パッシブ測定に基づく経路別エンドエンド品質測定法

長谷川 亨　　大岸 智彦　　阿野 茂浩

(株)KDDI研究所

IP ネットワークの通信インフラとしての役割の高まりとともに，エンドエンド通信に対して高品質なサービスを提供することが期待されている．この実現には，エンドエンドの経路の品質を常に監視し，品質劣化とその原因を検出する手法が必要である．これに対して，筆者らは，TCP 通信を対象として，ネットワークを流れるパケットをパッシブに観測することにより，経路毎のエンドエンドの TCP 品質を測定する手法を考案した．以下では，提案手法の詳細と，商用 IP ネットワークにおける品質測定実験結果に基づく評価について述べる．

Path-based Passive End-to-End Performance Measurement Method

Toru Hasegawa, Tomohiko Ogishi and Shigeiro Ano

KDDI R&D Laboratories

As the Internet has become an infrastructure for the global communication, the traffic engineering of ISPs (Internet Service Providers) becomes important to constantly provide high quality service. The recent studies make it clear that collecting performances of individual paths (a set of links between two end points) is inevitable to the traffic engineering. If an ISP detected a low performance path, it could add either a bandwidth or a new link to the bottleneck link of the path. In the literature, active and passive measurement methods were proposed to characterize individual path performances. However, all methods require many measurement tools to be set at various links in the Internet. The larger the number of measured paths becomes, the larger the number of tools becomes. In order to solve the scalability problem, we have explored a path-based measurement method using a single passive measurement tool. The method extracts TCP performances of individual paths in the following way: The measurement tool captures traffic aggregates on a backbone link, and extracts TCP performances of individual flows. Then, it collects them into a path performance using the IP address information of an ISP network. We have also performed the measurement experiment at an ISP network. As the result, the proposed method is useful to characterize individual path performances of an ISP network.

1. Introduction

As the Internet has become an infrastructure for the global communication, the performance degradation has become a serious problem. For ISPs (Internet Service Providers), the traffic engineering becomes important to constantly provide high quality service. The recent studies [1,2] make it clear that collecting either performances or traffic demands of individual paths (a set of links between two points) is inevitable to the traffic engineering. If an ISP detected a low performance path, it could add either a bandwidth or a new link to the bottleneck link of the path.

There are two approaches to measure individual path performances: active measurement and passive measurement. By injecting test packets, active measurements obtain responses to the test packets. The responses are used to extract performances such as packet loss, packet delay and packet delay jitter [3]. Many active measurement experiments to find traffic characteristics of Internet paths were performed. Although the experiments were successful, they are difficult to apply the ISP traffic engineering due to the following two drawbacks. First, active measurements have the potential problem to add significant test traffic load. Second, since the traffic engineering needs to collect performances or traffic demands for many paths, ISPs must set many active measurement tools over an ISP network.

In order to avoid the first drawback, passive measurements are hopeful because of no test traffic load. A few passive measurement methods [1, 2] were proposed to collect a matrix of traffic demands between many points over an IP network such as an ISP network and an Internet backbone. The methods consist of the two functions: flow measurement and path characteristics extraction. First, passive measurement tools, e.g., NetFlow [4] of routers and traffic monitors [5], are set at many links that connect the IP network and external networks so that the incoming traffic and outgoing traffic are measured. At these links, traffic amounts of individual flows, which are specified by source and destination IP addresses, are collected. Second, a matrix of traffic demands between the measured links is created using both the collected flow information and the IP network numbers of the external networks.

Although the above passive measurements collect individual path performances, they require a large measurement infrastructure that consists of many passive measurement tools. This reduces the scalability of the methods, and as the result, it is not easy to use the method in a daily traffic engineering of ISP. However, if the necessary measurement points were few, the passive measurement could be used for the daily ISP traffic engineering.

In order to solve the lack of scalability, we propose a path-based passive measurement method where a passive measurement tool is set at a link where many flows are aggregated. In order to evaluate the method, we have applied the passive measurement tool that extracts end-to-end TCP performances [6] to the link where many mail flows are aggregated in an ISP network. TCP packet retransmission rates are extracted for individual paths to dial up access points and they are analyzed from many viewpoints.

In this paper, we describe an empirical study on applying a path-based passive measurement method to an ISP network. The rest of paper is structured as follows. In section 2 and section 3, we describe the methodology of the proposed method and the implementation, respectively. In section 4, we present the experiences on applying the implemented tool to measure path performances of a commercial ISP network. In section 5, we evaluate the proposed method.

2. Measurement Methodology

The design principles of the proposed method are as follows.

(1) Passive Measurement of Traffic Aggregates

The first design criterion is to make measurement points of links as few as possible. In this paper, we consider a single measurement point. If many flows are aggregated, we can collect performances of paths between many pairs of points. Figure 1 shows the methodology and motivation. In an ISP network, traffic flows are aggregated on links to a backbone network. An example is a link that connects an ISP backbone and a center where mail and DNS servers are located. Since all dial up users access the mail servers, we can see all flows transferred through dial up access points. This makes it possible to analyze all path behaviors where the concentration links tend to become bottlenecks.

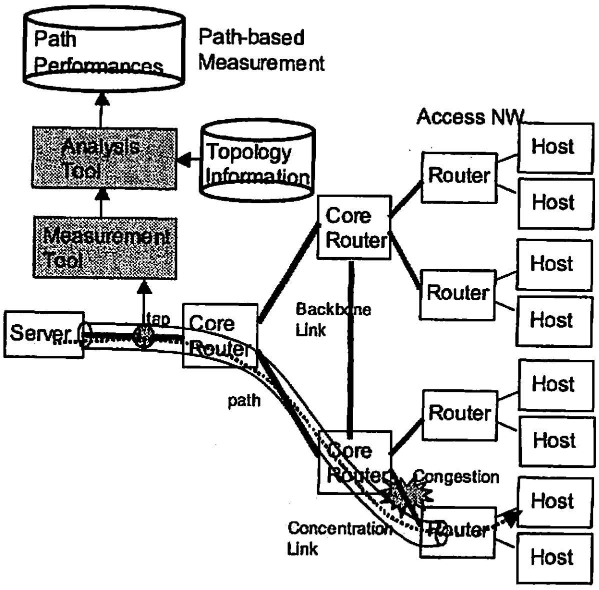

Figure 1 Network Configuration for Measurement Method

(2) Path-based Performance Extraction Using Topology Information

We define a path as a pair of source and destination IP network numbers, where an IP network number is specified by an IP address and a net mask. The measurement and analysis tools are used as shown in Fig.1. The measurement tool extracts performances for each flow which is determined by source and destination IP addresses. After the extraction, the analysis tool collects the performances on the same path using the topology information. An example of collection is as follows: The topology information is a table of IP network numbers of dial up access points. When a link between a mail center and a backbone is measured, we can extract performances of all paths between a mail center and all access points. If two hosts are accommodated at an access point, we can extract an average performance for the two hosts.

(3) TCP Performance Extraction

IP performances such as packet loss and delay [3] are difficult to extract because the extraction requires two measurement points. Instead of them, TCP performance parameters such as TCP throughput and TCP packet retransmission rate are used. Since TCP traffic is currently dominant, we consider that TCP performances are feasible to characterize path performances. The measurement tool extracts TCP performances of each flow analyzing captured TCP packet traces according to the TCP state machine.

3. Implementation

3.1 Overview

The method consists of the measurement tool and the analysis tool, which we call Internet Performance Monitor and Internet Performance Analyzer, respectively. The structure is illustrated in Fig. 2. In the rest of paper, we call them just as the monitor and analyzer, respectively.

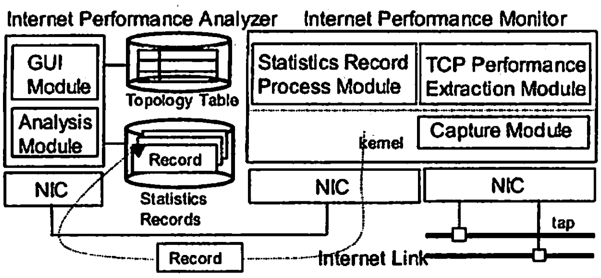
Internet Performance Analyzer
Internet Performance Monitor
GUI Module
Topology Table
Analysis Module
Record
Statistics Records
NIC
Statistics Record Process Module
TCP Performance Extraction Module
Capture Module
kernel
NIC
NIC
tap
Record
Internet Link

Figure 2 Structure of Monitor and Analyzer

The monitor and analyzer operate simultaneously in order to extract TCP performance for individual paths. The monitor extracts TCP performances without storing captured packet headers into a disk. It does not record a performance parameter for each flow (each TCP connection), but record a statistics record for a fixed interval which contains those for flows of the interval. This reduces the size of disk space.

A path is specified by source and destination IP network numbers; therefore, the topology table is a table of a pair of source and destination IP network numbers. After receiving statistics records from the monitor in a fixed interval, the analyzer sums up the statistics records whose source and destination IP addresses are in the range of the source and destination IP network numbers of an individual path.

3.2 Internet Performance Monitor

3.2.1 Overview

The monitor (Internet Performance Monitor) is an extension to the prototype whose implementation details are described in [6]; therefore, we briefly describe the implementation. The monitor has been implemented as software running on PC unix. The structure of the monitor, as illustrated in Fig. 2, consists of the three modules: the capture module, the TCP performance extraction module and statistics record process module. The capture module, which runs in the kernel, captures packets from a tapped link, collects headers and sends the headers to the TCP performance extraction module. The TCP performance extraction module extracts performance parameters for an individual TCP connection and creates a record in the main memory of the PC. The statistics record module sums up the records for the same source and destination IP addresses, and sends the statistics records to the analyzer in a fixed interval.

3.2.2 TCP Performance Parameter Extraction

Using the state transition table, the monitor emulates TCP behavior in response to TCP packets at the tapped point. It extracts a TCP packet retransmission rate and a TCP throughput as performance parameters. TCP packet retransmission rate calculation example is shown in Fig. 3. Every when a new TCP packet is captured, its sequence number is compared with the maximum sequence number which the monitor sees. When the sequence number of packet is less than the maximum sequence number, the DATA packet is regarded as a retransmission. In addition, TCP throughput is calculated in the following expression: throughput = ((sequence number of the last DATA packet + the size of the last DATA packet) – sequence number of the first DATA packet) / the elapsed time from the first DATA packet to the last one.

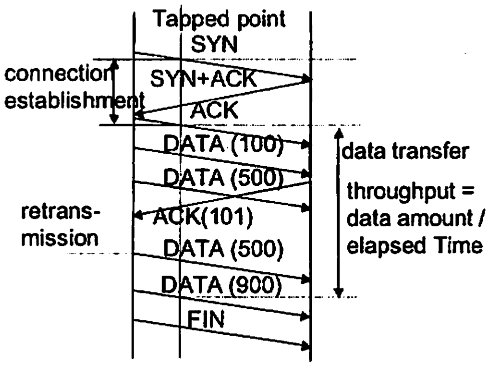

Figure 3 TCP Throughput and Retransmission Rate Calculation

3.3 Internet Performance Analyzer

(1) Statistics Record

Statistics records are created in a fixed interval, e.g., every 10 minutes, for all the pairs of source and destination IP network numbers. If some TCP connections happen between the same pair in the interval, all extracted parameters are summed up to the record. The statistics records of an interval are illustrated in Fig. 4, and a statistics record contains the following information:

- Connection Number: total number of detected TCP connections
- DATA Packet Number: total number of DATA packets which include both successfully transferred packets and retransmitted packets
- Retransmitted Packet Number: total number of

retransmitted DATA packets

- Data Transfer Duration: total duration of data transfer phases, and so on.

(2) Path-based analysis

The topology information is a table of a pair of source and destination IP network numbers. The table is written by users and is stored as a file. The analyzer reads both the topology table and the statistics records, and calculates the TCP throughput, TCP packet retransmission rate and so on for all the pairs of source and destination IP network numbers. The analyzer generates time series data of them in the user specified time unit, e.g., 10 minutes, 30 minutes, hour, day, week and so on, as shown in Fig. 4. The analyzer generates them both in the text format and in the graphical format.

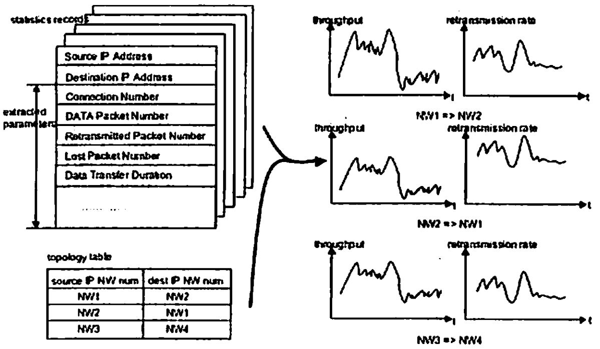

Figure 4 Statistics Records, Topology Table and Analysis Results

4. Measurement Experiments

We performed the measurement experiments at a commercial ISP network in order to answer the following questions.

- Can TCP performances characterize a path in an ISP network?
- Can the proposed method extract individual path performances?

4.1 Measurement

The measurement experiments were performed at a commercial ISP network. The monitor (Internet Performance Monitors) is connected to a mirror port of a router that connects the mail center and the backbone network. The monitor captures all packets between the mail servers and dial up users of all access points. The number of total access points is 141. The access points accommodate hosts using the following access links: analogue telephone links and narrow-band ISDN links.

The topology table is based on access points of dial up users. TCP throughputs, TCP retransmission rates and TCP packet pseudo-loss rates are extracted for down links from the mail center to the access points.

The interval of statistics record creation is 10 minutes.

In the rest of this section, we call TCP throughput, TCP packet retransmission rate and TCP packet pseudo-loss rate just as throughput and retransmission rate, respectively.

4.2 TCP Performances as IP Network Performance Metric

In order to answer the first question, we analyze the statistics records collected on 6 days of February 2001. The total mail traffic of the days is about 4 G bytes.

(1) TCP Throughput and TCP Retransmission Rate

We compare the average throughputs and the retransmission rates over the days for several access points. However, we do not see a strong correlation between them. This may be due to the access link bandwidth variation of an access point. Since an access point provides several kinds of access links such as analogue telephone links and ISDN links, the throughput not only depends on the path performance, but also on the bandwidth of access links. For example, the link bandwidths are 28.8 Kbps, 32 Kbps, 56 K bps and 64 Kbps. We conclude that the throughput cannot characterize a path performance to a dial up user access point.

(2) Retransmission Rate and Loss Rate

Since TCP retransmits packets that are not lost by mistake, the TCP packet retransmission rate is greater than the IP packet loss rate, as pointed out by [7,8]. Therefore, we check how many TCP packets are retransmitted by mistake. Table 1 shows the result. The ratio of the lost packets to the total retransmitted packets is shown for each day. We consider that a TCP packet is lost when three duplicate ACK packets are observed.

Table 1 Ratio of Lost Packets to Retransmitted Packets

Day	1st day	2nd day	3rd day	4th day	5th day	6th day	total
Ratio (%)	59.3	32.2	25.3	28.2	41.6	39.3	38.2

The similar evaluation was performed in 1994 and 1995 [7]. The ratios are calculated for the two large packet traces that contain 2,800 and 18,000 TCP communication traces in the Internet [7]. The ratios for the two traces are 44 % and 17 %.

The comparison between the two experiments is difficult due to the different conditions. However, we observe that the above ratios of the two experiments are similar. We consider that the retransmission rate is an appropriate metric for characterizing path.

4.3 Retransmission Rates of Individual Access Points

4.3.1 Overview

In order to answer the second question, we measured the mail traffic on another day (one day in August 2001). Since as of February, we did not know the IP network numbers of access points, we did the experiment again. The collected statistics records are analyzed in order to answer the following questions:

- Can the proposed method measure (see) enough mail traffic to analyze performances of the paths to 141 access points? If the mail traffic were not transferred on the path constantly, it would be difficult to extract performances of the path.
- Can TCP packet retransmission rates characterize path performances to access points?

The mail traffic of the day is about 2 G bytes, and the mail number of the day is more than 300,000. From the collected statistics records, we create the two time series data of mail traffic amounts and retransmission rates for each access point. The retransmission rate is calculated at each 10 minute interval, and it is the average retransmission rate of TCP connections that exist during the 10 minute interval. In this paper, we do not analyze retransmission rates of individual TCP connections.

We observe that access points are classified in the three groups according to the mail traffic amounts. Each group roughly corresponds to the size of city where an access point is located. We define the three groups, i.e., large, medium and small access points, according to the mail traffic amounts of individual 10 minute intervals.

- Large Access Point: The intervals whose traffic amounts are more than 50 K bytes are more than 50 %, and the intervals whose traffic amounts are not zero are more than 80%.
- Medium Access Point: The intervals whose traffic amounts are more than 50 K bytes are more than 20 %, and the intervals whose traffic amounts are not zero are more than 80 %.
- Small Access Point: The other access points other the above two types of access points.

Since a mail traffic of a 10 minute interval need be large in order to calculate the retransmission rate correctly, we adopt 50 K bytes as the threshold. If the mail traffic is more than 50 K bytes, we consider that the calculated retransmission rate of the interval is reliable.

4.3.2 Performance Characteristics of Small Access Points

The time series data of traffic amounts and retransmission rates of all the small access points are almost the same. The mail traffic is observed during the half of the day, but mail traffic amounts of most 10 minute intervals are less than 50 K bytes. However, among 7 intervals, the mail traffic amounts of 5 intervals are less than 1 K bytes. In other words, the retransmission rates of the 5 intervals are not reliable; therefore, we remove the intervals whose mail traffic amounts are less than 50 K bytes. As a result, the retransmission rates of only two 10 minute intervals are larger than 3 %. As for the other small access points, we see almost the same situation. From this observation, we conclude that there is no performance problem at the small access points.

4.3.3 Performance Characteristics of Medium and Large Access Points

(1) Quiescent and Busy Periods

After seeing the time series data of medium and large access points, we observe that there are two kinds of periods: quiescent and busy periods. This observation is very similar to the results of TCP retransmission rate measurement over the Internet [7]. In this section, we define the two periods in the following way. The definitions are a little bit different from those of [7]. The quiescent interval is the 10 minute interval where the retransmission rate is 0 % or where the mail traffic is less than 50 K bytes. We assume that if a mail traffic were less than 50 K bytes, the load of path would be very light, and the retransmission rate would be 0 % as the result. On the contrary, the busy interval is a 10 minute interval where the retransmission rate is more than 0%. The ratio of total quiescent intervals to the total intervals is about 87.7 %. From the above observation, we consider that a path performance is determined by a retransmission rate of busy intervals.

From the above discussion, we consider that a path performance is determined by busy periods. The difference of the four kinds of access points is the number of busy intervals, and the average retransmission rate of the busy intervals.

(2) Path Performance and Location

If the backbone link were heavily loaded, the retransmission rates of the access points that are accommodated by the same backbone link would be large. In order to know the correlation between path performances and locations, we compare the busy intervals of access points which are geographically close. We pick up all the intervals whose retransmission rates are more than 10 %, and also pick up the intervals of the same time for the access points that are located in the same prefecture. (Since we do not know the precise topology of the ISP network, we assume that the access points of the same prefecture is accommodated by the same backbone link.) We calculate the ratio of the intervals whose retransmission rates are more than 10 % to the total intervals of the same time and the same prefecture. What the ratio is high means that when a retransmission rate of an access point is high, the retransmission rates of the close access points are also high. However, the ratio is just 12.5 %. From the result, we consider that the performances of individual access points are independent, and the path performance does not depend on the backbone link performance. In other words, the path performance to the access point characterizes the behavior of concentration link between the backbone and the access point.

5. Discussion

(1) Applicability to Intra-domain Traffic Engineering

The proposed passive measurement method is considered be able to characterize performances of paths if the traffic is constantly transferred. As long as an ISP network is a target, the mail traffic is a good candidate to detect performance problems that are caused by the bandwidth shortages. Since intra-domain routing is well controlled for the intra-domain traffic, e.g., the mail traffic of dial up users, in ISP networks, bandwidth shortages of links are one of main causes of performance degradations. The results of section 4 implicate that concentration links tend to run short of bandwidths. The TCP packet retransmission rate is a good parameter to detect path performance degradation, which helps an ISP operator to detect a bottleneck link.

(2) Comparison with Related Work

TCP performances of end-to-end Internet paths were studied first in 1994 and 1995 [7]. TCP bulk transfers are performed between many measurement points where demon programs called as NPD are installed. This study is epochal, and it finds many characteristics of end-to-end Internet performances. This study also makes it cleat that TCP performances are useful to analyze individual path performances. Our work is motivated by this study; however, our work is different from the following viewpoints: First, our work is based on passive measurement although this study is based on active measurement. Although this study needs to set many active measurement tools, our method just uses one passive measurement tool. Second, our method is applied to an ISP network where the routing and provisioning are controlled in a unified manner. We find that paths of ISP networks have similar characteristics to those of Internet paths.

6. Conclusion

We have proposed a path-based passive TCP performance measurement method, and implemented the measurement tool that achieves the method. Our work is different from the previous studies of path performance characterization in the number of measurement points. Our method requires just a single measurement point on which many traffic flows are aggregated. In the proposed method, the tool taps the link where many traffic flows are aggregated, and extracts TCP performances such as TCP throughput and TCP packet retransmission rates in real time on flow basis. The tool collects flow performances into a path performance using the source and destination IP network numbers. We also applied the implemented tool to measure performances of paths to individual dial up access points in a commercial ISP network. The results show that the proposed path-based passive measurement method is useful to characterize individual path performances of ISP networks

References

[1]: A. Feldmann, A. Greenberg, C. Lung, N. Reingold, J. Recford and F. True, "Deriving Traffic Demands for Operational IP Networks: Methodology and Experiences," Proceedings of ACM SIGCOMM 2000, August 2000.

[2]: S. Bhattacharyya, C. Diot, D. Jorets and N. Taft, "Pop-Level and Access-Link-Level Traffic Dynamics in Tier-1 POP," Proceedings of ACM SIGCOMM Internet Measurement Workshop 2001, November 2001.

[3]: V. Paxson, G. Almes, J. Mahdavi and M. Mathis, "Framework for IP Performance Metrics," RFC 2330, IETF, May 1998.

[4]: Cisco Systems Inc., "NetFlow FlowCollector Homepage,"http://www.cisco.com/univercd/cc/td/doc/product/rtrmgmt/nfc/.

[5] C. Fraleigh, C. Diot, B. Lyles, S. Moon, P. Owezarski, D. Papagianaki and T. Tobagi, "Design and Deployment of a Passive Monitoring Infrastructure," Proceedings of Passive and Active Measurement Workshop (PAM2001), April 2001.

[6]: T. Kato, T. Ogishi, A. Idoue and K. Suzuki, "Performance Monitor for TCP/IP Traffic Analyzing Application Level Performance," Proceedings of ICCC'99, November 1999.

[7]: V. Paxson, "End-to-end Packet Dynamics," IEEE/ACM Transactions on Networking," Vo. 7 (3), pp. 277-292, 1999.

[8]: H. Balakrishnan, V. Padmanabhan, S. Seshan, M. Stemm and R. Katz, "TCP Behavior of a Busy Internet Server: Analysis and Improvements," Proceedings of Infocom'98, March 1998.