Monitoring and Prediction of Transport Protocols Throughput for Complex Networks

Alexander A. Sannikov, Olga I. Bogoiavlenskaia, Iurii A. Bogoiavlenskii
Department of Computer Science,University of Petrozavodsk
Lenin St., 33, Petrozavodsk, Republic of Karelia, 185640, Russia
sannikov@cs.karelia.ru

Abstract

End-to-end network throughput has great influence for performance of applications. Availability of forecasted performance metrics will be very useful in Stream-media services, distributed systems and other. Using these metrics we can redesign network, avoid network congestion and schedule data access in distributed systems. Aim of this work is development of the system for analysis and prediction of network performance metrics, with good accuracy and well detailed for different types of analysis such as time series based. Typical performance metrics are bandwidth, round trip time and its jitter, and finally packet loss rate. All of them are avaliable at transport layer, therefore method of passive measurement on end-hosts has been chosen. For example in case of TCP-protocol we can collect data directly from congestion control mechanism. As well described in RFC 2722 [1], such system should have four subsystem. Those are meter, meter reader, manager and analysis application. Meters, or probes in our terminology, are used for registration of network activity and placed immediately on the end-hosts. Current prototype of probe oriented only for TCP working over IPv4 and for passive measurement. It is based on GetTCP system [2]. GetTCP extracts internal variables of TCP congestion control mechanism and delivers them to user-space for following transmission.

In addition to in [1] described functionality collectors are used not for only data transmission, but also for access control, data replication and preprocessing, collecting of information for complex routes. Distributed decision also allows us to collect data in whole subnetwork, therefore we can get more complete set of data. So as we can see collectors are much more complicated and smart than another nodes of this system. And finally, main difference from RFC decision is long-term storage of data, for long-duration trends analysis and periodical statistical reports.

Manager is small, universal tool for remote customization and control of each other nodes. It will provide ability to change collector's data access control and replication politics. Also it is a tool for configuration of routes in complex network analysis. Main aim of the manager is providing remote control of system without local access to all nodes.

Last class of subsystem is analysis applications. It is a whole class of application which get data from collector nodes, interpret them and generate reports. It will be a statistical report for some time interval, forecasted value of throughput between some hosts or other type of analysis. In contrast to probes and collectors, manager and analysis applications usually perform by end-user request. For data interchange between subsystem we develop special protocol working over UDP. It provides unified network interface for all subsystems. This protocol is split on four sublevels.

First level describes rules of data interchange between probes and collectors. It is based on subscription model and for receiving data collector should send subscription request to probe. After subscription collector begins to receive data about network flows in the format specified by the probe. At this level we can observe a significant data flow from probe to collector.

Next sublevel is controlling one. It describes setup interface for probes and collectors. Special feature of this level is the need in security, encryption and access control.

Third sublevel implements data interchange between collectors and analysis applications. By request from analyzer another side should transmit data corresponding to the request.

The data interchange between collectors (fourth sublevel) is most complicated and it is organized in the form of Peer-to-Peer-like network. Since probes generate large volume of data which should be processed, data are replicated between few collector nodes. As the result collected data will be divided between collectors and processed by them separately or data from different sources will be processed on different collectors and result of preprocessing will be replicated between them. This mechanism will reduce workload of all network nodes and replication of processed data will reduce load of channels. All data about traffic flows from one host to another are aggregated and averaged for some short time frame. As the result we have a set of aggregated averages of metrics for some time interval for all flows between whole subnetwork and remote host or only for predetermined paths.

Each record about flow contains information including source and destination hosts, time stamp and metrics itself, and represent an atomic unit for any kind of following analysis. In real networks we can found a problem when one application layer connection splits into several transport layer connections. For example let host A which is connected to Internet through proxy P_A placed outside it's LAN, tries to send data to host B. In this case we have two transport layer connection per one application layer, by this way we can't estimate performance of such configuration by usual means. But when placing probes on A and P_A we can get performance metrics for segments $A \rightarrow P_A$ and $P_A \rightarrow B$, and then determine performance of the whole path. As the result we have developed architecture of distributed system for collecting of network performance metrics, which is scalable and extendable for different sizes of subnetwork. Deployment of this system will be used in analysis and prediction of performance of real networks. Existing prototype shows full operability and provides good testbed for testing of logic for data replication and provides data for statistical analysis. It contains implementation of probe and collector, library for collector's data storage and library with implementation of common part of our network protocol. Also in future we plan to make internal data format more flexible and extendible. Current internal data format is only thing not amenable to expansion, other subsystem are quite extendable and scalable. At this moment network protocol allows to set $printf$-like format string in sub-header of package. Using this field we can change data format.

Index Terms: Passive Measurement, Network Performance, Quality of Service, Transport Protocol, Throughput Prediction.

REFERENCES

[1] N. Brownlee, C. Mills, G. Ruth, "Traffic Flow Measurement: Architecture," RFC 2722, Internet Engineering Task Force, Oct. 1999 http://tools.ietf.org/html/rfc2722

[2] V. A. Ponomarev, O. Yu. Bogoyavlenskaya, Yu. A. Bogoyavlenskiy Configurable Kernel-Level Monitoring System of the TCP Behavior in *Information technologies* 1 2010 p. 54 59