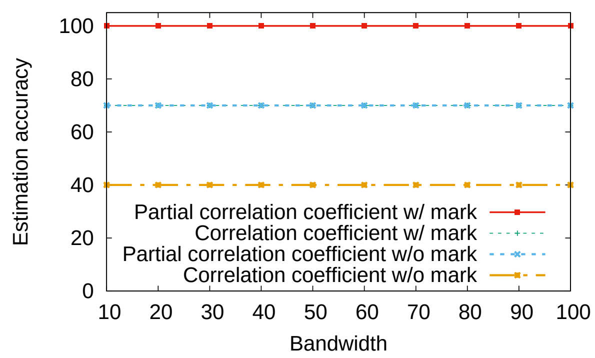
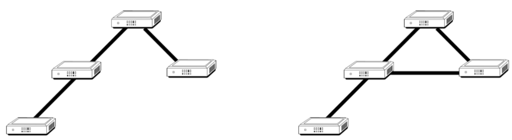


Figure 7: Estimation accuracy when a bandwidth is changed.



(a) non-loop topology (b) loop topology

Figure 8: Network topology. (a) Non-loop topology and (b) loop topology.

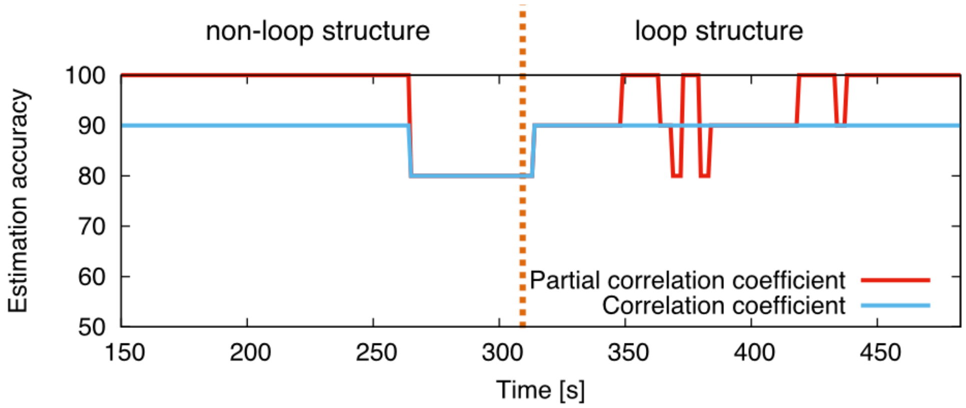


Figure 9: Estimation accuracy when the network topology changed to the loop structure.

accuracy when both the packet arrival time and the packet size as mark are used and when only the packet arrival time is used (Fig. 7). From the results, when the network topology is estimated by using the correlation coefficient using the packet size which is the mark information, the estimation accuracy shows a high value, and it was found that the network topology can be correctly estimated even in the experiment using real machines. Then, when the mark information is not used, the estimation accuracy worsens. The correlation coefficient results showed that the estimation accuracy was higher when using the mark information than when using only packet arrival time.

To estimate dynamically changing network topology, we estimate the network topology by dividing time series into the small temporal epochs. We set the length of the small temporal window is 50[s]. We show the results of estimation accuracy when the network topology is changed to loop topology (Fig. 9). The network topology is changed from non-loop topology to loop topology (Fig. 8) when the time is 315[s] (orange dotted line in Fig. 9). Because the length of the small temporal window is 50[s], the estimation accuracy is low when the network topology is changed. When the network topology is loop, the estimation accuracy occasionally becomes worse, because each HUB device receives same packets multiple time. However, the partial correlation coefficient exhibits higher estimation accuracy than the correlation coefficient.

4. Conclusion

In this paper, we proposed the method for estimating a HUB network topology from packet data. We confirmed that we can estimate the HUB network topology correctly. In addition, we show that the estimation accuracy becomes high when we use not only the packet arrived time but also packet size.

As a future work, to estimate a loop topology with high accuracy, we will improve the method by using distance information between packet data.

Acknowledgments

The research of K.K. was supported by JSPS KAKENHI Grant Number 15K21232.

References

[1] B. Schelter, M. Winterhalder, R. Dahlhaus, J. Kurths and J. Timmer, "Partial Phase Synchronization for Multivariate Synchronizing Systems," Phys. Rev. Lett., **96**:208103, 2006.

[2] D. Smirnov, B. Schelter, M. Winterhalder and J. Timmer, "Revealing Direction of Coupling between Neuronal Oscillators from Time Series: Phase Dynamics Modeling versus Partial Directed Coherence," Chaos, **17**:013111, 2007.

[3] S. Frenzel and B. Pompe, "Partial Mutual Information for Coupling Analysis of Multivariate Time Series," Phys. Rev. Lett., **99**:204101, 2007.

[4] M. Eichler, R. Dahlhaus and J. Sandkuhler, "Partial Correlation Analysis for the Identification of Synaptic Connections," Biological Cybernetics, **89**:289, 2003.

[5] K. Kuroda and T. Ikeguchi, "Reconstructing Connectivity of Nonlinear Dynamical Systems from Marked Point Processes,"*Proc. of NOLTA'11*, pp. 697–700, 2011.

[6] N. Otsu, "A Threshold Selection Method from Gray Level Histograms," *IEEE Trans. on Syst., Man and Cybernetics*, **9** 62-66, 1979.