Distributed Active Measuring Link Bandwidth in IP Networks*

Zhiping Cai, Jianping Yin, Fang Liu, Xianghui Liu, and Shaohe Lv

School of Computer, National University of Defense Technology, China
`caizhiping_nudt@163.com, jpyin@nudt.edu.cn, fangl_nudt@163.com,`
`liuxh@tom.com, chi.shaohe@gmail.com`

Abstract. Link bandwidth is obviously critical for numerous network management tasks. Taking into account the issues of measuring costs and network-wide view for large IP network, a distributed measuring system would be an ideal monitoring architecture for active measuring link bandwidth. In this paper, we address the problem of efficiently measure assignment, which optimizing goal is to reduce the cost of measuring all links bandwidth. We show that this problem is NP-hard and propose an approximation algorithm with approximation ratio 2. The effectiveness of our measuring algorithm is validated by simulations evaluation over a wide range of network topologies.

1 Introduction

Link bandwidth is obviously critical for numerous network management tasks, including identifying and relieving congestion points, proactive and reactive resource management and traffic engineering, as well as providing and verifying QoS guarantees for end-user applications. Some novel tools and infrastructures for measuring network bandwidth have been developed and proposed by researchers and industries, like as SNMP and RMON measurement probes [1], Cisco's NetFlow tools [2], the IDMaps [3], [4], packet-pair algorithms for measuring link bandwidth [5], [6] and the Pathchar [7] tool for estimating Internet link characteristics.

These measurement tools periodically query and collect detailed traffic data on packet flows for monitoring and measuring network flows and bandwidth usage. Unfortunately, probes processing queries can adversely impact routers performance and active probe message transfers can result in significant volumes of additional network traffic [8].

As an example, Pathchar [7] is unique in its ability to measure the bandwidth of every link on a path accurately while requiring special software on only one host. This mean it could easily be widely deployed. Although excellent as a testing tool, the problem with Pathchar is that it is slow and can consume significant amounts of network bandwidth [6]. The distance between measuring

---

[*] This work is supported by the National Natural Science Foundation of China under Grant No. 60373023.

H. Jin, D. Reed, and W. Jiang (Eds.): NPC 2005, LNCS 3779, pp. 295–302, 2005.

station and destination link is greater, the probe messages would consume more bandwidth. So using a distributed active measuring architecture can reduce the overall measuring cost and the impact on network bandwidth.

Another key consideration in using the distributed measuring architecture is these measurement tools can only measure links included in the routing trees of the measuring stations [9]. To measure all link of the network, it must install some distributed measuring stations. Thus, taking into account the issues of measuring costs and network-wide view for large service provider network, a distributed measuring system would be an ideal monitoring architecture for active measuring link bandwidth.

The costs of measuring the same link bandwidth are different in sending probe messages from different measuring stations. Hence, once we have selected a set of measuring stations, we need to determine the measurement strategies to minimize the measuring cost for measuring all link of a network. Our work focuses on optimizing the measure assignment for measuring all links bandwidths.

The main contributions of our work are as follows: We first show that the problem of minimizing the measuring costs in a given network with some given measuring stations is NP-hard. Then using greedy heuristics and dynamic programming, we propose an approximation algorithm with an approximation ratio 2. The effectiveness of our measuring algorithm is validated by simulations evaluation over a wide range of network topologies.

The paper is structured as follows. We bring forward the measure assignment problem and provide the integer programming formulation in the section 2. In next section, we give an approximation algorithm to solve the measure assignment problem. The result of simulations evaluation is shown in section 4. And we depict our further research in the last section.

1.1 Related Work

Y. Bejerano et al. [8] and J. Walz et al. [10] study link monitoring and delays in IP networks based on a single Network Operations Center(NOC) . In order to monitor links not in its routing tree, the NOC uses the IP source routing option to explicitly route probe packets along the links [8]. Unfortunately, due to security problems, many routers frequently disable the IP source routing option. Consequently, approaches that rely on explicitly routed probe packets for delay and fault monitoring may not be feasible in today's ISP and Enterprise environments [11]. On the other hand, the distributed monitoring infrastructures would be better than a single point-of-control due to reducing the measurement cost.

There is recently significant interest in developing network monitoring infrastructures that allow ISPs to monitor their network links [12]. A key consideration in the design of monitoring infrastructures is to develop low-cost solution. In particular, the idea of placing and operating monitors at all nodes in a network is not cost-efficient. Instead, there has been significant recent in replying on tomographic techniques that use only a few probing nodes (beacons) for monitoring the health of all network links [12], [13], [14], [15],[16].

Once the beacons are located, the smallest set of probes must still be determined. Our work focuses on determining and optimizing the probes assignment for measuring the link bandwidths. The majority of work on network tomography on either topology discovery [4], [17] or link delay monitoring [11]. Some recent research showed that active measurements can also be used to pinpoint failure in IP networks [11], [18]. For measuring link bandwidths, the measuring model and cost are different from that of topology discovery or link delay monitoring. We develop different strategies and algorithms based on different bandwidth measuring technology.

2 Bandwidth Measure Assignment Problem

2.1 Problem Formulation

A number of tools estimate network link bandwidth using Variable Packet Size (VPS) probing technology, like as Pathchar [7], Clink [19] and Pchar [20]. The key element of the technique is to measure the RTT from the measuring station to each hop of the path as a function of the probing packet size [9]. VPS uses the Time-To-Live field of the IP header to force probing packets to expire at a particular hop. The router at that hop discards the probing packets, returning ICMP "Time-exceeded" error messages back to the measuring station. The measuring station uses the received ICMP packets to measure the RTT to that hop.

We model the Service Provider or Enterprise IP network by an undirected graph $G(V, E)$, where the graph nodes V, denote the network routers and the edges, E, represent the communications links connecting them. For measuring the bandwidth of a link $e \in E$, a measuring station s must be selected firstly for sending probe message, where $s \in V$ such that e belongs to s's routing tree (i.e., $e \in T_s$). Consequently,the measuring station s must send two probe messages to the end-points of e, which travel almost identical routes except for the link e.

Once having selected a set S of monitoring stations, a measuring system designated for measuring the bandwidths of all network links has to find a measure assignment $M \subseteq \{m(s,u) | s \in S, u \in V\}$, where each message $m(s,u)$ represents a probe message that is sent from the measuring station s to node u.The measure assignment M are required to satisfy a covering assignment constraint which ensures that for every edge $e = (u,v) \in E$, there is a measuring station $s \in S$ such that $e \in T_s$ and M contains the messages $m(s,u)$ and $m(s,v)$. The covering assignment constraint essentially ensures that every link is measured by some stations. Note that although we only consider the problem of measuring all network links in this paper, our results also apply to the problem of measuring only a subset of links of interest.

We associate a positive cost $c_{s,t}$ with sending a probe message along the path $P_{s,t}$ between any pair of nodes $s, t \in V$. For every intermediate node $x \in P_{s,t}$ both $c_{s,x}$ and $c_{x,t}$ are at most $c_{s,t}$ and $c_{s,x} + c_{x,t} \geq c_{s,t}$. Typical example of this cost model are the fixed cost model, where all messages have the same cost, and the hop count model, where the message cost is the number of hops in its route. Moreover, we denote by $h_{s,t}$ the number of hops in path $P_{s,t}$.

Definition 1 (Measure Assignment). *Given an undirected graph* $G = (V, E)$*, where* V *denotes the set of nodes,* E *represents the edges between two nodes. Let* T_v *be a route tree for every node* $v \in V$*. And* $S \subseteq V$ *denotes a set of measuring stations. We say* $M \subseteq \{m(s,u) | s \in S, u \in V\}$ *is a Measure Assignment, if there is a measuring station* $s \in S$ *for every edge* $e = \{u, v\} \in E$ *such that* $e \in T_s$ *and* $m(s,u) \in M$*,*$m(s,v) \in M$*. The cost of a Measure Assignment* M *is* $COST_M = \sum_{m(s,u) \in M} c_{s,u}$*.*

To reduce the network burden, the measuring cost is preferable to as few as possible. We are interested in the following optimization problem.

Definition 2 (Measure Assignment Problem-MA). *Given an undirected graph* $G = (V, E)$ *and a routing tree* T_v*, for every node* $v \in V$*. Let* $S \subseteq V$ *denotes a set of measuring stations. The Measure Assignment problem is to determine the measure assignment with the minimum cost.*

2.2 Integer Programming Formulation for the MA Problem

Given an undirected graph $G = (V, E)$, where V denotes the set of nodes, E represents the edges between two nodes. Let T_v be a route tree for every node $v \in V$. And $S \subseteq V$ denotes a set of measuring stations. Let M denotes a Measure Assignment. The binary variable $x_{s,u}$ indicates whether there is a probe message from s to u in M. And the binary variable $y_{s,u,v}$ indicates whether the routing tree of s includes edge (u, v). We give the integer programming formulation of the measure assignment problem as follows.

$$\text{Min} \sum_{m(s,u) \in M} c_{s,u}$$

Subject to:

$$\sum_{s \in S} x_{s,u} x_{s,v} y_{s,u,v} \geq 1, \qquad \text{for each } \{u, v\} \in E \tag{1}$$

$$x_{s,u} \in \{0, 1\}, \qquad \text{for each } u \in V, s \in S \tag{2}$$

$$y_{s,u,v} \in \{0, 1\}, \qquad \text{for each } \{u, v\} \in E, s \in S \tag{3}$$

The first constraint makes sure that each edge can be measured from at least one measuring station which routing tree constains this edge.

2.3 Hardness of the MA Problem

The MA problem could be proved to be NP-hard by presenting a polynomial reduction from the well-known Vertex Cover problem [21] to the MA problem. The details of this proof are omitted due to space limitations.

Theorem 1. *Given a set of measuring stations* S*, the MA problem is NP-hard.*

3 Approximation Algorithm for Measure Assignment Problem

We give a 2-approximation algorithm for the measure assignment problem using dynamic programming strategy.

3.1 An Approximation Algorithm

For measuring the bandwidth of any edge $e \in E$, at least one station $s \in S$ must send two probe message, one to each end point of e. So the measuring cost is the sum of two probe message cost, i.e. $c_{s,u} + c_{s,v}$. Note that $c_{s,u}$ is zero while $s = u$. While the probe assignment M has contained one probe message $m_{s,u}$ or $m_{s,v}$, the measuring cost would be $c_{s,v}$ or $c_{s,v}$ respectively. We pick a station to minimize the measuring cost for every edge by using dynamic programming strategy. And the approximation algorithm is given as follows.
Algorithm $(G = (V, E), S \subseteq V, \{c_{s,u} | s \in S, u \in V\})$:

```
1. $M = \Phi$;
2. $E' = E$;
3. for each edge $(u, v) \in E$;
   (a) $Cost(u, v) = min_{(u,v)\in T_s}(c_{s,u} + c_{s,v})$;
4. while $(|E'| \neq 0)$
   (a) Pick $min_{(u^*,v^*)\in E'} Cost(u^*, v^*)$
   (b) $E' = E' - (u^*, v^*)$
   (c) $M = M \cup \{m(s^*, u^*), m(s^*, v^*)\}$
   (d) for each edge $(u', v') \in E' \cap T_{s^*}$,s.t. $u' \in \{u^*, v^*\}$ or $u' \in \{u^*, v^*\}$.
       Suppose that $u' \in \{u^*, v^*\}$.
       i. $Cost(u', v') = min_{(u,v)\in T_s}\{(c_{s,u'} + c_{s,v'}), c_{s^*,v'}\}$;
```

It is not hard to see that this algorithm is effectively equivalent to the following: start with $M = \Phi$ and $E' = E$. Compute the measuring cost for each edge. The measuring cost of each edge is the sum of two probe message cost, i.e. $c_{s^*,u} + c_{s^*,v}$, while s^* is the station which minimize the measuring cost. Pick one edge (u^*, v^*) from E' that achieves the minimum the measuring cost. Let it be measured by station s^*. Add these two probe messages $m(s^*, u^*)$,$m(s^*, v*)$ to the Measure Assignment M. And remove the edge u^*, v^* from E'. Adjust the measuring cost of these edges which incident with the picked edge. Repeat until M cover all links.

Note that there exists a implementation of this algorithm takes $O(|E|^2)$ time. Then we prove the algorithm is a 2-approximation algorithm.

Theorem 2. *The approximation ratio of the approximation algorithm is* 2.

Proof. In any measuring assignment, at least one probe message can be associated with each edge e. Let it be the message that is sent to the farthest endpoint of e from the measuring station. Let M'' be the optimal probe assignment and let

s''_e be the station that measures edge e in M''. So, in M'', the cost of measuring edge $e = (u, v)$ is at least $max\{c_{s''_e,u}, c_{s''_e,v}\}$. Let s^* be the selected station for measuring edge e in the assignment M returned by the approximation measure assignment algorithm. So we have the following inequality:

$$Cost(u, v) \le c_{s*_e,u} + c_{s*_e,v} \le c_{s''_e,u} + c_{s''_e,v} \le 2max\{c_{s''_e,u}, c_{s''_e,v}\}. \quad (4)$$

Thus,we have $COST_M \le 2COST_{M''}$. □

4 Simulations

In this section, we present simulation results of comparing the performance of the various algorithms that solve the measure assignment problem. The main objective of the simulations is to demonstrate that our proposed algorithmic solutions are not only theoretically sound but also they could give significant benefits over naive solutions in practice for a wide variety of realistic network topologies. The simulations are based on network topologies generated using the Waxman Model [22], which is a popular topology model for networking research. Different network topologies are generated by varying three parameters: (1)n, the number of nodes in the network graph; (2)α, a parameter that controls the density of short edges in the networks; and (3)β, a parameter that controls the average node degree.

We compare the performance of three algorithms: the naive random assignment algorithm, the simple probe assignment algorithm [16] and our approximation algorithm. The comparison is in terms of the total measuring cost. We denote the cost of measuring all link for these algorithms by $COST_r$, $COST_s$ and $COST_a$ respectively.

Table 1 presents one set of simulation results.We have obtained similar results for other parameter settings. The third and fourth columns in the table represent the maximum and average degree of the nodes in the generated network graph respectively. Our results indicate that using our approximation algorithm can reduce measuring cost. And the result of our algorithm is better than is better than the other two algorithms.

We have obtained the other simulation result by adjusting the number of measuring stations. We compute the measuring cost on having selected 50,75,100,125, 150 measuring stations respectively. From figure 1, we can know that the measuring cost would be reduced by adding the number of measuring stations. And our algorithm is better than the naive assignment.

Table 1. Comparisons of Measuring Algorithms on Different Topologies

n	α	β	Maximum Degree	Average Degree	$COST_r$	$COST_s$	$COST_a$
400	0.1	0.06	8	2.45	723	616	503
400	0.5	0.02	11	3.43	901	844	711
400	0.5	0.06	26	4.94	1988	1773	1475

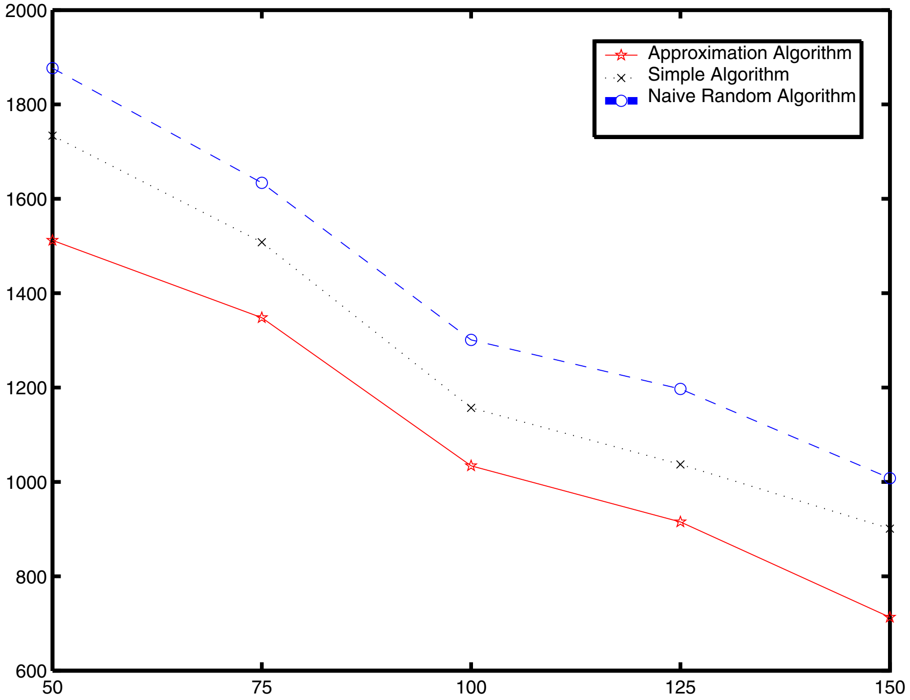

Fig. 1. Measuring Cost on Different Algorithms

5 Conclusion

In this paper, we have addressed the problem of efficiently measure assignment in IP networks. This problem is shown NP-hard. We have proposed an approximation algorithm with approximation ratio 2 to solve the measure assignment problem. Finally, we have verified the effectiveness of our approximation algorithms through simulations evaluation. This work is helpful to efficiently measure link bandwidth in IP networks.

Further research would be conducted to develop novel algorithms based on different measuring technology.

References

1. W. Stallings: SNMP, SNMPv2, SNMPv3, and RMON 1 and 2. Addison-Wesley Longman, Inc., 1999.
2. Cisco Systems: NetFlow Services and Applications, White Paper, 1999
3. P. Francis, S. Jamin, V. Paxson, L. Zhang, D.F.Gryniewicz, and Y. Jin: An Architecture for a Global Internet Host Distance Estimation Service. In Proc. IEEE INFOCOM 1999.

4. S. Jamin, C. Jin, Y. Jin, Y. Raz, Y. Shavitt, and L. Zhang: On the Placement of Internet Instrumentation. In Proc. IEEE INFOCOM 2000.
5. J.C.Bolot: End-to-End Packet Delay and Loss Behavior in the Internet. In. Proc. ACM SIGCOMM 1993.
6. K. Lai and M. Baker: Measuring Bandwidth. In Proc. IEEE INFOCOM 1999.
7. V.Jacobsen: Dynamic Distance Maps of the Internet Paths, ftp://ftp.ee.lbl.gov/pathchar, April 1997.
8. Breitbart Y., Chan CY., Garofalakis M., Rastogi R., Siberschatz A.: Efficiently Monitoring Bandwidth and Latency in IP Networks. In Proc. IEEE INFOCOM 2001.
9. R.S. Prasad, M. Murray, C. Dovrolis, K. Claffy: Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. IEEE Network, 2003, 17(6):27-35.
10. J. Walz, B. Levine: A Hierachical Multicast Monitoring Scheme. In Proc. Networked Group Communication 2000.
11. Y. Bejerano, R.Rastogi: Robust Monitoring of Link Delays and Faults in Networks. In Proc. IEEE INFOCOM'03, 2003.
12. R. Kumar, J. Kaur: Efficient Beacon Placement for Network Tomography. In Proc. ACM Internet Measurement Conference 2004, October 2004.
13. K. Claffy, T.E. Monk, D. McRobb: Internet Tomography, Nature, January, 1999.
14. J. D. Horton, A. Lopez-Ortiz: On the Number of Distributed Measurement Points for Network Tomography. In Proc. ACM SIGCOMM IMC'03. 2003,pp.204-209.
15. Kyoungwon Suh, Yang Guo, Jim Kurose, and Don Towsley. Locating Network Monitors: Complexity, Heuristics, and Coverage. In Proc. IEEE INFOCOM 2005.
16. M. Adler, T. Bu, R.K. Sitaraman, D. Towsley: Tree Layout for Internal Network Characterizations in Multicast Network. In Proc. Networked Group Comm,2001.
17. N. Spring, R. Mahajan, D. Wetherall: Measuring ISP Topologies with Rocketfuel. In Proc. ACM SIGCOMM'02. 2002.
18. H. X. Nguyen, P. Thiran: Active Measurment for Multiple Link Failure Diagnosis in IP Networks. In Proc. PAM'04. April, 2004.
19. A. B. Doweny: Using Pathchar to Estimate Internet Link Characteristics. In Proc. ACM SIGCOMM'99. Sept,1999.
20. CAIDA: http://www.caida.org/tools/. Oct, 2002.
21. Dorit S. Hochbaum: Approximation Algorithm for *NP*-Hard Problems. PWS Publishing Company,1997.
22. B.M.Waxman: Routing of Multipoint Connections. IEEE Journal on Selected Areas in Communications, 1988, 6(9):1617-1622.
23. Zhiping Cai, Jianping Yin, Fang Liu and Xianghui Liu: Distributed Monitoring Model With Bounded Delay For Evolving Networks. *Journal of Software*, 2005,11.
24. Zhiping Cai, Wentao Zhao, Jianping Yin and Xianghui Liu: Using Passive Measuring to Calibrate Active Measuring Latency. In Proc. ICOIN2005, *Lecture Notes in Computer Science 3391*, C.Kim(eds.), Springer-Verlag, 2005.
25. Zhiping Cai, Jianping Yin, Fang Liu, Xianghui Liu, Shaohe Lv: Efficiently Monitoring Link Bandwidth in IP Networks. In Proc. IEEE GLOBECOM 2005.