Selecting Measurement Paths for Efficient Network Monitoring on a User-Cooperative Scheme

Atsuo Tachibana   Shigehiro Ano

KDDI R&D   Laboratories
2-1-15 Ohara, Fujimino-shi, JAPAN
{tachi, ano}@kddilabs.jp

Masato Tsuru

Kyushu Institute of Technology
680-4 Kawazu, Iizuka-shi, JAPAN
tsuru@cse.kyutech.ac.jp

Abstract **A user cooperative monitoring scheme is considered in which many but variable users' off-the-shelf PCs are involved as measurement end-point nodes (called *beacons*) to actively measure the performance of end-to-end paths between beacons in order to detect end-to-end performance degradation and identify (narrow down) degraded network links. In the scheme, the measurement paths with beacons are dynamically selected by considering a trade-off between detection coverage and identification accuracy. From an ISP viewpoint, utilizing many users' PCs as a measurement infrastructure has a potential to greatly reduce deployment and maintenance costs while individual PCs may vary and are not always stable. The effectiveness and feasibility of our solution are suggested through model simulations.**

Index Terms—**Network monitoring, Network tomography, Path selection, Maximum coverage problem.**

I. INTRODUCTION

Network monitoring based on end-to-end performance measurements, which likely reflect user experience, is important for Internet Service Providers (ISPs) to provide network services in a reliable and efficient manner. In such monitoring, not only detecting performance degradation of network paths but also identifying causal troubled links is necessary. For this, network tomographic approaches have been studied, which infer network-internal performance characteristics by monitoring end-to-end multiple paths traversing the targeted network links instead of directly monitoring each of those links. However, it is still challenging to develop deployable solutions for large-scale actual network operations, due, e.g., to the high deployment and/or maintenance costs.

 In this paper, we propose a scalable solution that actively measures the performance of many end-to-end paths and identifies (narrows down) the causal links responsible for the performance degradation, utilizing many but variable users' PCs as measurement end-point nodes (called beacons). Because recent ISP services provide customer premise equipment (CPE) (e.g., set-top boxes, VoIP phones) to offer a variety of add-on services and hence the demarcation point is moving closer to the users, the user perspective approaches have become more important for today's network services. A relevant issue here is how to let many users cooperate in the measurement system. We believe, however, there are some possible scenarios with some trade-off such as scalability, reliability, and cost. For example, an ISP may raise reliable volunteers from their employees or an ISP may invite cooperative users by offering economic incentives. Note the proposed scheme is applicable not only to user PCs but also to CPEs and dedicated measurement boxes managed by the targeted ISP. However, a CPE-only approach may be insufficient. For instance, the locations of CPEs are limited within the ISP, and thus inter-AS links may not be monitored. In addition, CPEs are often designed with the minimum development cost and do not have enough resource for monitoring.

II. RELATED WORK

Recent research on network tomography has been successful in estimating the performance characteristics of distinguishing links of some interest and/or in identifying degraded links solely by multiple end-to-end path measurements without investigation of packet-level correlation, called performance tomography. A basic model of performance tomography can be formulated as linear equation Eq. 1. (e.g., a survey [1])

$$\mathbf{Y} = \mathbf{A}\,\mathbf{X} \qquad (1)$$

where $\mathbf{Y}$ is the vector of end-to-end performance measurement observables, $\mathbf{X}$ is the vector of unknown quantities to be estimated, e.g., average (or minimal) packet delays or log{1 - loss rate} on individual links, and $\mathbf{A}$ is a matrix (referred to as *routing matrix*) which associates $\mathbf{Y}$ with $\mathbf{X}$ reflecting the network topology. Routing matrix $\mathbf{A} = (a_{ij})$ encodes whether link l_j belongs to path (or route) r_i, i.e., a_{ij} takes a value of 1 if path r_i includes link l_j, and 0 otherwise. Here, if $\mathbf{A}$ is a full column rank, we can solve Eq. 1 and obtain $\mathbf{X}$ directly, but unfortunately $\mathbf{A}$ is typically rank-deficient. In response to this, various methods have been used to estimate $\mathbf{X}$ by obtaining or assuming additional information (e.g., [2, 3]).

 As the other important part for effective large-scale network measurements, it has been studied how to determine good placements of beacons and active measurement paths among them under some resource constraints (e.g., [4, 5]). In a variety of studies, however, few are found on designing a possible user-cooperative scheme that actively measures the performance of many end-to-end paths and identifies the links responsible for the performance degradation with dynamically variable beacons. Most of existing designs assume fixed

(stable) beacons and propose complex processes of path selection, while effective, that may take more than several dozen minutes in a large network.

III. THE PROPOSED SCHEME

A. Overview of the proposed scheme

Our measurement framework consists of servers and beacons. The servers determine the network paths (pairs of beacons) to be measured based on the information of available beacons and a known routing matrix of the targeted network. Each beacon measures the end-to-end (one-way) network performance with other beacons under the control of the servers. The measurement results are periodically collected in the servers and analyzed. Hereafter we assume a single central server for simplicity, but multiple or hierarchal servers can be considered.

We repeat the following four steps every measurement period.

(i) Each beacon monitors its own internal states and resources, such as CPU and memory usage, to judge its availability. Sophisticated resource characterizations are found in literature (e.g., [5]).

(ii) The server calculates the routes among available beacons. Instead of inefficiently employing traceroute in a full-mesh manner, we suppose that the management information of the targeted ISP's internal topology is available, e.g., by monitoring Interior Gateway Protocol (IGP).

(iii) The server selects a set of measurement paths under measurement load constraints and requests the corresponding beacons to monitor the paths. Since it is essential to promptly detect performance degradation by comprehensively monitoring the network and to identify the degraded links in order to take efficient action to mitigate or analyze the performance degradation, the measurement paths are selected based on two criteria given in Sec. III-B.

(iv) The beacon measures one-way packet losses and delays between other beacons by transmitting a series of probing packets for a certain measurement period (e.g., 60s). After completing measurement, the beacon uploads the results to the server and the server checks whether or not some specific performance metrics are worse than the predefined thresholds. If performance degradation is detected on at least one measurement path, the server tries to identify the causal links responsible for the degradation. Among various existing methods, in our simulation in Sec. IV-C we adopt a simple approach in [6] that infers a possible set of the degraded links in which the number of degraded links is minimized.

B. Selecting measurement paths

In the proposed scheme, as the number of paths simultaneously monitored increases, both the network coverage (the set of links monitored in the targeted network) and the identification accuracy (the range of inferred causal links possibly responsible for the performance degradation) are likely to be improved, but at the same time, the incurred load on the network and beacons must increase. Therefore, we should select a small set of paths to measure at one

measurement period and change the set of paths adaptively to accomplish comprehensive network monitoring and to narrow down the identification of degraded links over multiple measurement periods. Note that we set one measurement period so short that the performance degradation will continue for multiple measurement periods based on the observation that Internet congestions have some constancy [7].

On the other hand, there is a trade-off relationship between these two different objectives under common constraints. Here we adopt a simple method to relax this trade-off by switching two objectives at each measurement period if needed. Given the measurement results of the previously selected set of paths, we first ascertain whether performance degradation occurs or not. If degradation is not detected on any path, we proceed to the path selection for comprehensive network monitoring at the next measurement period. Otherwise (i.e., degradation is detected) we proceed to the causal degraded link inference by using an existing tomographic approach. In this inference, if the location of the causal links are narrowed down and identified, we proceed to the path selection for comprehensive network monitoring at the next measurement period. On the other hand, if the current measurement results are insufficient to identify degraded links, we proceed to the path selection for narrowing down the identification at the next measurement period. Optionally, after a set of sufficient paths for accurate identification is selected, additional paths can be selected to increase network coverage as long as the constraints allow.

1) Maximizing network coverage

Finding a set of measurement paths at a measurement period to maximally cover the entire set of network links is a classic maximum coverage problem. To maximize the network coverage over multiple measurement periods, we introduce "weight" of each link to indicate the importance of being covered at the next measurement period, and numerically solve a weighted version of the maximum coverage problem that maximizes the sum of weights. The weight is updated every measurement period to increase the network coverage over multiple measurements. Namely, the weight of a link that was rarely covered in the past measurements should be higher and vise-versa. We adopt here an exponential weighted moving average to reflect time elapse as follows.

$$w_{j,t} = \alpha\, q_{j,t-1} + (1 - \alpha)\, w_{j,t-1} + \delta \qquad (2)$$

where $q_{j,t}$ is the binary function that takes a value of 1 if link l_j is NOT covered at time t, and 0 otherwise. The constant α ($0 < \alpha < 1$) is a smoothing parameter that controls a contribution of the latest result $q_{j,t-1}$ (0 or 1) to the moving average value, and small positive constant δ (e.g., δ = 1e-5) keeps the value of $w_{j,t}$ always positive.

Suppose the targeted network contains m links to be monitored. At a measurement period t, n beacons are available (i.e., $n(n-1)$ candidate paths) and weight $w_{j,t}$ is assigned to link l_j. For each candidate path (a pair of available beacons), let route r_i be a set of coverable links in the targeted network on the path and the whole collection of paths be $R = \{r_1, r_2, \ldots, r_{n(n-1)}\}$. The problem is extracting a set of paths G from R to maximize the sum of weights $W(G)$ contained in their union.

Two constraints are considered. One is the load on each beacon: the maximum number of simultaneous measurements on each beacon b_k should be less than the threshold B_k. B_k (k=1, 2, .., n) may statically be set based on the capability of the beacon or may dynamically be set based on the current resource conditions. The other is the measurement load on the network: the total number of overlapped measurement paths at link l_j is limited to be less than the threshold L_j (j=1, 2, …, m). Suppose binary function $C(i, k)$ that takes a value of 1 if beacon b_k is used (as a sender or a receiver) for a measurement on path r_i and 0 otherwise, and $C'(i, j)$ that takes a value of 1 if path r_i traverses link l_j and 0 otherwise. The weighted version of the maximum coverage problem is formulated as follows.

$$\text{maximize} \quad W(G) = \sum_{l_j \in \cup_{r_i \in G}} w(l_j) \tag{3}$$

$$\text{s.t.} \quad \forall b_k, \ \sum_{r_i \in G} C(i, k) \leq B_k, \quad \forall l_j, \ \sum_{r_i \in G} C'(i, j) \leq L_j \tag{4}$$

For the maximum coverage problem, the greedy algorithm has good approximation [8]. Hence, we use a greedy heuristic in which we start with an empty set G for selected paths and iteratively pick the path that contains the uncovered links to maximize the total weight under the constraints.

2) Maximizing the identification accuracy

To infer the causal links responsible for performance degradation by a network tomographic technique with a limited number of measurement paths, selecting paths so as to increase the number of linearly independent paths (i.e., the number of basis of a routing matrix for the selected paths) is essential. The set of links shared by a set of degraded measured paths includes the candidates of causal links but the characteristics of individual links in the set cannot be determined. In order to locate the real causal degraded link from the candidates, it is essential to add new paths so that the set of candidate links is divided into multiple subsets i.e., the number of links that are shared by different (linearly independent) sets of paths increases.

Because the constrained basis selection problem is NP-hard [4], we utilize a similar solution in Sec. III-B-(1), i.e., a greedy heuristic again. We iteratively add paths (rows) that expand the basis of a routing matrix constructed by the selected paths. As the number of paths is limited to meet measurement constraints, the resulting routing matrix may be insufficient to provide a good identification of degraded links. In such a case, the candidates of degraded links should be narrowed down over the following measurement periods or network operators have to conduct other investigations to locate the real degraded links.

To properly handle the order of selected path in the greedy algorithm, we introduce another weight $w'_{j,t}$, which is assigned to each link to indicate the preference to be identified. To identify degraded links promptly in a fair fashion when performance degradations occur at multiple locations, $w'_{j,t}$ is defined as the time duration in which link l_j has not been identified (distinguished), and calculated repeatedly at every measurement period based on the past history of the inference results by Eq. 5. Let $s_{i,t}$ be the sum of $w'_{j,t}$ for all links

contained in each path r_i (Eq. 6). Note that $w'_{j,t}$ of all links included in non-degraded paths is zero and we can calculate $s_{i,t}$ only for the paths that include at least one candidate link.

$$w'_{j,t} = \begin{cases} w'_{j,t-1} + 1 & \text{if } l_j \text{ is NOT identified at } t \ (t > 1), \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

$$s_{i,t} = \sum_{l_j \in r_i} w'_{j,t} \tag{6}$$

We sort all paths that have positive value of $s_{i,t}$ by $s_{i,t}$ and have the ordered set of paths R_o. At the beginning, we set an empty set G' and a null routing (sub-)matrix $\mathbf{A_s}$ whose rows and columns correspond to selected paths and included links, respectively. Then we iteratively add paths (rows) that expand the basis of $\mathbf{A_s}$ to $\mathbf{A_s}$ until all possible paths under the measurement load constraints are selected or the rank of $\mathbf{A_s}$ becomes the same as the rank of the full-size (full-mesh) routing matrix $\mathbf{A}$.

IV. SIMULATION EVALUATION

A. Setup

We used two random topologies of Waxman model and Barabasi-Albert model (BR1 and BR2, respectively) and one real-world ISP backbone topology in Japan (BB1) as illustrated in Table 1. On each topology, we randomly classified all (directed) links into three types (high-, medium- and low-bandwidth links) with different measurement load constraints L_j=50, 10, 5, respectively. As the measurement constraints on beacons, we simply adopted uniformly one, i.e., $\forall k$, B_k=1, meaning that each beacon can send or receive the probing packets with only one beacon. To model the statistical distribution of beacons (user PCs) connected to each router, we assigned the number of routers u each of which is connected by v different beacons to be approximated by a function $u = a \cdot v^{-0.73}$, where 0.73 reflects the geographical diversity on the number of broadband service contracts in Japan [9], and constant a is set in proportion to the total number of beacons varying at each simulation. Parameter α in Eq. 2 is experimentally set to 0.5. Initial weight $w_{j,0}$ and $w'_{j,0}$ of all links is set to δ and 0, respectively.

TABLE 1 NETWORK TOPOLOGY

	#routers	#links	# classified links (high, med, low)
BR1	1000	4000	1000, 1000, 2000
BR2	1000	7934	2000, 2000, 3934
BB1	100-1000 [(*)]	2078	500, 1000, 1078

[(*)] For topology BB1, only the order of the number of routers is given due to reasons of company confidentiality.

B. Evaluation of network monitoring coverage

We evaluated the cumulative coverage of monitored links over multiple measurement periods as follows.
1. On a given topology of routers with link load constraints, we randomly connect beacons based on a model shown in subsection A. Note a shortest hop route is assumed as a path between beacons.
2. In path selection, the weights of links in Eq. 2 reflect a past measurement history. Therefore, to avoid influence of initial setting, we first perform dummy simulation for

twenty measurement periods before starting evaluation.

3. Then we perform simulation for three measurement periods to evaluate the cumulative network coverage.
4. We repeat steps 1-3 with a different number of beacons.

Figure 1 shows the cumulative coverage on three networks. Thanks to dynamically changeable paths in our method, the cumulative coverage over two and three measurement periods becomes essentially improved.
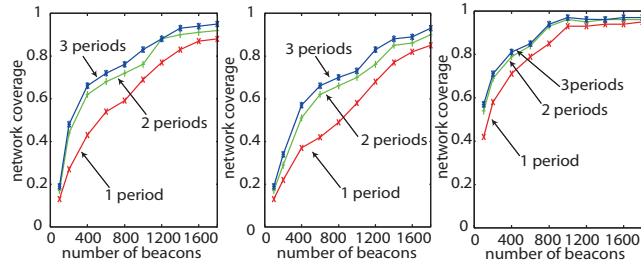

Fig.1. Network coverage (From left to right: BR1, BR2, BB1).

C. Evaluation of identification accuracy of degraded links

We evaluated how accurately the degraded links are identified in each process of comprehensive network monitoring and of narrowing down the identification over multiple measurement periods as follows.

1'. After the same steps 1 and 2 in Sec. IV-B, we randomly choose "the targeted links" assumed to be degraded from all links contained in at least one of selected measurement paths. We assign a high packet loss rate of 0.05 to the targeted links and low packet loss rates uniformly distributed between 0 and 0.001 to other links. An end-to-end packet loss rate of each measurement path to be observed is calculated from the loss rates of individual links along the path.
2'. Path degradation is detected using a packet loss rate threshold: 0.01. Since degradation is always detected in this setting, the network tomographic inference of causal links based on [6] is performed. Then the links inferred as being degraded are recorded.
3'. Then we check whether the path selection for narrowing down the identification is used or not in the algorithm. If all targeted links are inferred with the best possible granularity in step 2', we stop this simulation. Otherwise, the next set of paths is selected for narrowing down the identification and the next inference for the targeted links is performed based on the measurement results of the selected paths. Then the links inferred as being degraded are recorded again. Until all targeted links are inferred, we continue over three measurement periods.
4'. We repeated above steps 1'-3' with different sets of targeted links for ten times. We performed the same simulations with different numbers, i.e., 1, 5, and 10, of the targeted links (in step 1') on the same topology.

Figure 2 shows the average number of links that are incorrectly identified as being degraded in case of five targeted links. The results suggest the path selection for narrowing down the identification is usable and effective to decreasing the number of non-degraded links especially in case of a large number of beacons. Note the results with single

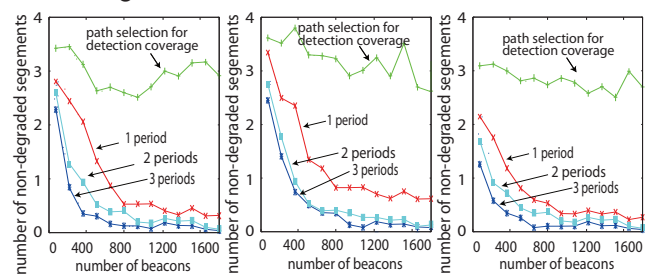and ten targeted links showed similar tendencies.

Fig.2. Identification accuracy (From left to right: BR1, BR2, BB1).

D. Evaluation of computation time

We evaluated the average computation time of our algorithm of selecting measurement paths through experiments on a PC with Intel(R) Xeon(TM) 3.4GHz CPU. To find a set of paths for maximizing the network coverage with 500 beacons on BR1 topology with 4000 links, our algorithm took about 30 seconds in average. In contrast, the average calculation time to find a set of paths for maximizing the identification accuracy in the case of 10 targeted links took only a few seconds (without the option of finding additional paths to increase network coverage). The results suggest a feasibility of the proposed path selection algorithm in a large-scale network.

V. CONCLUDING REMARKS

In this paper we have proposed a user cooperative measurement scheme utilizing many users' off-the-shelf PCs as measurement beacons. In the scheme, the measurement paths are dynamically and adaptively selected by considering operational constraints, so as to maximize detection coverage and identification accuracy. The effectiveness of our solution is demonstrated through model simulations.

We already verified our scheme through measurement experiments on Japan's commercial Internet with a prototype implementation with Windows OS beacons and about 500 volunteer users. Evaluation through more comprehensive Internet-wide experiments remains future work.

References

[1] R. Castro et al. "Network Tomography: Recent Developments," Statistical Science, 19(3):499-517, 2004.

[2] D. Ghita et al. "Netscope: Practical Network Loss Tomography," *Proc. IEEE INFOCOM*, 2010.

[3] H. H. Song et al, "NetQuest: A Flexible Framework for Large-Scale Network Measurement," *Proc. ACM SIGMETRICS* 2006.

[4] Y. Zhao et al, "Towards Efficient Large-Scale VPN Monitoring and Diagnosis under Operational Constraints," *Proc. IEEE INFOCOM*, 2009.

[5] H. H. Song and P. Yalagandula, "Realtime: Real-time End-to-end Network Monitoring in Large Distributed Systems," *Proc. IEEE COMSWARE* 2007.

[6] A. Tachibana et al., "Locating congested links over the internet based on multiple end-to-end path measurements," *IEICE Trans. Commu.*, E89-B(4):1099–1109, 2006

[7] Y. Zhang et al., "On the constancy of Internet path properties," *Proc. SIGCOMM IMW*, 2001.

[8] A. Pathria, "Analysis of the greedy approach in covering problems," 1998.

[9] Ministry of Internal Affairs and Communications. Information and Communications Statistics Database. http://www.soumu.go.jp/johotsusintokei/english/.