Internet Traffic Measurement

Aniket Mahanti

I. Internet Traffic Measurement

Internet measurement can be categorized based on network monitor placement (edge network vs core network), measurement/analysis tools used (hardware-based vs software-based), probing mechanism (passive vs active), and the number of vantage points (single viewpoint vs multiple viewpoints) [2], [3], [8].

A. Edge Network vs Core Network

Figure 1 shows a schematic diagram of the Internet and its entities. ISPs are organized into three tiers: Tier-1 (continental backbone networks), Tier-2 (regional networks), and Tier-3 (access networks). Tier-1 ISPs are large organizations that peer with other Tier-1 ISPs to form an Internet backbone network. Examples of Tier-1 ISPs are AT&T and Tata Communications. Tier-2 ISPs peer with other Tier-2 ISPs and purchase connectivity from Tier-1 ISPs. An example of a Tier-2 ISP is Comcast. Tier-3 ISPs act as resellers, and purchase connectivity from other networks (Tier-1 or Tier-2) to connect to the Internet. Customers buy access to the Internet from Tier-3 ISPs through various access technologies such as cable, DSL, fiber optic, or WiMAX. Previously, content providers were generally connected to the Internet through Tier-3 ISPs. This trend seems to be changing, with large content providers (such as Google) building their own private WAN or being connected directly to Tier-1 networks [6], [7].

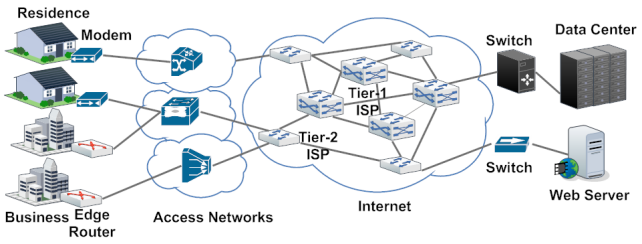

Fig. 1. Edge networks and the Internet core

Internet measurements can be performed at various locations in the Internet. Measurements can be performed at individual end systems (left side of Figure 1) to understand performance, service quality, and usage. End user host machines can be easily instrumented with software to capture incoming and outgoing traffic. Several commercial and customizable tools (e.g., `tcpdump`[1], Wireshark[2]) exist for this purpose. While these measurements serve a limited purpose when the sample size is small, such measurements are a good starting point for performing large-scale studies. Also, these studies are often performed for understanding the characteristics of new access technologies. For example, Falaki *et al.* [4] studied smartphone traffic characteristics with the data collected from 43 users. On the other end (right side of Figure 1), server logs from content providers can provide several insights into the trends of Internet traffic. Such data are considered proprietary and are not shared outside the enterprise.

Internet measurements can be taken at edge networks such as campus or enterprise networks. These networks consist of several thousands of users, although their demographic may be skewed towards one category of users (e.g., students in a campus network). Such measurements provide a large data sample to understand Internet traffic characteristics. Measurements at edge networks are easier to undertake because these networks are maintained by a single organization. Clearance to capture traffic containing private information may be easier to obtain. Internet measurements in the core network are more difficult, primarily due to administrative, privacy, and proprietary concerns. ISPs collect logs of traffic flowing through their routers, however, such data is considered proprietary and are not shared outside the organization. Data from core networks aggregates traffic from hundreds of thousands of users, and traffic is more representative due to the inherent diversity of users.

B. Hardware-based vs Software-based Analysis

Internet traffic measurement tools can be classified into hardware-based or software-based tools. Hardware-based tools such as network analyzers are specialized equipment for measuring and analyzing traffic in a network. Off-the-shelf network analyzers come with measurement devices (such as wireless probes or high-speed network cards) and software for real-time analysis and visualization of Internet traffic (e.g., WildPackets OmniPeek Network Analyzer[3]). Such products are often geared towards network operators for managing and troubleshooting networking issues, and may not be well-suited for research purposes.

Software-based tools such as `tcpdump`, Wireshark, `Bro`[4], and `ntop`[5] are widely used by networking researchers for Internet measurement. Such tools use kernel-level modification to network interfaces of commodity network cards to capture traffic. This approach is inexpensive and provides greater functionality for customizing measurement and analysis. Custom analysis scripts are available (and scalable) for analyzing huge volumes of data, which may not be possible with commercial applications. This approach also allows for offline traffic analysis, which is useful for traffic characterization and allows for appropriate privacy considerations to be taken into account. Online analysis is more suited for time-sensitive applications such as intrusion detection or application identification.

[1] http://www.tcpdump.org/
[2] http://www.wireshark.org/
[3] http://www.wildpackets.com/products/omnipeek_network_analyzer
[4] http://bro-ids.org/
[5] http://www.ntop.org/

C. Passive vs Active Measurement

Passive network measurement is performed by listening to all traffic passing through routers or hosts. This approach requires that the traffic passing through is unaltered and there is minimal disruption to the operation of the network. Passive measurement at an edge network is done by connecting a network monitor to the edge router. All incoming and outgoing traffic to the edge network is replicated and sent to the network monitor.

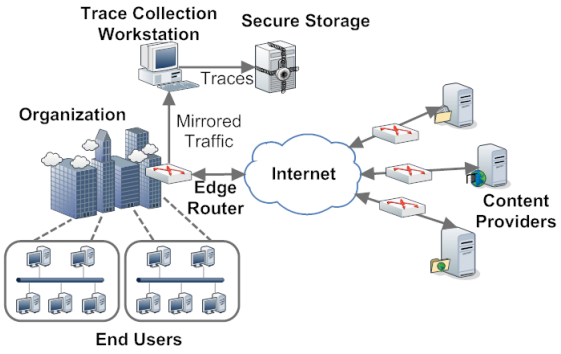

Fig. 2. Passive measurement

Figure 2 illustrates the principles of passive measurement. Depending on the size of an organization, passive measurement involves capturing large amounts of data. Data storage is of crucial concern. Collecting the most appropriate data can reduce the size of the resulting dataset; instead of collecting full packet traces with payload, only the transport-layer and application-layer headers may be saved [1]. Sampling may be employed if analysis of the full data is not required [3]. Compression of the data can also drastically reduce the size of the resulting datasets. Compression also assists when exporting the data from the network monitor to the storage server. The time of export should be chosen appropriately to have minimum disruption to normal operation of the network. Due to the non-intrusive nature of passive measurements, they provide an accurate representation of network traffic.

Passive measurement mechanisms are also built into routers, switches, and other network devices. These mechanisms, such as Management Information Base (MIB) and Simple Network Management Protocol (SNMP) polling, only provide coarse-grained aggregate information, like total bytes transferred, total lost packets, and interface statistics [3]. These statistics are collected by polling the devices at coarse-grained time intervals (e.g., 5 min). The large polling time intervals and lack of detailed information make this approach infeasible for continuous monitoring and export of data.

Privacy is an important concern in passive Internet measurement. Appropriate measures should be put in place such that individual users cannot be identified and associated with the data collected. This entails translating user-identifiable information such as IP address or Medium Access Control (MAC) address into an anonymous user id. This step may be performed during data collection or when post-processing the data.

Active measurements involve generating special probe traffic from one host to another host. Probe traffic could contain small UDP packets with little or no payload. Figure 3 illustrates the principles of active measurement. Active measurement perturbs the flow of normal traffic. It is used to understand end-to-end performance like latency and bandwidth availability. Well-known active measurement tools are `ping` and `traceroute`. Another example of active measurement is requesting a Web page from a Web server to measure the load time.

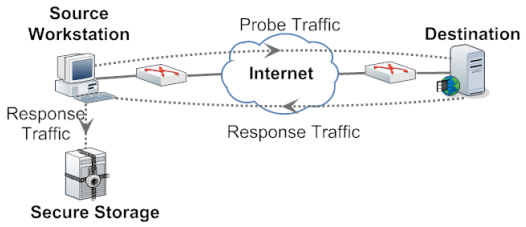

Fig. 3. Active measurement

Active measurements typically do not require large storage, and it is used to complement passive measurements. Privacy is also not a concern with active measurement. The type of probe packet (using User Datagram Protocol (UDP) instead of Internet Control Message Protocol (ICMP)), size of the probe packet, time of request, and frequency of requests are important factors in active measurements.

D. Single Viewpoint vs Multiple Viewpoint

Performing measurements from multiple observational viewpoints is important for understanding the "big picture" behind performance, traffic characteristics, and usage of Internet applications. Both passive and active measurements can be used in tandem to measure different aspects of Internet applications and networked systems. Measurements from multiple viewpoints enhance the representativeness of the results by reducing skews and anomalies in the data. For example, to study traffic characteristics of Web applications (e.g., YouTube), passive measurements can be performed at edge networks to understand local usage characteristics, along with active measurements (through Application Programming Interface (API) calls) to glean global usage trends [5].

REFERENCES

[1] M. Arlitt and C. Williamson. Understanding Web Server Configuration Issues. *Software: Practice and Experience*, 34(2):163–186, 2004.

[2] M. Crovella and B. Krishnamurthy. *Internet Measurement: Infrastructure, Traffic and Applications*. Wiley, 2006.

[3] N. Duffield. Sampling for Passive Internet Measurement: A Review. *Statistical Science*, 19(3):472–498, 2004.

[4] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin. A First Look at Traffic on Smartphones. In *Proc. ACM SIGCOMM Internet Measurement Conference*, Melbourne, Australia, November 2010.

[5] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube Traffic Characterization: A View from the Edge. In *Proc. ACM SIGCOMM Conference on Internet Measurement*, San Diego, USA, October 2007.

[6] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse? In *Proc. Passive and Active Measurement Conference*, Cleveland, USA, April 2008.

[7] C. Labovitz, S. Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet Inter-domain Traffic. In *Proc. ACM SIGCOMM Conference*, New Delhi, India, August/September 2010.

[8] C. Williamson. Internet Traffic Measurement. *IEEE Internet Computing*, 5(6):70 –74, November/December 2001.