Link Quality Classifier with Compressed Sensing Based on ℓ_1-ℓ_2 Optimization

Takahiro Matsuda, *Member, IEEE,* Masaaki Nagahara, *Member, IEEE,* and Kazunori Hayashi, *Member, IEEE*

***Abstract*—Network tomography is an inference technique for internal network characteristics from end-to-end measurements. In this letter, we propose a new network tomography scheme to classify communication links into lower or higher quality classes according to their link loss rates. The two-class classification is achieved by the estimation of link loss rates via *compressed sensing*, which is an emerging theory to obtain a sparse solution from an underdetermined linear system, with regarding link loss rates in the higher quality class as 0. In the proposed scheme, we implement compressed sensing with an ℓ_1-ℓ_2 optimization, where the cost function is defined as a sum of ℓ_1 and ℓ_2 norms with a *mixing parameter*, which enables us to control the threshold between the lower and higher quality classes.**

***Index Terms*—Network tomography, compressed sensing, ℓ_1-ℓ_2 optimization.**

I. Introduction

NETWORK tomography is an inference technique for internal network characteristics such as link loss rates and link delays from end-to-end measurements. In this letter, we consider a network tomography problem to estimate link loss rates, which is referred to as *loss tomography*. There have been a lot of loss tomography schemes [1], [2], which focus on estimating loss rates of all links correctly. In some cases of practical network design or management problems, however, only a limited number of links typically have high loss rates and it would be more important to identify links with higher loss rates than to estimate every link loss rate in the network.

Taking advantages of the sparsity of the links with higher loss rates, some pioneering works have applied *compressed sensing* to network tomography problems [3]–[5]. Compressed sensing is an emerging theory in signal/image processing for acquiring a sparse vector from a small number of linear measurements [6], [7]. Since network tomography problems can be formulated by a set of linear equations [2], the idea of compressed sensing can be naturally introduced to the problems. Xu et al. [5] discuss the sufficient number of measurement paths for identifying sparse vectors of link qualities. Firooz and Roy [4] provide conditions on routing matrices to estimate k-sparse vectors, which contain at most k non-zero components, and an upper bound on estimation error.

In this letter, we propose a two-class link quality classifier based on estimated link loss rates by means of compressed sensing. The proposed scheme achieves two functions; rough estimate of link loss rates and classification of links. Our contribution is that, by the introduction of an ℓ_1-ℓ_2 optimization [8] for an implementation of compressed sensing, we provide a link quality classifier, which can adjust the threshold between lower and higher quality classes. The ℓ_1-ℓ_2 optimization, where the cost function is defined as a weighted sum of ℓ_1 and ℓ_2 norms, is commonly used to cope with noisy measurements in the context of compressed sensing, while it is also used to obtain sparse representations as in LASSO [9] and basis pursuit denoising in wavelet theory [10]. Our motivation mainly corresponds to the latter. Intuitively, ℓ_1 and ℓ_2 norms represent the measure of the sparsity and the estimation error of link loss rates, respectively, and the trade-off between them can be controlled by the weight, which we call *mixing parameter* in this letter. According to the system requirement, we can control the threshold of the link loss rate for the classification by using the mixing parameter, which will be demonstrated via simulation experiments.

II. System Model

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a directed acyclic network, where $\mathcal{V}$ and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denote sets of nodes and links, respectively. Moreover, let $\mathcal{V}_{\rm S} \subset \mathcal{V}$ and $\mathcal{V}_{\rm R} \subset \mathcal{V}$ denote sets of source and receiver nodes, respectively. We define $\mathcal{W} = \{path_{s,r}; s \in \mathcal{V}_{\rm S}, r \in \mathcal{V}_{\rm R}\}$ as a set of all paths and $path_{s,r} = \{path_{s,r}^{(l)};\ l = 1, 2, \ldots, |path_{s,r}|\}$ as a set of paths from source node $s \in \mathcal{V}_{\rm S}$ to receiver node $r \in \mathcal{V}_{\rm R}$, where $path_{s,r}^{(l)} = \{(s, v_{s,r}^{(l,1)}), (v_{s,r}^{(l,1)}, v_{s,r}^{(l,2)}), \ldots, (v_{s,r}^{(l,|path_{s,r}^{(l)}|-1)}, r)\} \subset \mathcal{E}$ represents the lth path in $path_{s,r}$ and $v_{s,r}^{(l,k)} \in \mathcal{V} \setminus \{s, r\}$ $(k = 1, \ldots, |path_{s,r}^{(l)}| - 1)$ are intermediate nodes in the path.

We reformulate $\mathcal{W}$ and $\mathcal{E}$ as $\mathcal{W} = \{w_1, w_2, \ldots, w_M\}$ and $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$, respectively, where $M = |\mathcal{W}|$ and $N = |\mathcal{E}|$ denote the numbers of paths and links, respectively. We assume that packets are lost randomly on each link e_j $(j = 1, 2, \ldots, N)$ with link loss rate p_{e_j}. Thus, a packet transmitted on a path w_i $(i = 1, 2, \ldots, M)$ is successfully received at its receiver node with rate $1 - q_{w_i} = \prod_{e_j \in w_i}(1 - p_{e_j})$. We define $\boldsymbol{y} = (y_1\ y_2\ \cdots\ y_M)^\top$ and $\boldsymbol{x} = (x_1\ x_2\ \cdots\ x_N)^\top$ as

$$y_i = -\log(1 - q_{w_i}) = -\sum_{e_j \in w_i} \log(1 - p_{e_j}),$$

$$x_j = -\log\left(1 - p_{e_j}\right),$$

where $\top$ denotes the transpose operator. Then, we obtain

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{A} \in \{0, 1\}^{M \times N}$ represents the *routing matrix* of $\mathcal{W}$, i.e., (i, j) components $a_{i,j}$ $(i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, N)$

Manuscript received July 27, 2011. The associate editor coordinating the review of this letter and approving it for publication was Y.-D. Lin.

T. Matsuda is with the Graduate School of Engineering, Osaka University, Osaka, 5650871, Japan (e-mail: matsuda@comm.eng.osaka-u.ac.jp).

M. Nagahara and K. Hayashi are with the Graduate School of Informatics, Kyoto University, Kyoto, 6068501, Japan (e-mail: {nagahara, kazunori}@i.kyoto-u.ac.jp).

Digital Object Identifier 10.1109/LCOMM.2011.082911.111611

in $\boldsymbol{A}$ are set to $a_{i,j} = 1$ if $e_j \in w_i$, and $a_{i,j} = 0$ otherwise. In this letter, we assume that link states are stationary, i.e., link loss rates do not change while the proposed scheme is applied.

III. Loss Tomography via Compressed Sensing

A. Overview of the Proposed Link Quality Classification

In general, network tomography techniques are categorized into *active measurement* and *passive measurement* schemes, where the former obtains end-to-end measurements by injecting probe packets into the network and the latter does by collecting existing packets in the network. We consider an active measurement scheme and assume that $\boldsymbol{A}$ is given in advance by establishing paths of probe packets administratively.

The purpose of the proposed scheme is to classify links according to estimated link loss rates obtained from $\boldsymbol{y}$. Let $K_{\mathrm{S}}^{(w_j)}$ and $K_{\mathrm{R}}^{(w_j)}$ denote the numbers of packets transmitted from the source node and received at the receiver node on the path $w_j \in \mathcal{W}$, respectively. A measurement $\hat{y}_j$ of y_j is then given by $\hat{y}_j = -\log(K_R^{(w_j)}/K_S^{(w_j)})$. The proposed scheme obtains the estimate $\hat{\boldsymbol{x}} = (\hat{x}_1\ \hat{x}_2\ \cdots\ \hat{x}_N)^\top$ of $\boldsymbol{x}$ from $\boldsymbol{A}$ and $\hat{\boldsymbol{y}} = (\hat{y}_1\ \hat{y}_2\ \cdots\ \hat{y}_M)^\top$ by means of compressed sensing explained in the following section. The estimate $\hat{p}_{e_j}$ ($e_j \in \mathcal{E}$) of p_{e_j} is obtained by $\hat{p}_{e_j} = 1 - \exp(-\hat{x}_j)$. According to estimated link loss rates, all links are classified into lower link quality class $\mathcal{E}_{\mathrm{L}} \subseteq \mathcal{E}$ and higher link quality class $\mathcal{E}_{\mathrm{H}} = \mathcal{E} \setminus \mathcal{E}_{\mathrm{L}}$, where $e_j \in \mathcal{E}_{\mathrm{L}}$ if $\hat{p}_{e_j} = 0$ and $e_j \in \mathcal{E}_{\mathrm{H}}$ if $\hat{p}_{e_j} > 0$.

B. Compressed Sensing

When $M \geq N$, we can obtain $\boldsymbol{x}$ by solving (1). In active measurement schemes, however, $M < N$ is desirable in order to reduce traffic load, which means that the system of linear equations (1) is ill-posed. In this letter, we use compressed sensing to solve the problem. Compressed sensing enables us to obtain a true solution, which is known to be sparse a priori, from an underdetermined linear system by means of ℓ_1 norm optimization. To be specific, under the condition that almost all components of $\boldsymbol{x}$ are zero, $\hat{\boldsymbol{x}}$ is obtained by

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \ \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{2}$$

where ℓ_1 norm $\|\cdot\|_1$ is defined as $\|\boldsymbol{z}\|_1 \triangleq \sum_{i=1}^n |z_i|$ for $\boldsymbol{z} = (z_1\ z_2\ \cdots\ z_n)^\top \in \mathcal{R}^n$. Berinde and Indyk [11] show that if a routing matrix is the bi-adjacency matrix of a $(2k, \epsilon)$-expander graph with left-degree d, linear programming (2) estimates a k-sparse vector.

C. ℓ_1-ℓ_2 Optimization

For the direct application of compressed sensing to loss tomography, one of the key issues is the sparsity of $\boldsymbol{x}$, which will not be true in practical networks, since this requires that almost all links have loss rates of exactly zero. However, if we are interested in the identification of a limited number of links with higher loss rates, we can still apply compressed sensing by approximating the elements of $\boldsymbol{x}$ corresponding to lower link loss rates to be zero. This approximation, as well as the fact that only measurements $\hat{\boldsymbol{y}}$ of $\boldsymbol{y}$ is available, naturally leads us to an ℓ_1-ℓ_2 optimization, where the linear constraint in the original problem is relaxed to an inequality and is further replaced with a penalty in the cost function.

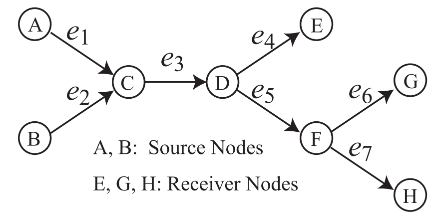

Fig. 1. Network topology with 8 nodes and 7 links.

With the ℓ_1-ℓ_2 optimization, $\boldsymbol{x}$ is estimated by solving the following optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\hat{\boldsymbol{y}} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 ,$$

where λ denotes the mixing parameter. $\|\cdot\|_2$ represents ℓ_2 norm and $\|\boldsymbol{z}\|_2$ is defined as $\|\boldsymbol{z}\|_2 \triangleq (\sum_{i=1}^n z_i^2)^{1/2}$. While the ℓ_1-ℓ_2 optimization problem can be solved effectively with various algorithms [8], we employ the iterative shrinkage-thresholding algorithm (ISTA) [8], [12] for simplicity.

The mixing parameter λ governs the trade-off between the estimation error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2$ and the sparsity of $\boldsymbol{x}$ [8]. This indicates that λ could be used to control the threshold of the link quality classifier. Namely, the larger λ is, the higher the threshold of link loss rate becomes.

IV. Performance Evaluation

A. Simulation Environment

In this section, we evaluate the proposed scheme with simulation experiments. We developed the simulation environment with MATLAB [13]. Fig. 1 shows the network topology with 8 nodes and 7 links used in the simulation experiments. There are two source nodes and three receiver nodes in the network topology. Packets are transmitted on 4 different paths $\mathcal{W} = \{w_i; i = 1, 2, 3, 4\}$, where $w_1 = \{e_1, e_3, e_4\}$, $w_2 = \{e_2, e_3, e_4\}$, $w_3 = \{e_1, e_3, e_5, e_6\}$, and $w_4 = \{e_1, e_3, e_5, e_7\}$. The corresponding routing matrix satisfies the 1-identifiable condition in [4, Table I]. We assume that packets are lost randomly on each link $e_j \in \mathcal{E}$ ($j = 1, 2, \ldots, 7$) with probability p_{e_j}. We also assume that one link has higher link loss rates than other links. In each iteration of simulation experiments, we randomly choose a link $e_{\mathrm{bad}} \in \mathcal{E}$, and assign link loss rates α to e_{bad} and β to all other links. We set $\alpha \in \{0.01, 0.05, 0.1\}$ and $\beta = 0.001$.

We evaluate the performance of the proposed scheme with estimated link loss rate $\hat{p}_{e_{\mathrm{bad}}}$, false negative rate $P_{\mathrm{FN}} = \Pr(\hat{p}_e = 0 |\ e = e_{\mathrm{bad}})$, and false positive rate $P_{\mathrm{FP}} = \Pr(\hat{p}_e > 0 |\ e \in \mathcal{E} \setminus \{e_{\mathrm{bad}}\})$. $\hat{p}_{e_{\mathrm{bad}}}$ is obtained by averaging the estimated loss rate of the corresponding link over N_{run} simulation runs. In the following simulation results, we set $K_{\mathrm{S}} = 1000$, $N_{\mathrm{run}} = 10000$. In order to validate the proposed scheme, we also estimate $\hat{p}_{e_{\mathrm{bad}}}$ by solving (2) with linear programming based on the *simplex method* as in [4].

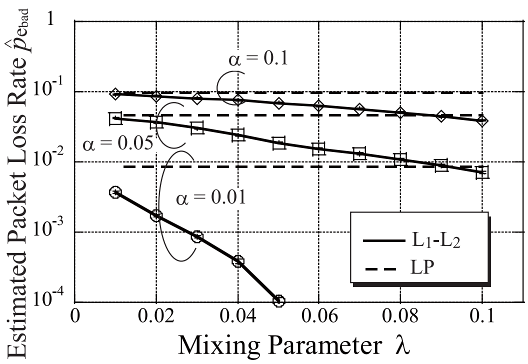

Fig. 2. Estimated link loss rate $\hat{p}_{e_{\mathrm{bad}}}$ vs. mixing parameter λ. 95% confidence interval is shown for each plot.

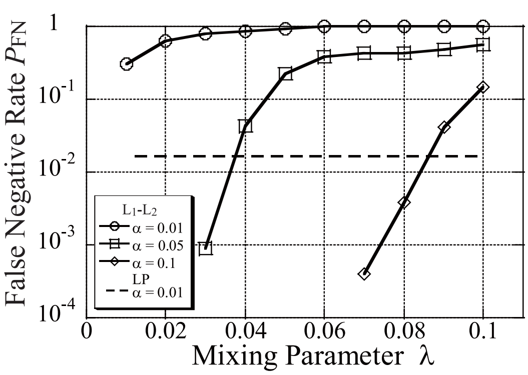

Fig. 3. False negative rate P_{FN} vs. mixing parameter λ. When using linear programming for $\alpha = 0.05, 0.1$, no false negative events are observed in the simulation experiments.

B. Simulation Results

In the following figures, "L_1-L_2" and "LP" correspond to the performance of the proposed scheme and linear programming (2), respectively. Fig. 2 shows estimated link loss rate $\hat{p}_{e_{\mathrm{bad}}}$ vs. the mixing parameter λ. We observe that $\hat{p}_{e_{\mathrm{bad}}}$ in the proposed scheme approaches $\hat{p}_{e_{\mathrm{bad}}}$ in linear programming as λ decreases. Fig. 3 shows P_{FN} vs. the mixing paramater λ. We observe that the proposed scheme achieves lower P_{FN} for smaller λ. While linear programming shows better performance in terms of $\hat{p}_{e_{\mathrm{bad}}}$ and P_{FN}, the proposed scheme can adjust the link quality classification capability according to mixing parameter λ. For example, when we set $\lambda = 0.03$, links with $\alpha \geq 0.05$ are detected with $P_{\mathrm{FN}} < 10^{-3}$. On the other hand, when we set $\lambda = 0.07$, links with $\alpha \geq 0.1$ are detected with $P_{\mathrm{FN}} < 10^{-3}$. Further, the proposed scheme has an advantage in the false positive rate. Fig. 4 shows P_{FP} vs. λ. We observe that the proposed scheme achieves lower P_{FP} than linear programming. This result agrees with the fact in [14] that in general, ℓ_1-ℓ_2 optimization has solutions with fewer nonzero components than linear programming.

V. Conclusion

In this letter, we proposed a link quality classifier using compressed sensing and evaluated its performance via simulation experiments. With the mixing parameter, the proposed scheme can adjust the threshold of the link quality capability according to the system requirement. Some technical issues remain in the proposed scheme. One is the evaluation of the

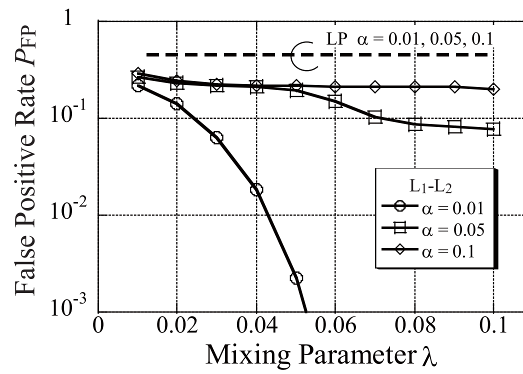

Fig. 4. False positive rate P_{FP} vs. mixing parameter λ.

performance when there are several higher loss rate links. This issue is closely related to established paths and network topologies. Actually, in [4], the k-identifiability problem is discussed for routing matrices in randomly generated networks. We also consider that in order to validate the compressed sensing approach to network tomography problems, the proposed scheme should be evaluated in larger network topologies. Because these issues are beyond the scope of this letter, we leave them as future works.

Acknowledgment

This research was supported in part by Grant-in-Aid for Young Scientists (B) of Japan Society for the Promotion of Science under Grant No. 22700069, No. 22760317, and No. 21760289.

References

[1] R. Cáceres, N. G. Duffield, J. Horowitz, and D. F. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2462–2480, Nov. 1999.

[2] M. Coates, A. O. Hero III, R. Nowak, and B. Yu, "Internet tomography," *IEEE Signal Process. Mag.*, vol. 19, no. 3, pp. 47–65, May 2002.

[3] M. H. Firooz and S. Roy, "Network tomography via compressed sensing," in *Proc. IEEE GLOBECOM*, Dec. 2010.

[4] M. H. Firooz and S. Roy, "Link delay estimation via expander graphs," arXiv:1106.0941v1, June 2011. Available: http://arxiv.org/abs/1106.0941.

[5] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," arXiv:1008.0919v1. Available: http://arxiv.org/abs/1008.0919, Aug. 2010.

[6] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[8] M. Zibulevski and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.

[9] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[11] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," Tech. Rep. MIT-CSAIL-TR-2008-001, MIT, Cambridge, MA, Jan. 2008.

[12] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math*, vol. 57, no. 11, pp.1413–1457, Nov. 2004.

[13] MATLAB, http://www.mathworks.com.

[14] J. J. Fuchs, "On sparse representation in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.