Characterization of Internet Traffic - Interesting Facts and Figures

Lukas Karrer, 1. 5. 2000

Abstract

The Internet is growing rapidly. Users, traffic level and complexity have exploded in recent years. The more the Internet is growing, the stronger rises a need for detailed investigation of Internet traffic.
I will summarize two papers, which have examined Internet traffic in terms of wide area traffic pattern and characteristics as well as end-to-end packet dynamics. The findings in these two papers [1] [2] are based on large-scale studies conducted on the „real world“ Internet backbone in 1995 and 1997. My focus will be on interesting facts and figures. Both studies reveal astonishing facts which help in understanding „what is going on“ in today’s Internet.

Part I: End-To-End Packet Dynamics

In Part I, I will discuss the results from a large-scale study of packet dynamics conducted by tracing 20‘000 TCP bulk transfers between 21 selected sites throughout the world. Recording the traffic of both sender and receiver, the measurements not only allow to investigate the prevalence of unusual network events but also asymmetries exhibited in the Internet. Unusual network events cover mainly pathological behavior such as packet reordering, packet loss and packet corruption.

Even though routers employ FIFO Queues, packet reordering was surprisingly common. 12% out of all recorded transfers experienced at least one packet delivered out of order. Usually, reordering involved only very few packets. Further investigation revealed, that out-of-order delivery is highly site-dependant and asymmetric. One site experienced reordering in 15% out of all packets transferred in one direction, but almost none the other way around. Out-of-order delivery does not depend on packet loss, but correlates to route fluttering. These figures are large enough to question the FIFO model, but do not seem to have significant impact on TCP performance.
A further pathological behavior is packet replication. Only one site suffered replication, which turned out to be due to improper configuration of a link level device. As no other site experienced packet replication, this phenomenon does not need to be considered in any modeling.
Packet corruption however, is a pathology, which impacts heavily on TCP. An even distributed 0.02% out of all packets had inconsistent checksums. A corruption rate of 1 in 5000 is quite surprising, as link-layer checksums should detect erroneous transfers. Analyzing the collected data more closely, large data packets were much more prone to incorrect checksums than ACK or smaller data packets. It seems that

corruption occurs inside routers due to cache inconsistencies or botched DMA, where this pathological behavior goes undetected. The rate of occurrence is quite alarming. Using TCP's 16bit checksum, an average of one bad packet in 300 million is erroneously accepted by receiving TCP. Calculating with today's size of the Internet, certainly many corrupt packets are accepted each day, resulting in distorted data.

As the last pathological behavior, I will take a look at packet loss. The collected data shows that about 5.2% out of all packets were lost on their way. As a result, about half the connections experienced packet loss. Packet losses too, is highly asymmetric and site dependant. One connection for example, experienced packet loss of over 65% in one direction. Even though performance was seriously impeded, TCP was able to succeed in such an "unfriendly" environment.

Other interesting findings are, that packet loss is extremely bursty. Further, two third out of all packets sent were queued within at least one router. The amount of ACK packet loss, which amounts to be about the same number as data packet loss, indicates, that packet loss does not correlate to the varying window size.

Conclusions of Part I

The Internet exhibits a wide range of different behavior – there are no typical aspects of packet dynamics. Common assumption frequently made for simulations are faulty. Assumptions such as FIFO queuing, path symmetries or independent loss are violated quite frequently. Simulating the Internet is not easy!

Part II: Wide Area Traffic Patterns and Characteristics

Part II focuses on the characterization of commercial IP traffic in terms of traffic volume, duration and patterns as well as protocol distribution. Two measurements, over both a 24 hour and 7 day time period on MCI Worldcom's IP backbone, provided data for analysis. A custom-built PC, which was connected to a splitted optical OC3 fiber, was used for data acquisition.

IP traffic on MCI's domestic backbone followed a clear and predictable 24-hour pattern. Byte volume increased 500% between the hours of 5 AM to 10 AM, hitting a peak of 50 Mbits / second at noon. This pattern repeated daily, with a 20% reduction of byte traffic over weekends. A similar pattern could be observed by looking at packet volume. Interesting enough, the total packet count in one direction of the link was about 50% smaller than in the other direction, whereas total byte count did not show this behavior. This leads to one important conclusion – byte count cannot be used for packet count predictions. During peak hours, there existed as many as 240‘000 concurrent sessions over the link.

Packet size also varied over time and amounted to an average of about 170 bytes per packet during the night and 200 bytes during daytime. Plotting packet size in a histogram, we can observe a predominance of small packets. About 40% of all packets were 40 bytes and smaller, indicating TCP ACK, SYN, FIN packets. Other peaks occurred at around 552 bytes and 1500 bytes. The mode observed at 552 bytes per packet is probably due to missing PATH MTU Discovery. 10% of all packets were 1500 bytes of size. This indicates Ethernet attached hosts.

Similar numbers as stated above, apply to MCI's international backbone between the UK and the US. As expected, byte volume in the direction leaving the US is about double as the volume entering the US. A clear asymmetry in backbone traffic.

Next, I would like to go into details on traffic composition. Focusing on IP protocols, we observe that TCP was by far the predominant protocol, accounting for over 95%

of byte volume. In terms of byte volume, UDP is almost negligible. In terms of packet volume however, UDP, specifically DNS traffic, accounted for about 10%.

Web traffic dominated as the largest Internet application, with byte traffic accounting for more than 75% of overall traffic. Other major applications, such as FTP, SMTP and NNTP made up for the rest. Traffic volume originating from these protocols is somewhat higher during night hours, but rarely exceeded the 10% margin.

These figures are quite stunning! In a similar study conducted in 1995, only 2 years earlier, Web traffic dominated, but the distribution was much more balanced. WWW accounted for only 21% of overall volume, NNTP 14%, SMTP and Telnet still 8%. It is interesting, how a so-called killer application can change the landscape of the Internet in such an impressive way.

Conclusions of Part II

I wonder what influence multimedia protocols will have in the future. In 1997‘s study, RealPlayer traffic averaged for less than 1% of bytes and packets. I predict multimedia like IP telephony and video to be the next killer application. I wonder how much percent of traffic these protocols will account for in the near future!

Part III: Doing your own Network analysis

In the last part of my talk, I would like to introduce a tool to perform your own network analysis. Ntop is an open-source Unix tool that shows the network usage, similar to what the popular top command does. Based on libpcap, a commonly used packet capture library, ntop reports in a similar fashion as described in the last part.

Amongst other things, ntop

- sorts network traffic according to many protocols
- shows network traffic sorted according to various criteria
- displays traffic statistics
- shows IP traffic distribution among the various protocols
- analyses IP traffic and sorts it according to the source/destination
- displays IP traffic subnet matrix (who's talking to who?)
- reports on IP protocol usage sorted by protocol type

Ntop is accessed via web browser or command line.

More information and downloads are accessible via ntop’s web page at http://www.ntop.org

[1] **End-to-End Internet Packet Dynamics**

Vern Paxson, Network Research Group,
Lawrence Berkeley National Laboratory
University of California, Berkeley
vern@aciri.org

[2] **Wide-Area Internet Traffic Patterns and Characteristics**

Kevin Thompson, MCI Coorperation, Reston VA
kthomp@mci.net