DCE: A NOVEL DELAY CORRELATION MEASUREMENT FOR TOMOGRAPHY WITH PASSIVE REALIZATION

Peng Qin[1,2], Bin Dai[1], Benxiong Huang[1] and Guan Xu[1]

[1] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China
[2] China Academy of Electronics and Information Technology, Beijing, China

ABSTRACT

Tomography is important for network design and routing optimization. Prior approaches require either precise time synchronization or complex cooperation. Furthermore, active tomography consumes explicit probing resulting in limited scalability. To address the first issue we propose a novel Delay Correlation Estimation methodology named DCE with no need of synchronization and special cooperation. For the second issue we develop a passive realization mechanism merely using regular data flow without explicit bandwidth consumption. Extensive simulations in OMNeT++ are made to evaluate its accuracy where we show that DCE measurement is highly identical with the true value. Also from test result we find that mechanism of passive realization is able to achieve both regular data transmission and purpose of tomography with excellent robustness versus different background traffic and package size.

KEYWORDS

Network Tomography, Delay Correlation Measurement, Passive Realization.

1. INTRODUCTION

Network tomography [1] studies internal characteristics of Internet using information derived from end nodes. One advantage is that it requires no participation from network elements other than the usual forwarding of packets while traditional traceroute method needs response to ICMP messages facing challenge of anonymous routers [2,3].

Many literatures choose delay to calculate correlation between end hosts for tomography. However, they require either precise time synchronization or complex cooperation. Moreover, active tomography consumes quantities of explicit probing bandwidth which results in limited scalability.

In this paper we propose a novel Delay Correlation Estimation approach named DCE with no need of cooperation and synchronization between end nodes. The greatest property is that we only need to measure the packet arriving time at receivers. To further reduce bandwidth consumption a passive mechanism using regular data flow is developed.

We do extensive simulations in OMNeT++ to evaluate its accuracy. Results show that the correlation of delays $\sigma^2_{d_a,d_b}$ measured by DCE is highly identical with the true value σ^2_s on shared path. By altering background traffic and package size we see that passive mechanism has

DOI:10.5121/ijdps.2015.6301

excellent robustness and is able to achieve regular data transmission as well as purpose of tomography.

1.1. Contributions

- We propose DCE to estimate delay correlation. This method needs no special cooperation or synchronization and avoids issues using RTT, making it largely different from prior tomography tools [4] and [5].
- We develop a passive mechanism for realization, which is efficient for bandwidth saving.
- Extensive simulations in OMNeT++ demonstrates its accuracy and robustness.

2. RELATED WORK

Y. Vardi was one of the first to study network tomography [6] that can be implemented in either an active or passive way. Active network tomography [7,8,9,10] needs to explicitly send out probing messages to estimate the end-to-end path characteristics, while passive network tomography [11,12,13,14] infers network topology without sending any explicit probing messages.

Article [4]describes delay tomography which however, needs synchronization and cooperation between sender and receiver. In [5] authors develop Network Radar based on RTT trying to solve these issues. However, two reasons distort the measurement accuracy. One is due to the variable processing delay at destination nodes and the other is its violation of a significant assumption that return paths of packet are uncorrelated while actually they overlaps.

In addition, delay correlation can be further used for topology recovery [15,16], which is important to improve network performance.

In [17], the concept of DCE was briefly introduced without detailed analysis and evaluation. This paper includes the full-fledged version of the DEC method together with a comprehensive analysis of practical cases and experiment results.

3. DELAY CORRELATION ESTIMATION

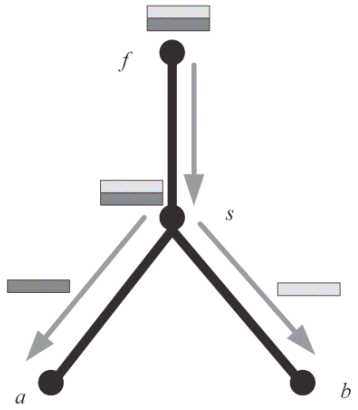

Figure 1. The tree structure: a sender *f* and two receivers *a,b*

A simple model we use in this paper is shown in *Fig.1*. The routing structure from sender *f* to receivers *a,b* is a tree rooted at *f*. Otherwise, there is routing loop which must be corrected. Assume that router *s* is the ancestor node of both *a* and *b*. Assume that the sender uses unicast to send messages to receivers, and assume that packets are sent in a back-to-back pair. For the *k*-th pair of back-to-back packets, denoted as *a* and *b*, sent from *f* to *a* and *b*, respectively, we use the following notations:

- $t_a(k)$: the time when *a* receive a^k in the *k*-th pair.
- $t_b(k)$: the time when *b* receive b^k in the *k*-th pair.
- $d_a(k)$: the latency of a^k along the path from *f* to *a*.
- $d_a(k)$: the latency of b^k along the path from *f* to *b*.
- $t_f(k)$: the time when *f* sends the *k*-th pair of packets.

Using network model of *Fig.1* we obtain *Eq.*(1) for a^0 (we start the index with 0 for convenience):

$$t_a(0) = t_f(0) + d_a(0). \quad (1)$$

Similarly, for a^k, we have

$$t_a(k) = t_f(k) + d_a(k). \quad (2)$$

Let *Eq.*(2) – *Eq.*(1) and $\delta_a(k) \equiv t_a(k) - t_a(0)$. We can obtain *Eq.* (3) :

$$\delta_a(k) = (t_f(k) - t_f(0)) + (d_a(k) - d_a(0)). \quad (3)$$

Denote the time interval between two consecutive pairs of packets as δ. We assume that δ is a constant for simplicity at this moment, and relax this assumption later. In this case we use $k\delta \equiv k \cdot \delta$ to replace $t_f(k)$-$t_f(0)$ in *Eq.*3. We then have:

$$d_a(k) = \delta_a(k) - k\delta + d_a(0). \quad (4)$$

Let $\delta'_a(k) \equiv \delta_a(k) - k\delta$. *Eq.* (4) can be transformed into *Eq.* (5) :

$$d_a(k) = \delta'_a(k) + d_a(0). \quad (5)$$

We can achieve similar result at receiver *b* as in *Eq.*(6):

$$d_b(k) = \delta'_b(k) + d_b(0). \quad (6)$$

where $\delta'_b(k) \equiv \delta_b(k) - k \cdot \delta$.

To estimate the correlation between $d_a(k)$ and $d_b(k)$, we introduce the following lemma.

Lemma 1. Assume that ζ, η are two random variables, and $\chi = a\zeta + b$, $\gamma = c\eta + d$, where a,b,c,d are constants and a,c have the same symbol, thus we have $\sigma^2_{\zeta,\eta} = \sigma^2_{\chi,\gamma}$.

Proof.

$$\sigma^2_{\chi,\gamma} = \frac{E(\chi - E\chi)(\gamma - E\gamma)}{\sqrt{D\chi}\sqrt{D\gamma}} \quad (7)$$

$$= \frac{E(a\zeta + b - aE\zeta - b)(c\eta + d - cE\eta - d)}{\sqrt{a^2 D\zeta}\sqrt{c^2 D\eta}}$$
$$= \frac{acE(\zeta - E\zeta)(\eta - E\eta)}{|a||c|\sqrt{D\zeta}\sqrt{D\eta}}$$
$$= \frac{E(\zeta - E\zeta)(\eta - E\eta)}{\sqrt{D\zeta}\sqrt{D\eta}}$$
$$= \sigma^2_{\zeta,\eta}$$

Since $d_a(0)$ and $d_b(0)$ can both be regarded as constants in one serial of probing packets, based on Lemma 1, we have the following theorem:

Theorem 1. The correlation between delay variables $d_a(k)$ and $d_b(k)$ is equal to the correlation between variables $\delta'_a(k)$ and $\delta'_b(k)$, which means,

$$\sigma^2_{d_a(k),d_b(k)} = \sigma^2_{\delta'_a(k),\delta'_b(k)}. \quad (8)$$

Based on the measurements of $\delta'_a(k)$ and $\delta'_b(k)$, we can calculate the correlation of delays along the path from *f* to *a* and along the path from *f* to *b*, denoted as $\sigma^2_{\delta'_a,\delta'_b}$., using *Eq.*(9):

$$\sigma^2_{\delta'_a,\delta'_b} = \frac{1}{n-1}\sum_{k=1}^{n}[\delta'_a(k) - \overline{\delta'_a}][\delta'_b(k) - \overline{\delta'_b}]. \quad (9)$$

where δ'_a is the sample mean of $\delta'_a(k)_{k=1}^n$ for *i=a,b.*

Theorem 2. $\sigma^2_{\delta'_a,\delta'_b}$ in *Eq.*(9) is an unbiased estimator of the correlation on shared path.

Proof. First of all we show that $\hat{\sigma}^2_{d_a,d_b}$ (not $\hat{\sigma}^2_{\delta'_a,\delta'_b}$) is an unbiased estimator of the correlation on shared path (*f,s*). Let λ_a λ_b denote the mean time latency of $path_a$, $path_b$ and let $\overline{d}_a, \overline{d}_b$ denote the sample mean correspondingly; true correlation $\sigma^2_{d_a,d_b} = E[(d_a(k) - \lambda_a)(d_b(k) - \lambda_b)]$ To prove that $E[\hat{\sigma}^2_{d_a,d_b}] = \sigma^2_{d_a,d_b}$ we analyze the expectation of $E[(d_a(k) - \overline{d}_a)(d_b(k) - \overline{d}_b)]$:

$$E[(d_a(k) - \overline{d}_a)(d_b(k) - \overline{d}_b)]$$
$$= E[(d_a(k)d_b(k)] - \frac{1}{n}\sum_{i=1}^{n}E[(d_a(k)d_b(i)] \quad (10)$$
$$-\frac{1}{n}\sum_{i=1}^{n}E[(d_a(i)d_b(k)] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E[(d_a(i)d_b(j)]$$

Since delays of the $i^{th}$and k^{th} pair are independent and

$$\sigma^2_{d_a,d_b} = E[(d_a(k) - \lambda_a)(d_b(k) - \lambda_b)] = E[d_a(k)d_b(k)] - \lambda_a\lambda_b$$

We obtain *Eq.* (11)

$$E[d_a(k)d_b(i)] = \begin{cases} \lambda_a\lambda_b & k \neq i \\ \lambda_a\lambda_b + \sigma^2_{d_a,d_b} & k = i \end{cases} \quad (11)$$

Substituting *Eq.*(11) into *Eq.*(10) we obtain

$$E[(d_a(k) - \bar{d}_a)(d_b(k) - \bar{d}_b)] = (\frac{n-1}{n})\sigma^2_{d_a,d_b}$$

Therefore, $\hat{\sigma}^2_{d_a,d_b}$ is an unbiased estimator of the correlation on shared path as is shown in *Eq.*(12)

$$E[\hat{\sigma}^2_{d_a,d_b}] = \frac{1}{n-1}\sum_{k=1}^{n} E[(d_a(k) - \bar{d}_a)(d_b(k) - \bar{d}_b)] = \frac{1}{n-1}n(\frac{n-1}{n})\sigma^2_{d_a,d_b} = \sigma^2_{d_a,d_b} \quad (12)$$

According to $\sigma^2_{d_a,d_b} = \sigma^2_{\delta'_a,\delta'_b}$ in *Theorem 1* we prove it.

3.1. Discussion of δ

3.1.1. δ is a constant

A complete delay correlation estimation algorithm (DCE) is summarized in *Algorithm 1* if the time interval δ is a constant.

Algorithm 1 Delay Correlation Estimation Algorithm

Require:
Given the time interval δ
Ensure:
1: **for** k = 0 : n **do**
2: Use *Eq.* (1) to *Eq.* (3) to measure δa(k) in the k-th transmission;
3: Use δ′a(k) = δa(k)−kδ to calculate δ′ a(k) in the k-th loop;
4: Similarly, measure δb(k) and calculate δ′b(k) =δb(k) − kδ;
5: **end for**
6: With all the values of δ′a(k) and δ′b(k), use *Eq.*(9) toobtain the correlation $\sigma^2_{\delta'_a,\delta'_b}$, which is equivalent to $\sigma^2_{d_a,d_b}$ between hosts a and b according to *Theorem 1*.

3.1.2 δ is not a constant

If the time interval δ is not a constant then $t_f(k) - t_f(0) \neq k\delta$. In this case using $k\delta$ to replace $t_f(k) - t_f(0)$ is inappropriate and we choose $\delta_f(k)$ to denote $t_f(k) - t_f(0)$ in *Eq.*(3), then *Eq.*(4) can be rewritten to *Eq.*(13).

$$d_a(k) = \delta_a(k) - \delta_f(k) + d_a(0). \quad (13)$$

Correspondingly, *Eq.*(5) and *Eq.*(6) can be replaced with $\delta'_a(k) \equiv \delta_a(k) - \delta_f(k)$ and $\delta'_b(k) \equiv \delta_b(k) - \delta_f(k)$ *respectively.*

Note that $t_f(k)$ is a timestamp contained in the packet, and thus $\delta_f(k) = t_f(k) - t_f(0)$ is readily available in practice.

3.2. A mechanism for passive realization

To reduce explicit probing we propose a mechanism for passive realization of *Algorithm 1* in real networks.

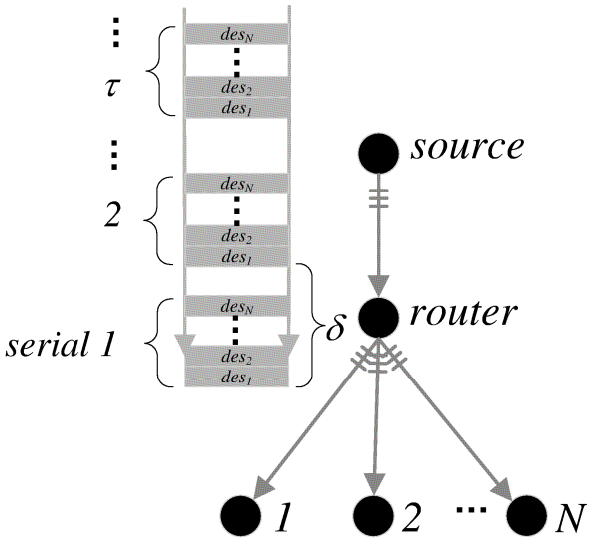

Figure 2: The mechanism for passive tomography with source and N end hosts.

As *Fig.2* shows passive mechanism works as follows. In practical networks (for example, P2P networks) if N end hosts request common contents from a *source*, it will distribute packets. In this situation source first chooses the *No.1* requested data block which is duplicated into packet *serial 1* and sent out to all *N* hosts simultaneously guaranteeing that there exist two successive packets in a back-to-back manner. An indicator (**IR**) is needed to tell if the received packet at each host belongs to the back-to-back pair. If serial of *No.1* is sent completely *source* repeats to the next until all requested contents are received by *N* hosts. As regular data flow proceeds transmitting we change destination address of the current two successive packets when delay correlation between the corresponding host pair has been measured (if number of packets sent to them with indicator **IR** reaches τ , where τ is a tunable threshold).

One may naturally raise two questions: first is which two successive packets in one serial are chosen to add an **IR**? while the other is how to guarantee that in each transmission the two successive packets are in a back-to-back manner? A simple solution can address both of them. The basic idea is that we divide packets into small size. This can satisfy both the need of back-to-back and regular data transmission. In fact our experiment results show that any two successive packets in a serial distributed to N hosts can be chosen to add **IR** as long as the time interval δ is appropriate.

3.3. Arrangement of IR

Based on above argument that each successive pair of packets can be regarded as back-to-back, in one transmitting serial at most N-1 delay correlations are measured (shown on top of Fig.3). After N-1 times switching of destination addresses and IRs all delay correlations between N hosts can be obtained.

In this way the complexity is only O(N) compared with O(N2) to measure N(N−1)/2 correlations.

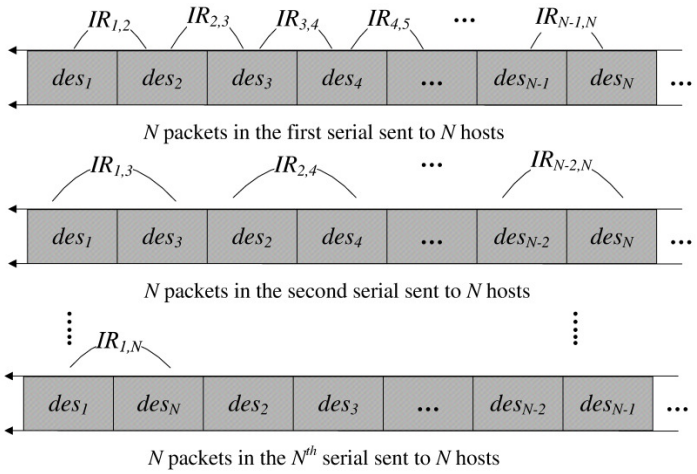

Figure 3: Arrangement of ***IR***. $IR_{i,j}$ in each serial indicates blocks destined to *i, j* should be successive.

4. SIMULATION RESULTS

4.1. Dynamic Network Tomography

We use OMNeT++ for simulations [18] to demonstrate the correctness and robustness of passive DCE tomography. We generate a network shown in *Fig.4*. Nodes of BG are randomly and dynamically selected for producing background traffic while others are the source and client nodes. When two client nodes *a, b* request contents from source node *f*, *it* sends regular data in a back-to-back manner. In this case route algorithm determines a multicast tree with root f and leaves *a, b*.

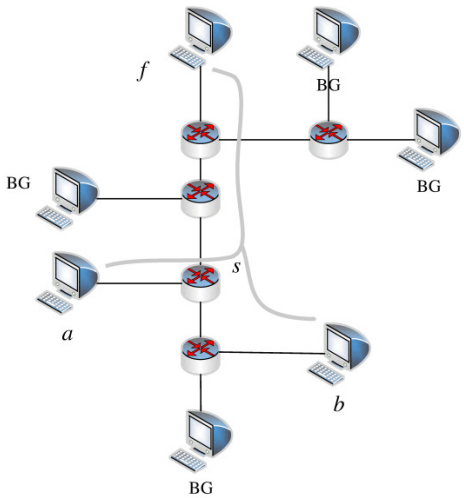

Figure 4: Structure for DCE test

We set packet size hundreds of bytes to satisfy both regular transmission and back-to-back property. One advantage is that since it is smaller than Maximum Transmission Unit (MTU) we avoid delay for package segmentation. We set bandwidth of each link value of 100Mbps; the

background traffic pattern conforms to the Poisson distribution, whose expectation value could be set from 1MBps to 12MBps. We also change the size of packet from 100 bytes to MTU to see its influence on DCE measurement.

4.2. Results

Using DCE methodology we set τ to be 1550 and observe over 1500 timing samples for receiver pair.

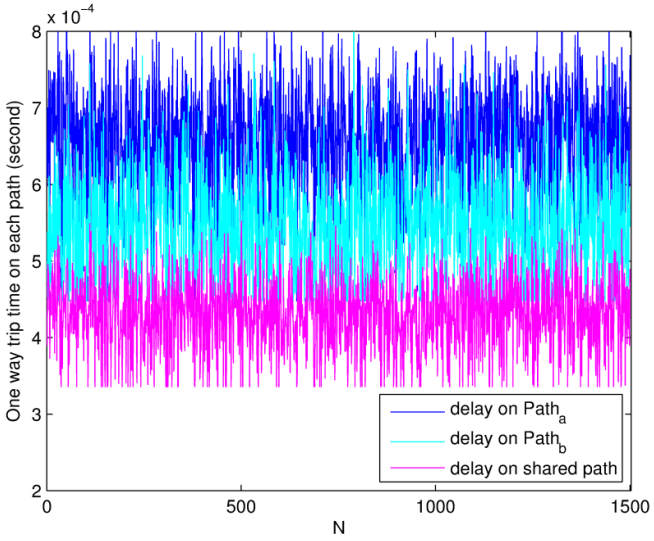

Figure 5: One way trip time latency on patha, pathband shared path (f,s)

Fig.5 depicts the one way trip latency in our environment. One example of the average delay on $path_a$, $path_b$ and shared path are 0.6526ms, 0.5478ms and 0.4317ms respectively. In some case errors may happen to the timestamp as the variation of background traffic. Therefore, we ignore packets beyond twice the average delay on each path.

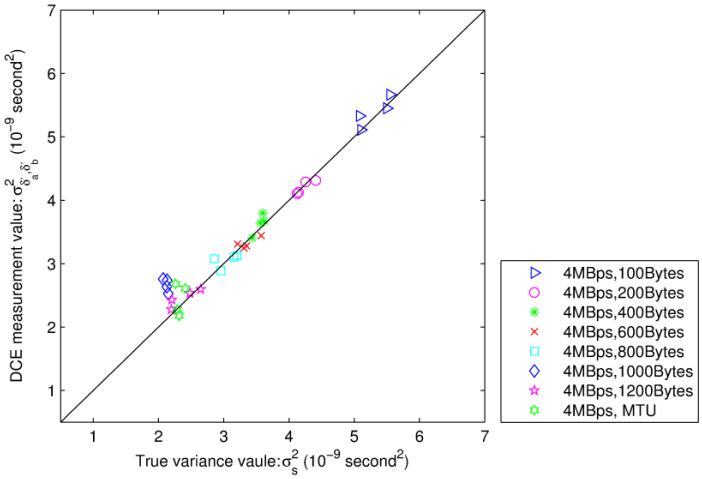

Figure 6: DCE measurement covariance $\sigma^2_{\delta'_a,\delta'_b}$ versus the directly measured delay variance on the shared path (f,s)

Fig.6 shows the DCE covariance $\sigma^2_{\delta'_a,\delta'_b}$ versus true value σ^2_s on shared path *(f,s)* with different packet size. Value $\sigma^2_{\delta'_a,\delta'_b}$ is calculated using the arriving time of package at *a,b* while σ^2_s is calculated directly from delay on (*f,s*). According to Theorem 1 we know that $\sigma^2_{d_a,d_b}$ is equal to $\sigma^2_{\delta'_a,\delta'_b}$ thus, ideally $\sigma^2_{\delta'_a,\delta'_b}$ and σ^2_s should be identical and fall onto the 45 degree line. Taking test result with package size of 800 Bytes for example we see the estimated value is always locating nearby the true value with slight difference which demonstrates the correctness of DCE tomography. If we change size of data packet from 100 bytes to MTU, delay correlations measured by DCE are always desired. This demonstrates that passive mechanism is able to achieve both data transmission and tomography.

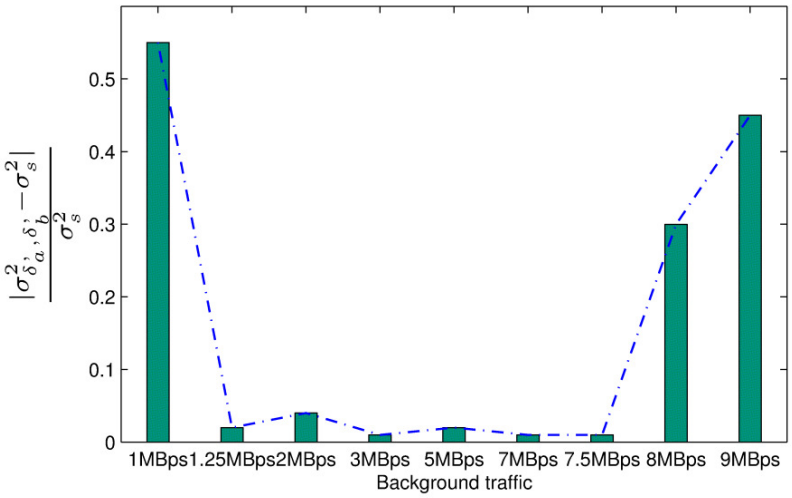

Figure 7: $\frac{\sigma^2_{\delta'_a,\delta'_b}-\sigma^2_s}{\sigma^2_s}$: DCE measurement error versus different background traffic

To see its robustness with different background traffic, in *Fig.7* packet size is fixed to be MTU. In this case performance of DCE tomography is perfect with the percentage error below 4% when the expectation value of background traffic is within the region [1.25MBps,7.5MBps]. However, when it increases to 8MBps, error percentage increases sharply. This is because in this situation network's performance becomes worse and some shared paths between higher level routers are congested heavily, which destroys the back-to-back property of packets. Note that when the expectation value of background traffic is relatively small (below 1MBps) delay correlation caused by queuing on routers will not be significant thus the performance also degrades.

5. CONCLUSIONS AND FUTURE WORK

In this paper we propose a novel tomography method named DCE to estimate delay correlation with no need of synchronization and cooperation between end hosts. We also develop the passive mechanism to further save bandwidth. Extensive simulations demonstrate the correctness of DCE. Moreover, passive realization is able to achieve both purpose of tomography and data transmission with excellent robustness versus different background traffic and package size.

In future, we plan to utilize the DCE measure for topology inference and implement it in practical network environment.

ACKNOWLEDGEMENTS

This work was supported by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China under Grant no.2012BAH93F01, the Innovation Research Fund of Huazhong University of Science and Technology, no. 2014TS095 and the National Sciece Foundation of China under Grant no. 60803005.

REFERENCES

[1] A. Chen, J. Cao, T. Bu, Network tomography: Identifiability and fourier domain estimation, IEEE *Transactions on Signal Processing* 58 (2010) 6029 – 6039.
[2] M. H. Gunes, K. Sarac, Resolving anonymous routers in internet topology measurement studies, in: Proceedings of the IEEE *INFOCOM* 2008, Phoenix, Arizona, USA, 2008, pp. 13 – 18.
[3] X. Zhang, C. Phillips, A survey on selective routing topology inference through active probing, IEEE *Communications Surveys and Tutorials* 14 (2012) 1129 – 1141.
[4] Y. Tsang, R. D. Nowak, Network delay tomography, IEEE *Transactions on Signal Processing* 51 (2003).
[5] Y. Tsang, P. Barford, R. Nowak, Network radar: Tomography from round trip time measurements, in: Proceedings of the 4th *ACM SIGCOMM conference on Internet measurement*(IMC 04), 2004,
[6] Y. Vardi, Network tomography: estimating source-destination traffic intensities from link data, Journal of *the American statistical association* 91 (1996) 365 – 377.
[7] R. Caceres, N. G. Duffield, J. Horowitz, D. F. Towsley, Multicast-based inference of network internal loss characteristics, IEEE *Transactions on Information Theory* 45(7) (1999) 2462 – 2480.
[8] F. L. Presti, N. G. Duffield, J. Horowitz, D. Towsley, Multicast-based inference of network-internal delay distributions, IEEE/ACM *Transactions on Networking* 10 (2002).
[9] M. Rabbat, R. Nowak, M. Coates, Multiple source, multiple destination network tomography, in: Proceedings of the IEEE *INFOCOM*, Piscataway, NJ, USA, 2004, pp. 1628 – 1639.
[10] A. Krishnamurthy, A. Singh, Robust multi-source network tomography using selective probes, in: Proceedings of the IEEE *INFOCOM* 2012, Orlando, Florida, USA, 2012, pp. 1629 – 1637.
[11] J. Cao, D. Davis, S. V. Wiel, B. Yu, S. Vander, W. B. Yu, Time-varying network tomography: router link data, *Journal of the American statistical association* 95 (2000) 1063 – 1075.
[12] V. N. Padmanabhan, L. Qiu, H. J. Wang, Passive network tomography using bayesian inference, *Microsoft Research* (2002).
[13] F. Ricciato, F. Vacirca, W. Fleischer, J. Motz, M. Rupp, Passive tomography of a 3g network: Challenges and opportunities, in:Proceedings of the IEEE *INFOCOM*, 2006.
[14] H. Yao, S. Jaggi, M. Chen, Passive network tomography for erroneous networks: A network coding approach, IEEE *Transactions on Information Theory* 58(9) (2012) 5922 – 5940.
[15] B. D. Eriksson, P. Barford, R. Nowak, Toward the practical use of network tomography for internet topology discovery, in: Proceedings of IEEE *INFOCOM* 2010, San Diego, California, USA,2010.
[16] J. Ni, H. Xie, Tatikonda, Yang, Efficient and dynamic routing topology inference from end-to-end measurements, IEEE/ACM *Transactions on Networking* (2010).
[17] P. Qin, B. Dai, B. Huang, G. Xu, K. Wu, A survey on network tomography with network coding, *Communications Surveys Tutorials*, IEEE PP (2014) 1 – 1.
[18] OMNet++4.3.1, Omnet++, http://www.omnetpp.org/, 2013. The homepage of *OMNet++*.pp. 175 – 180.

Authors

Peng Qin received the B. S. Degree in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, P. R. China, in 2009. His research interests are in the areas of network tomography, network measurement, p2p network and applications of network coding.

Bin Dai received the B. Eng, the M. Eng degrees and the PhD degree from Huazhong University of Science and Technology of China, P. R. China in 2000, 2002 and 2006, respectively. From 2007 to 2008, he was a Research Fellow at the City University of Hong Kong. He is currently an associate professor at Department of Electronics and Information Engineering, Huazhong University of Science and Technology, P. R. China. His research interests include p2p network, wireless network, network coding, and multicast routing.

Benxiong Huang received the B. S. degree in 1987 and PhD degree in 2003 from Huazhong University of Science and Technology, Wuhan, P. R. China. He is currently a professor in the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, P. R. China. His research interests include next generation communication system and communication signal processing.

Guan Xu received the B.S. degree in Electronics and Information Engineering from Huazhong University of science and technology, Wuhan, P. R. China, in 2008. He is currently a PhD Candidate in the Department of Electronics and Information Engineering at the Huazhong University of science and technology. His research interests are in the areas of practical network coding in P2P network, IP switch networks and SDN networks with emphasis on routing algorithms and rate control algorithms.