



# **A packet quantitative balance study on IP access network**

**Yan Yang**

**April 24, 2007**





# Introduction

- Rapid traffic growth adds complexity to network management. To provide users with high quality services, a profound understanding of the **network behavior and its characteristics** is badly needed.
- Network measurement is based on **network metrics**.
- **Advantages and disadvantages for transport layer (especially for TCP) oriented and IP layer oriented metrics**——more semanteme, easy for analysis and modeling, stateful, difficult to calculate on real time VS stateless, easy to obtain on real time, difficult for modeling.



# Related work

- Currently, various metrics have been proposed for network management and inspection, but they are insufficient. Many mechanisms, especially in the field of network security to counter SYN flooding attacks, are all stateful, that is, maintaining the states of all the TCP connections is needed, which largely degrades the end-to-end TCP performance and is inadequate for real time requirement.
- About quantitative balance studies in network, there have been some findings. From the byte view(BPS) it is not symmetrical between the traffic that comes into and goes out of a network, however, it may be quite different from the packet view(PPS). [10] shows a macroscopical quantitative balance of TCP packets which are used to control TCP sessions, but the quantitative balance is aimed at TCP's control packets.



# Our Work

- Define the quantitative balance metrics for access network and TCP connection respectively and convert the analysis of the former to the calculation of the latter.
- Study the quantitative balance between the packets that come into and go out of the network from the perspective of TCP connection and IP access network.
- Based on the above research and by referring to the methods used to establish boundary of medical metrics, we attempt to identify the normal reference range for the latter metric. This normal reference range is made up of two boundary lines called “yellow” and “red”, so the whole range is divided into three intervals that do not intersect—“red”, “yellow” and “green”, through which we make it a utility metric.





# Data Source

---

- Trace1: 100Mbps link of WIDE backbone across the Pacific Ocean in January 7th,2005. 24 hours.
- Trace2: collected by a platform for traffic measurement in high-speed network – WATCHER, which is run in the boundary of a provincial network of CERNET. November 10th, 2005 and 24 hours.



# ratio of bidirectional packets in TCP connections

- It can be inferred from TCP mechanism(especially Sliding Window Protocol) that the number of packets in both directions has a proportional relation. So we make a statistic of packets in trace1 in terms of complete TCP connection.
- Fig.1 shows the percentage of ratio of bidirectional packets in different ranges.

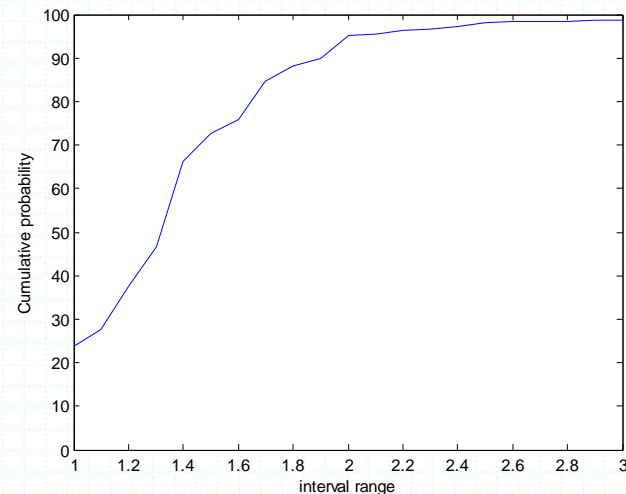


Fig.1The percentage of ratio of bidirectional packets in different ranges.

## Definitions and properties of the quantitative balance metrics of packets

- **Definition 1** Given a time slot  $t$ , denote the number of packets that the access network receives  $\lambda_{in}(t)$  and the number of packets that the access network sends  $\lambda_{out}(t)$ , define the function  $x(t) = \frac{\lambda_{in}(t)}{\lambda_{out}(t)}$  as the ratio of in-and-out packets of the access network during the time period  $t$ .
- **Definition 2** Denote the TCP connection between the access network and the outside network  $li$ , the number of packets that the access network receives in this connection  $\lambda_{in}(li)$  and the number of packets that the access network sends in this connection  $\lambda_{out}(li)$ , define the function  $R(li) = \frac{\lambda_{in}(li)}{\lambda_{out}(li)}$  as the ratio of in-and-out packets of connection  $li$ .



## Definitions and properties of the quantitative balance metrics of packets

- **Definition 3** Given a time slot  $t$ , denote the number of packets that the access network receives  $\lambda_{in}(t)$  and the number of packets that the access network sends  $\lambda_{out}(t)$ , define function  $d(t) = \frac{\lambda_{in}(t) - \lambda_{out}(t)}{\lambda_{in}(t) + \lambda_{out}(t)}$  as the quantitative balance metric of bidirectional packets for access network during the time period  $t$ . Obviously,  $d(t) = \frac{x(t) - 1}{x(t) + 1}$ .
- **Definition 4** Denote the TCP connection between the access network and the outside network  $li$ , the number of packets that the access network receives in this connection  $\lambda_{in}(li)$  and the number of packets that the access network sends in this connection  $\lambda_{out}(li)$ , define function  $D(li) = \frac{\lambda_{in}(li) - \lambda_{out}(li)}{\lambda_{in}(li) + \lambda_{out}(li)}$  as quantitative balance metric of bidirectional packets for connection  $li$ .

# Property of the metric

- **Theorem 1:** Suppose the ratio of number of packets in two directions of a TCP connection is within the range  $[a, b]$ . If  $t$  is large enough and only TCP volume is considered, then the range of  $d(t)$  is  $[\frac{a-1}{a+1}, \frac{b-1}{b+1}]$ .

Proof ignored.

- select  $a=1/2$  and  $b=2$ , then  $d(t) \in [-1/3, 1/3]$  with the error 5%.
- set a coefficient to reduce the effect of UDP traffic, so that the range of is enlarged to  $[-0.35, 0.35]$ —the “red” cordon





## The method to evaluate the health of network based on the quantitative balance metric of bidirectional packets

- According to the medical method, most of the normal values measured should be within this normal reference range, where “most of” is regularly set 80%,90%,95% and 99%.
- In the case of unknown distribution, set the percentage index of  $\alpha$  and  $\beta$  as the boundary line of the normal reference range. Using  $d(t)$  to evaluate the quantitative balance of packets, there are  $(1+\alpha\%-\beta\%) \times 100\%$  networks whose metric is out of the normal reference range.  $\alpha$  and  $\beta$  can be set according to the network.
- We will give a normal reference range of “health” in terms of  $d(t)$  based on trace2 and regard this normal reference range as another cordon of this metric(“yellow cordon”).



## The method to evaluate the health of network based on the quantitative balance metric of bidirectional packets(Cont.)

### —Instance:

❖ We take the 8-hour-long data(12:00~20:00) from trace 2 and carry on analysis on the time granularity of 4 hours to satisfy the condition of “a time slot large enough”. So the number of individuals of the sample turns to be  $81 \times 2 = 162$ . The frequency distribution is shown in Fig.2. (Set the groups 14 and group gap 0.05)

❖ According to the above definition and method, in the case of unknown distribution, we set  $\alpha=2.5$  and  $\beta=97.5$ , that is the percentage index of 2.5 and 97.5, to get the normal reference range of quantitative balance metric of bidirectional packets for access network  $[-0.189, 0.236]$  — **“yellow Cordon”**.

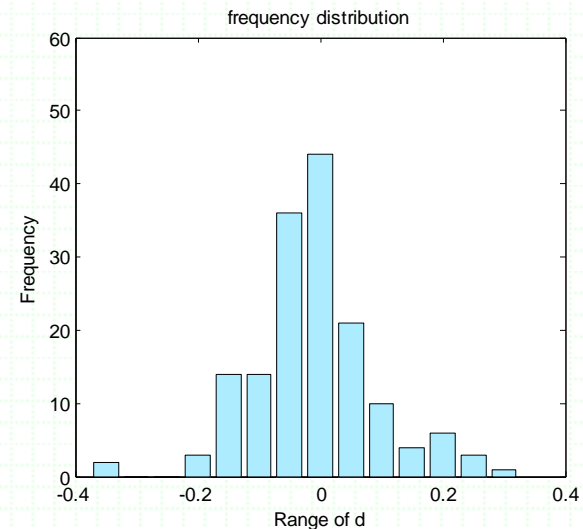


Fig.2 The frequency distribution of metric d



# Other issues

In real circumstance, the following issues should be considered as well:

- For a certain access network, when  $d(t)$  is within some range steadily, it is an indication of its network constitution---servers and users.
- For some access networks, the priority of the incoming and outgoing traffic may be different.---Discard packets with low priority in some direction according to the balance degree that  $d(t)$  represents and the service level agreement as well.
- The scale of access network---network scale of small or middle is appropriate, e.g., campus networks or access networks with smaller scale.



# Analysis of Time Granularity

— We take the traffic between 00: 00~09: 00 in trace2 for example. Fig.3 shows the number of in-and-out packets of the network at the time granularity of 300s.

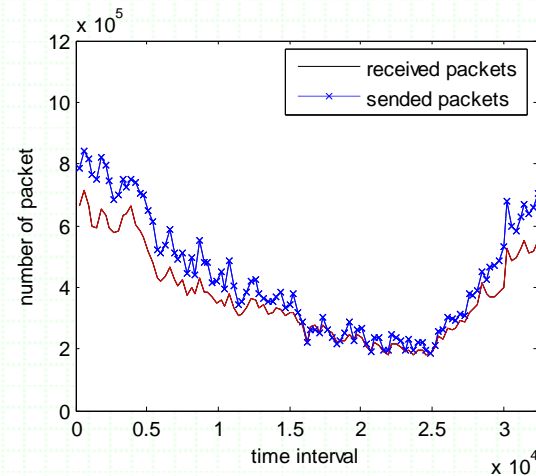


Fig.3 The number of in-and-out packets of the network in the time bin of 300s during 00:00~09:00



# Analysis of Time Granularity

—All the above analysis is based on the condition that the time granularity is large enough while the time granularity of 4 hours is unacceptable for the utility of the metric.

We refer to the idea in “*Internet Traffic Characterization*” (CLAFFY K C.’s Ph D dissertation) and try to obtain a reasonable time granularity to calculate the metric in application.

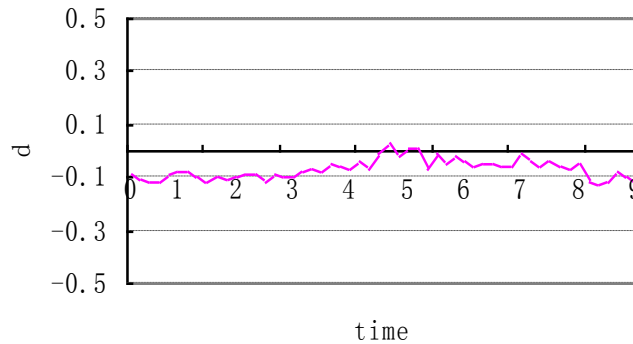


Fig.4 Time Granularity of 600s

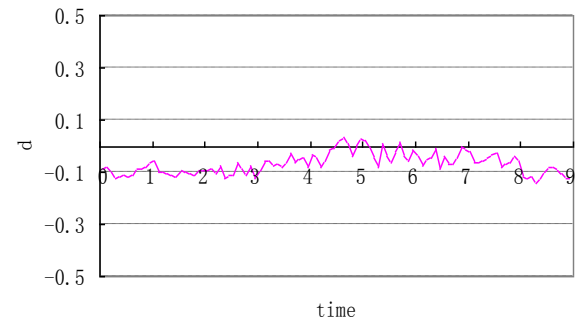


Fig.5 Time Granularity of 300s

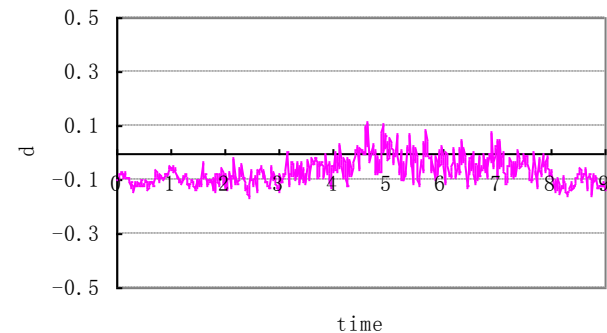


Fig.6 Time Granularity of 60s



# Conclusion

- **Define reasonable and effective metrics**---the demand of Internet measurement standardization, which is fundamental to the research of network behavior. **The difficult** lies in the use of these metrics rather than defining them---the acquirement of metrics(including two phases of parameter acquirement and calculation) and the other is to identify the boundary line of normal range.
- **This paper** analyzed the range of the metric based on the characteristics of TCP connection and discussed the method of mapping it to IP layer oriented metric . We also give the range (access network oriented metric) on the ground of real trace and statistical results. Furthermore, we improved it by referring to medical statistics to get a more accurate range , through which we got the “red cordon” and “yellow cordon” for the health range of the metric thus made it a practical utility.
- **The significance** of research in this paper lies in the method to obtain the boundary line of the metric rather than the boundary itself. The boundary line, especially the “yellow cordon” can be adjusted according to the running state of the network; And we hope that this kind of method can be applied in similar research for metrics, which is also a key point for our **further research**.



# Reference

- [1] Marina Fomenkov, Ken Keys, David Moore, KC Claffy. "Longitudinal study of Internet traffic in 1998-2003." In *Winter International Symposium on Information and Communication Technologies (WISICT)*, Cancun, 2004
- [2] D. Moore, G. Voelker and S. Savage, "Inferring Internet Denial of Service Activity", *Proceedings of USENIX Security Symposium' 2001*, August 2001.
- [3] J. Lemon, "Resisting SYN Flooding Dos Attacks with a SYN Cache", *Proceedings of USENIX BSDCon'2002*, February, 2002.
- [4] D. J. Bernstein and Eric Schenk, "Linux Kernel SYN Cookies Firewall Project", <http://www.bronzesoft.org/projects/scfw>.
- [5] Check Point Software Technologies Ltd. SynDefender: <http://www.checkpoint.com/products/firewall-1>.
- [6] Netscreen 100 Firewall Appliance, <http://www.netscreen.com/>.
- [7] V. Paxson. Framework for IP Performance Metrics, May 1998. Internet RFC 2330.
- [8] WIDE MAWI WorkingGroup, "MAWI Working Group Traffic Archive", <http://tracer.csl.sony.co.jp/mawi/samplepoint-B/20050107/>
- [9] <ftp://tracer.csl.sony.co.jp/pub/mawi/tools/tcpd-tools.tar.gz>
- [10] Gong Jian, Peng Yanbing, et al.. Macroscopical Quantitative Balance of TCP Packets. *Chinese Journal of Computers*, 2006, 29(9), 1561~1571
- [11] Cheng Guang, Ding Wei, Gong Jian. General Platform for Traffic Measurement in High-speed Network-WATCHER. In *Proceedings of CSIT*, 2004, 122~128
- [12] CLAFFY K C. *Internet Traffic Characterization*: [Ph D dissertation ]. San Diego : University of California, 1994.



Thanks !

