Locating High-loss Links for OpenFlow Networks by Multiple Hosts to Probe Packets

著者	Nguyen Minh Tri, Shibata Masahiro, Tsuru Masato, Kawaguchi Akira
journal or publication title	2021 23rd International Conference on Advanced Communication Technology (ICACT)
page range	455-460
year	2021-03-10
URL	http://hdl.handle.net/10228/00008383

doi: https://doi.org/10.23919/ICACT51234.2021.9370812

Locating High-loss Links for OpenFlow Networks by Multiple Hosts to Probe Packets

Nguyen Minh Tri*, Masahiro Shibata*, Masato Tsuru*, Akira Kawaguchi†
*School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan
†The City College of New York of The City University of New York, U.S.
tri.nguyen-minh414@mail.kyutech.jp, shibata@cse.kyutech.ac.jp, tsuru@cse.kyutech.ac.jp, akawaguchi@ccny.cuny.edu

***Abstract*—We previously proposed a measurement framework for OpenFlow-based networks to promptly locate high-loss links with a small load incurred by the measurement on both the data-plane (e.g., the number of transmissions of probe packets on each link) and the control-plane (e.g., the number of accesses to switches) until locating all high-loss links. One of key components is the multicast measurement route of probe packets traversing all links in both directions. However, the previously proposed Eulerian cycle-based measurement route scheme called the backbone-and-branch tree (BBT) that uses only a single measurement host (MH) may build a too long measurement path in a large network, resulting in a low measurement accuracy and an intolerance to very high-loss, e.g., failure, links located in upstream of a measurement path. Therefore, in this paper, we newly propose an enhancement of the BBT with multiple MHs, called BBT-mMH, which can control the measurement path lengths to maintain an acceptable measurement accuracy with a small overhead on both the control-plane and data-plane. The numerical simulation demonstrates potential benefits of our proposal.**

***Keywords*—active measurement, link loss rate, multicast, OpenFlow, flow statistics, SDN**

I. INTRODUCTION

As practical realization of Software Defined Networking (SDN) technology, the OpenFlow-based networks become more widespread because of the capabilities for flexible and dynamic traffic engineering especially in cloud computing and in any form of geographically-distributed but centrally-managed computing and services [1]. Such a dynamic traffic engineering to manage the network service qualities requires actively measuring all links to monitor, detect and locate performance-degraded links.

In OpenFlow, the control plane and data plane are decoupled. In the data plane, each switch forwards packets based on per-flow rules managed by a controller and records the statistical information (flow-stats) of each flow passively. In the control plane, the flow-stats can be collected by the controller to monitor the network traffic. However, such flow-stats collection by accessing to each switch has a trade-off between the measurement accuracy and the load incurred on switches and the control network. FlowSense [2] can calculate the network utilization without additional measurement cost, but it cannot trace quickly changed links. In PayLess [3], the authors proposed a dynamic algorithm to balance the request frequency and accuracy. The active measurement approach is also considered in OpenFlow networks, in which a measurement host sends and receives probe packets to measure the packet loss, delay, and the round-trip-time, mainly based on unicast probe packets. In [4], the authors presented a probing scheme that can cover all links in both directions with minimizing the number of flow entries. For datacenter networks, an effective probe matrix is designed to locate real-time failures in [5].

Several recent works challenged to investigate the optimality on network-topographic approaches to identify link performance metrics from end-to-end measurements of probing packets among measurement hosts. In [6], a problem of placing measurement hosts and selecting measurement paths to identify link metrics from end-to-end measurements is considered to minimize the number of measurement hosts. The proposed method can robustly identify all link metrics under an existence of disruption of measurement paths due to predictable and unpredictable link failures. In [7], a problem of identifying failure nodes from end-to-end measurements, and upper bounds on the maximum number of identifiable nodes are analytically provided given the number of measurement paths and different constraints on the network topology, the routing scheme, and the maximum path length. Note that it can be applied to not only failure nodes but failure links by modeling it as logical nodes. However, although such theoretical results on capabilities and limitations on network-topographic approaches give useful insights generally, they cannot be directly applicable to our hybrid approach explained in this paper, in which packet probing on multicast measurement paths traversing every link only once are combined with probed packet monitoring on selected switches.

Differently from those existing work, we proposed a measurement framework using actively multicasted probing packets combined with passively recorded flow-stats information in order to efficiently and promptly locate high-loss links on an OpenFlow-based full-duplex network. An initial work was reported in [8], followed by some extensions by our group [9], [10]. A Bayesian-based network tomography is complementarily used in [9] in order to refine candidates for high-loss links to optimize the retrieval order of accesses to switches. In [10], the results of past measurements are used in order to place high-loss-prone/failure-prone links at the end parts of route tree to increase the measurement accuracy and also decrease the necessary number of accesses to switches. Both extensions aim to adaptively improve the measurement routes and the

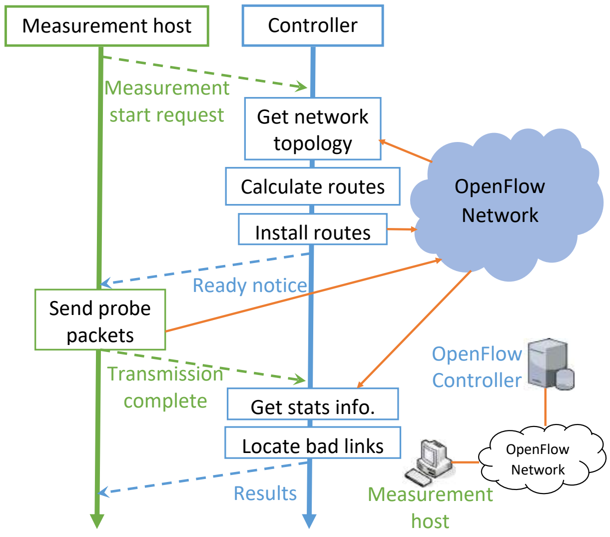


Fig. 1. Measurement process to locate bad links [8]

order of accesses in a repeated monitoring scenario, although they are still based on the baseline route scheme in [8]. An essential extension on a better route scheme was proposed in [11] as explained in Section 3-A later. However, that improved route scheme is still for a single measurement host so that it is limited in terms of controllability of the measurement path lengths, especially in a large-scale network. In this paper, therefore, we newly propose a further improvement of the route scheme by introducing multiple measurement hosts.

II. SYSTEM OVERVIEW

Our framework is for OpenFlow-based full-duplex networks consisting of the OpenFlow controller (OFC) and OpenFlow switches (OFS), with the measurement host (MH) that sends a series of multicast probe packets traversing all links in the network. We assume only the standard functions of OpenFlow. An MH is directly connected to an OFS (called "measurement node"). The input port of the measurement node connecting the MH is called "root port".

The measurement process starts when the MH sends a measurement start request message to the OFC, as in Fig. 1. Then, the OFC obtains network topology, calculates probe packet routes, and installs them to OFSs. After that, the MH starts actively sending the probe packets. Each probe packet (or a copy of it as multicasting) starting from the root port travels through each link once and only once (in each direction of a full-duplex link separately) and is discarded at a leaf port (an input port of the last OFS of the measurement flow). A measurement path from the root port to a leaf port is called "terminal path". A sequence of adjacent directed links along a path is called "segment" as a part of a terminal path. The number of directed links on a terminal path is the terminal path length. The number of probe packets arriving at an individual input port on each OFS is passively recorded as flow-stats at each OFS and then, if needed, is collected by the OFC. Finally, the OFC calculates the packet loss rate on a link or a segment between two ports by taking the difference between the numbers of arriving probe packets at those two ports; the packet loss rate is compared with a threshold to detect a high-loss link or a high-loss segment. This process is recursively performed until locating all high-loss links.

Two key components of the framework should be carefully designed: (i) the multicast measurement route of probe packets traversing all links in both directions, and (ii) the order of access from the OFC to OFSs in selectively collecting the flow-stats. To avoid a concentration of probe packets at links near the MH, especially in large networks, each probe packet should traverse each link one and only once. Two possible options are a unicursal-based unicast route and a tree-based multicast route. On the other hand, to reduce the number of accesses to OFSs, the order of accesses should be dynamically decided so as to narrow the segments which likely include lossy links to finally locate the high-loss links quickly, as explained in Sec. 4 later.

On designing the measurement route in (i), the terminal path length affects the efficiency in locating lossy links as well as the accuracy (reliability) of measurement results when a number of probe packets are lost at links located in upstream of a terminal path. As a good aspect, long terminal paths allow a small number of terminal paths, i.e., a small number of leaf ports needed to initially access, resulting in a small number of accesses to OFSs to locate all high-loss links. On the other hand, in a long terminal path, when a number of links on a path are not actually lossy but with a light loss rate, an accumulated loss rate over those links will cause a wrong decision to narrow a segment that likely includes actual lossy links, resulting in an unnecessary increase of accesses to OFSs. Furthermore, also as a bad aspect, a long terminal path requires a large number of probe packets to operate accurately because an incidental reduction of probe packets in upstream, e.g., due to heavy packet losses at a failure link, will degrade the reliability of measurement results for links in its downstream. Therefore, it is required to properly control the terminal path lengths of measurement route mainly depending on the targeted loss rate to be located and the expected distribution of loss rates for other normal lightly lossy links in the network.

III. ROUTE SCHEME DESIGN

A. The Backbone-and-Branch Tree Route Scheme (BBT)

In our previous work [11], we proposed a route scheme called the backbone-and-branch tree route scheme (BBT) and showed its advantage to two extreme cases, a unicursal-based unicast route and a shortest path tree-based multicast route. In this subsection, the BBT scheme is briefly explained as illustrated in Fig. 2. The Eulerian cycle algorithm is used to build backbone paths in the original undirected graph (network). Since an Eulerian cycle exists if and only if the graph consists of only even-degree vertices, first we need to remove all links incident to odd-degree vertices (nodes) temporarily, called "omitted links". Note that an Eulerian cycle is not unique. Then, we generate a backbone cycle by using the Eulerian cycle algorithm to cover all remaining undirected

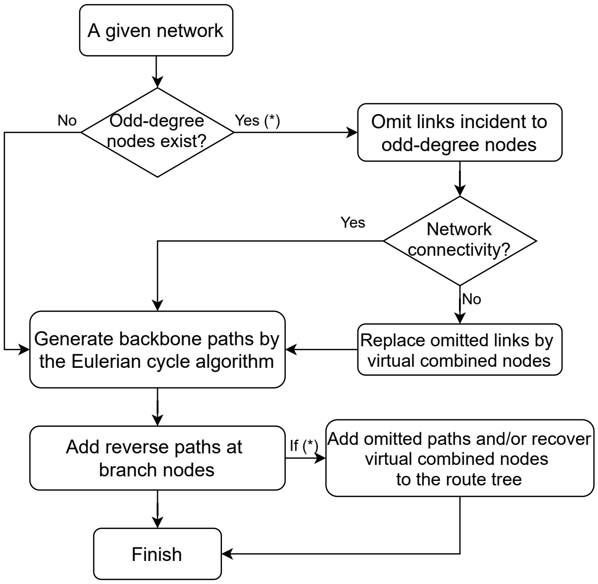


Fig. 2. BBT route scheme flowchart

links. From the generated backbone cycle, in [11], we build one or more backbone paths. Suppose a simple example of an ideal network topology with no odd-degree nodes in Fig. 3d. There are two options. The BBT T1 has one backbone path which is the full Eulerian cycle, illustrated by the bold line in Fig. 3a. The BBT T2 has two halves of the Eulerian cycle as backbone paths to avoid too long terminal paths, see the bold line in Fig. 3b.

After building the backbone path(s), we divide each backbone path into multiple backbone segments with almost the same length. At the end node of each segment, called branch node, the reverse direction segment of route on the backbone path is added as extension of the route toward the measurement node, called the reverse path. This is necessary because both directions of each full-duplex link should be traversed by a measurement path. Each reverse path has the same length with its backbone segment but the opposite direction, see dashed lines in Fig. 3a-c. Finally, we integrate additional paths of temporally removed (omitted) links into the route tree. Those operations eventually construct a route tree consisting of multiple terminal paths for multicast measurements.

As discussed in Sec. 2, the terminal path length should be controlled. In the previously proposed BBT route scheme, the terminal path length is determined by the number of backbone paths which is limited by the degree of the measurement node (an OFS accommodated with the MH). For example, if the degree is two, we can build two backbone paths at maximum, which may lead to long terminal paths. In the next subsection, to make it possible to reduce the terminal path lengths in response to the network topology and link conditions flexibly, we explain a new route scheme based on the BBT by introducing multiple MHs.

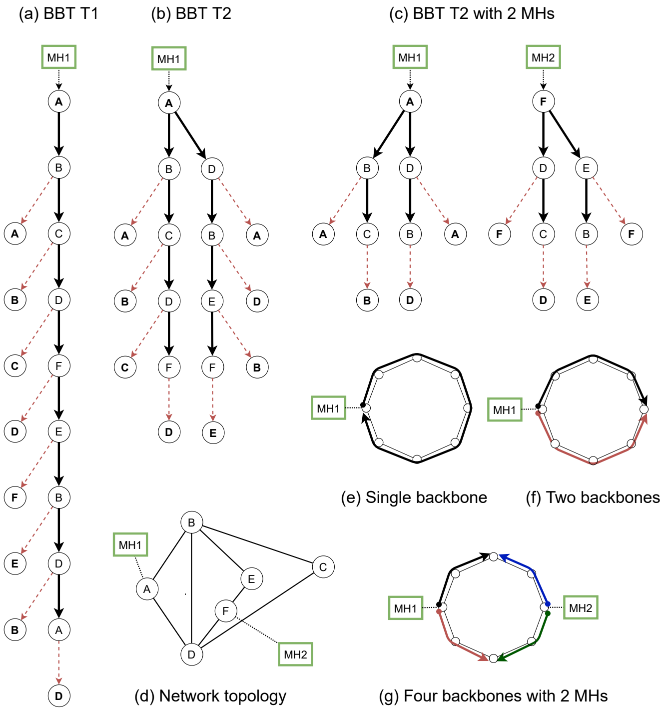


Fig. 3. Examples in BBT and BBT-mMH route schemes

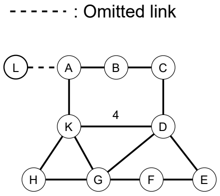
B. The BBT with Multiple Measurement Host (BBT-mMH)

As shown in the BBT scheme, we can build one or two backbone paths from a measurement node with the degree of two, see Fig. 3e-f. If there are multiple measurement nodes in the Eulerian cycle, we can build at least two backbone paths from each of them. Two backbone paths from two measurement nodes neighboring on the Eulerian cycle meet together at the middle. By using more than one measurement nodes, a long Eulerian cycle can be divided into multiple short backbone paths, see Fig. 3g.

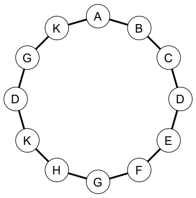
To reduce the maximum length of backbone paths, the measurement nodes will be located uniformly (with an equal distance) on the Eulerian cycle. For example, in the network topology Fig. 3d with two measurement hosts MH1 at A and MH2 at F with the degree of two, the route tree is shown in Fig. 3c. From MH1, there are two backbone paths: the A-B-C path and the A-D-B path. From MH2, two backbone paths are the F-D-C path and the F-E-B path. Then, for each backbone path, the reverse path is added. It is worth noting that the number of terminal paths does not change in BBT and BBT-mMH as shown in Fig. 3a-c.

C. The Location and Number of MHs

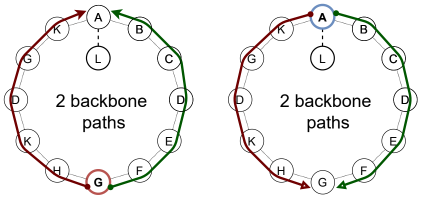
A possible option to reduce the length of backbone paths is to place a measurement node at a node with a higher degree. In general, the degree of measurement node determines the maximum number of backbone paths from this node. Therefore, by selecting measurement nodes with a degree more than two, we can construct more backbone paths. For example, in a network as Fig. 4a, we have its Eulerian cycle in Fig. 4b.



(a) Network topology

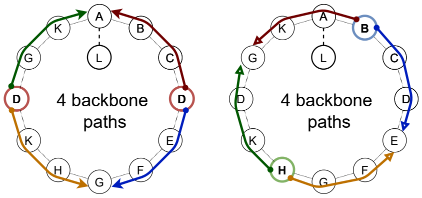


(b) Eulerian cycle



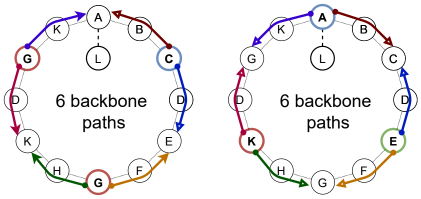
(c) One MH at a omitted link opposite

(d) One MH at a omitted link



(e) One MH at a 4-degree node

(f) Two MHs at 2-degree nodes



(g) Two MHs at 4-degree & 2-degree nodes

(h) Three MHs

Fig. 4. Example of the MH selection

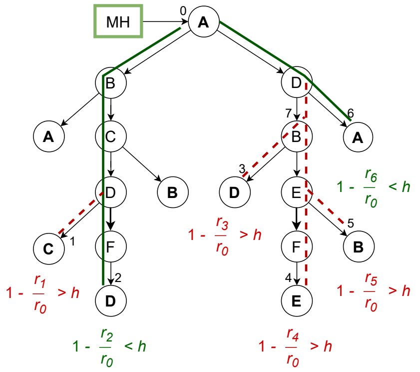


Fig. 5. Example of the accesses to switches in locating high-loss links

If the MH is connected to a 2-degree node, we only construct two backbone paths from this node, as in Fig. 4d. On the other hand, from 4-degree measurement node D, we can build four backbone paths, see Fig. 4e. If an MH can be placed at a high-degree node, we can get the benefit equivalent to using multiple MHs connecting to a low-degree node, i.e., we can reduce the number of MHs while keeping the number and length of terminal paths, see Fig. 4e, f and Fig. 4g, h.

Another factor giving an impact on the route scheme is the relationship between an MH and omitted links' locations. Each omitted link is added into the route tree as an individual terminal path. The length of this terminal path includes the backbone segment from the MH to the omitted link and the omitted links' length. Therefore, MHs should be selected at nodes that are closed to omitted links to reduce the lengths of terminal paths including those omitted links as well as the average length of terminal paths of the route tree. For example, in Fig. 4, two extreme locations of the omitted link are examined, i.e., at the beginning (Fig. 4d, h) and at the end (Fig. 4c, g) of backbone paths in a measurement route tree. The number of terminal paths and the path length of route schemes is shown in Table I. Note that the implicit link distance is 4 in this topology; that is, in Fig. 4b, each displayed link between two named nodes includes 4 hidden undirected links connecting 3 unnamed intermediate nodes, thus the total number of links is 52 (104 in both directions).

By using multiple MHs, we can reduce the length of terminal paths and overcome the shortcomings of a long path as discussed in Sec. 2. Moreover, the degree and location of measurement nodes can be considered to reduce the number of MHs and the maximum and average lengths of the terminal paths. In conclusion, the proposed route scheme can control the lengths of terminal paths, which will affect the measurement accuracy depending on an average loss rate of links along a terminal path and a desired threshold value of link loss rate in the target network.

IV. THE ORDER OF ACCESS TO SWITCHES

The OFC detects and locates high-loss links based on their packet loss rates. A link is regarded as high-loss if and only if its loss rate exceeds threshold value h; h is a design parameter that represents the target link quality to be maintained. The packet loss rate (PLR) of a segment (also a terminal path or a link) from ports i to j is defined as $PLR = 1 - \frac{r_j}{r_i}$, where r_i and r_j are the numbers of probe packets arriving at switch ports i and j, respectively. The OFC can collect r_i and r_j as the flow-stats by accessing the corresponding ports of some switches. Note that if the number r_i of probe packets used for this segment is not enough, the measured loss rate PLR may not be accurate. Assuming accurately measured PLR, if the PLR of a terminal path is less than h, this terminal path does not include a high-loss link. Otherwise, this terminal path is likely to include one or more high-loss links. The correlation among terminal paths in terms of the degree of packet loss can be used to narrow the search scope.

Fig. 5 illustrates an example of the process to locate high-loss links on BBT T2 measurement route in Fig. 3b. If a terminal path is high-loss but all other terminal paths sharing at least one link with the high-loss terminal path are not, the high-loss links are located within a segment between the leaf port and the nearest branch port of the single high-loss terminal path. The dashed line on the left part in Fig. 5 illustrates an example of this case. Then, the binary search algorithm is used to locate all high-loss links by accessing appropriate ports. On the other hand, if there are multiple high-loss terminal paths, the next port to access is the most commonly port shared by those paths. An example of this case is shown in the right part of Fig. 5, the OFC accesses the port 7 of node B.

TABLE I
THE NUMBER OF TERMINAL PATHS AND PATH LENGTH

	MHs	Paths	Average	Min	Max
BBT T2 Fig. 4c	1	13	19.08	8	32
BBT T2 Fig. 4d	1	13	17.23	8	28
BBT T4 Fig. 4e	1	13	12.62	8	20
BBT T4 Fig. 4f	2	13	12.00	8	16
BBT T6 Fig. 4g	2	13	10.46	8	16
BBT T6 Fig. 4h	3	13	9.85	8	12

MHs: The number of MHs
Paths: The number of terminal paths
Average: The average length of terminal paths.
Min: The minimum length. Max: The maximum length.

V. EVALUATION

To evaluate the performance of the proposed route schemes, we investigate the number of required accesses to OFSs and the measurement accuracy depending on the number of probe packets through simulation. We use the topology in Fig. 4a, and compare six route schemes as Fig. 3c-h. We assume the threshold of high-loss link is 0.03. In each measurement, 4 high-loss links including 2 "failure links" are set at some link positions randomly. The loss rate of a high-loss (but not failure) link is set in the range of $[0.04$ - $0.06]$ randomly and that of a failure link is 0.5. Other normal links have a light loss rate in the range of $[0$ - $0.01]$. All resulting values in Fig. 6 and 7 are averaged over $10,000$ measurement instances.

Fig. 6 shows the measurement accuracy depending on the number of probe packets. The measurement accuracy is the ratio of the number of measurements in which all 4 high-loss links are correctly located to the total number of measurements $(10,000$ in our setting). We see that a route scheme with longer terminal paths needs a larger number of probe packets to operate accurately. This is because the losses at upstream links of the long path can degrade the loss measurement at downstream links as expected in Sec. 2. If a small number of probe packets arrives at the upper port of a normal link, its measured PLR can exceed the threshold even by a very small number of losses accidentally happening. On the other hand, a small number of probe packets may result in the measured PLR of a high-loss link less than the threshold.

Fig. 7 shows the number of the required accesses from OFC to OFSs until the high-loss link location process ends in case of 4 high-loss links, depending on the number of probe packets. Note that the results of the location process are not always correct in case that the number of probe packets is less than about 220 for all schemes as shown in Fig. 6. A small number of probe packets needs a more number of accesses probably due to inaccurately measured PLRs. Furthermore, a route scheme with longer terminal paths likely needs more accesses. This is because an accumulated loss rate over multiple links in a long segment will accidentally exceed the threshold even if the segment does not include any high-loss link, which leads OFC to mistakenly and unnecessarily seek high-loss links in a wrong segment as expected in Sec. 2. Hence our proposed BBT-mMH scheme with shorter terminal paths can reduce the required accesses.

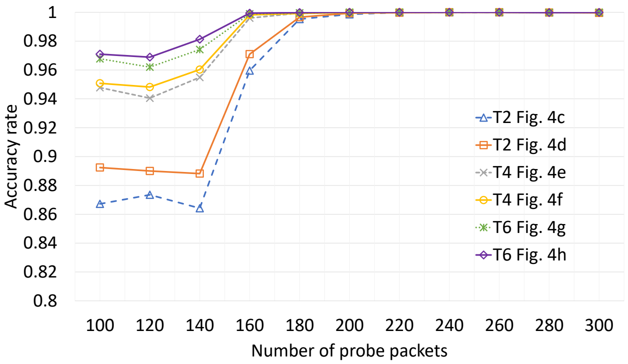


Fig. 6. The accuracy rate of the measurement

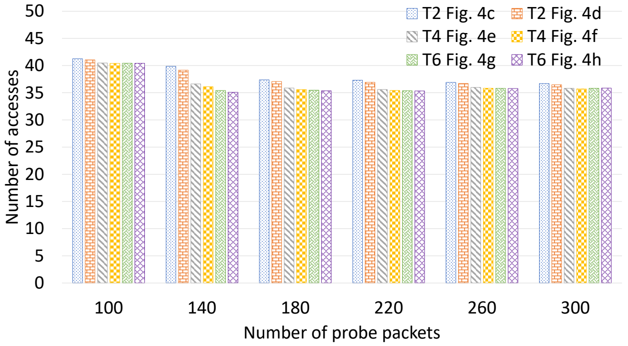


Fig. 7. The number of accesses to locate high-loss links

The comparison between single MH cases (c,d,e) and multiple MH cases (f,g,h) in Figs. 7 and 6 suggests that the use of multiple MHs with consideration on the locations and their node degrees benefits to realize an acceptably-high measurement accuracy with less number of probe packets and required accesses to OFSs. This is because the use of multiple MHs can control the length of terminal paths more flexibly compared with the use of only a single MH, although the use of multiple MHs involves additional operational costs.

VI. DISCUSSION ON BACKBONE PATHS

In the BBT route scheme (including BBT-mMH), backbone paths are constructed by using the Eulerian cycle algorithm after omitting some links if necessary to make the network consisting of even-degree nodes only. However, we also can construct backbone paths by using the Eulerian trail algorithm after omitting some links if necessary to make the network consisting of two odd-degree nodes and other even-degree nodes. An Eulerian trail starts at an odd-degree node and ends at another odd-degree node.

One possible advantage of the Eulerian cycle-based BBT scheme is a more flexibility of MH locations for starting backbone paths on a cycle than on a line due to the perfect symmetricity of cycle. For example, an MH placed at any node on a cycle can have two equal-length measurement paths covering the cycle (starting the MH node and ending the opposite node) while an MH placed only at the center node on

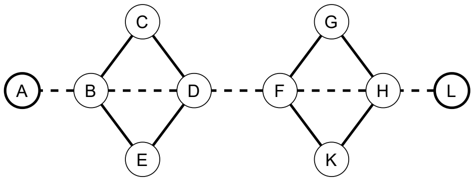


Fig. 8. An example of disadvantage in omitting links

a line can have two equal-length measurement paths. On the other hand, despite such a limitation, the Eulerian trail-based BBT scheme actually reduces the number of omitted links and can be an alternative way. In particular, for a network with two odd-degree nodes, the Eulerian cycle-based needs to omit several links among those two nodes while the Eulerian trail-based dose not omit any link. For example, in Fig. 8, with omitted links as dashed lines connecting two odd-degree nodes A and L, the remaining network will be disconnected and the process of building the measurement route becomes more complex, e.g., replacing the omitted links by a virtual combined node to construct an Eulerian cycle, and recovering the combined node into real nodes and links to complete the measurement route. Moreover, a number of tandemly adjacent omitted links may lead to a long terminal path (the path includes the omitted links). In this example, therefore, it may be reasonable to construct two backbone paths on an Eulerian trail between nodes A and L. However, in many cases, similar measurement routes can be finally constructed from the cycle-based and trail-based ones, and the performance difference is not large. Therefore, in this paper, we focus only on the Eulerian cycle-based BBT scheme. A better combination or selection of the Eulerian cycle-based and the Eulerian trail-based ones remains as future work.

VII. CONCLUDING REMARKS

On our framework of locating high-loss links in OpenFlow-based networks, we have proposed a new route scheme BBT-mMH for probing multicast packets. A benefit of the BBT-mMH to the previous BBT with a single MH, by reducing the lengths of terminal paths while keeping the number of terminal paths, was shown through simulation. However, in real networks, the possible locations of MHs may be limited and an additional cost should also be considered. How to implement multiple MHs at arbitrary locations in a cost-efficient manner remains as future work.

ACKNOWLEDGMENT

The research results have been achieved by the “Resilient Edge Cloud Designed Network (19304),” NICT, and by JSPS KAKENHI JP20K11770, Japan.

REFERENCES

[1] S. Jain, A. Kumar, S. Mandal, et al., “B4: Experience with a Globally-deployed Software Defined WAN,” in *Proc. ACM SIGCOMM'13*, 2013, pp.3–14.

[2] C. Yu, C. Lumezanu, Y. Zhang, et al., *“FlowSense: Monitoring network utilization with zero measurement cost,”* Lecture Notes in Computer Science, Springer, 2013, vol. 7799.

[3] S. R. Chowdhury, M. F. Bari, R. Ahmed, and R. Boutaba, “PayLess: A low cost network monitoring framework for Software Defined Networks,” in *Proc. 2014 IEEE NOMS*, 2014, pp. 1-9.

[4] M. Shibuya, A. Tachibana, and T. Hasegawa, “Efficient active measurement for monitoring link-by-link performance in OpenFlow networks,” *IEICE Trans. Commun.*, vol. E99B(5), pp. 1032–1040, 2016.

[5] Y. Peng, J. Yang, C. Wu, et al., “deTector: a Topology-aware monitoring system for data center networks,” in *Proc. the 2017 USENIX Annual Technical Conference*, 2017, pp.55–68.

[6] H. Li, Y. Gao, W. Dong, et al., “Taming Both Predictable and Unpredictable Link Failures for Network Tomography,” *IEEE/ACM Trans. on Networking*, vol. 26(3), pp. 1460–1473, June 2018.

[7] N. Bartolini, T. He, V. Arrigoni, et al., “On Fundamental Bounds on Failure Identifiability by Boolean Network Tomography,” *IEEE/ACM Trans. on Networking*, vol. 28(2), pp. 588–601, April 2020.

[8] N. M. Tri and M. Tsuru, “Locating deteriorated links by network-assisted multicast proving on OpenFlow networks,” in *Proc. the 24th ISCC*, 2019, pp. 701–706.

[9] T. Nakamura, M. Shibata and M. Tsuru, “On retrieval order of statistics information from OpenFlow switches to locate lossy links by network tomographic refinement,” in *Proc. INCoS*, 2019, pp. 342-351.

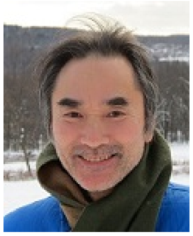
[10] S. Goto, M. Shibata and M. Tsuru, “Dynamic optimization of multicast active probing path to locate lossy links for OpenFlow networks,” in *Proc. 2020 ICOIN*, 2020, pp. 628–633.

[11] N. M. Tri, M. Shibata, and M. Tsuru, “Effective route scheme of multicast probing to locate high-loss links in OpenFlow networks,” *Journal of Information Processing*, 9 pages, to be published in Feb 2021.

Nguyen Minh Tri was born in 1987. He received his M.S. degree from University of Science, Vietnam in 2014. He is currently a Ph.D. student at Kyushu Institute of Technology, Japan. His research interest is network management and communication systems.

Masahiro Shibata was born in 1989. He received BE, ME, and DE degrees in Computer Science from Osaka University, in 2012, 2014, and 2017, respectively. Since 2017, he has been an Assistant Professor in Kyushu Institute of Technology. His research interests include network management distributed algorithms. He is a member of JPSJ and IEICE.

Masato Tsuru received B.E. and M.E. degrees from Kyoto University in 1983 and 1985, respectively, and then received D.E. degree from Kyushu Institute of Technology, Japan in 2002. He was an Associate Professor of Kyushu Institute of Technology from 2003 and has been a Professor since 2006. His research interests include performance measurement and management of computer communication networks. He is a member of the ACM, IEEE, IEICE, and IPSJ.



Akira Kawaguchi is a Professor and Head of the Computer Science Department of The City College of New York of The City University of New York. He has been working together with Kyushu Institute of Technology on a unique project to make the future internet more secure, resilient, and efficient under the support of the U.S. National Science Foundation. He received B.S. and M.S. degrees in Administration Engineering from Keio University and M.S. and Ph.D. degrees in Computer Science from Columbia University.