Performance Issues of Internet Protocol Versions

Gowtham Mamidisetti, G. Tej Varma

Abstract— In this paper, we develop an analysis method that matches DNS information so that we can compare and contrast performance over protocols for a variety of Internet services. Our initial analyses focus on the basic services that are accessed using protocols, observed client behaviors, and a presentation of performance characteristics of services using both IPv4 and IPv6. Our objective is to detect and expose differences by passive measurement without access to application traffic payloads. To demonstrate our method, Our method uses data collected on the IPv6, including both DNS requests/responses and flow export records for dual-stack hosts operating .Our method expose various performance characteristics of Internet services that support IPv6.

Keywords- IPv4, IPv6, DNS.

I. INTRODUCTION

In this paper we present a new method for assessing the performance of Internet services over IPv6. Our objective is to

(i) Accurately assess performance based on passive measurements. (ii) provide the capability to compare and contrast IPv6 performance with that of IPv4, and (iii) provide an assessment that is both independent of the end-hosts and the Internet services they access. We perform this assessment based on passive measurements gathered at two observation points: one at or near the clients recursive Domain Name System (DNS) resolver and the other at any point along the end-to-end path. Our approach does not need privileged knowledge of the Internet services nor special access to the end-hosts involved in the exchange of traffic. We then develop a framework and tools to detect and inspect performance differences between IPv6 and IPv4 for Internet services.

Our performance assessment method is predicated on the fact that client hosts, on IPv4 and IPv6, necessarily employ some mechanism to discover the IP address of an Internet service before interacting with it. For many types of port-based services and for most of those with early IPv6 support. When a service supports IPv4 and IPv6 simultaneously, clients use a common mechanism for both Internet protocol versions, such as the DNS.

Our method determines service performance in three steps: (1) measurement in two forms: (a) full capture of low-volume DNS query/response packets and (b) collection of 5-tuple IP flow export records with duration and byte count of high-volume application traffic; (2) classification of flows source and/or destination IP addresses by matching them to their corresponding domain names (when possible) based on query names and the resulting IP addresses in DNS responses;

Manuscript received January 15, 2014.

Gowtham.Mamidisetti, Assistant Professor, Department of Information Technology, Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India

Tej Varma, Assistant Professor, Department of Information Technology, Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India

(3) performance inference by constructing a distribution of flow bit rates and applying statistical techniques.

II. METHOD AND IMPLEMENTATION

Our goal is to develop a performance assessment framework with the following characteristics or features:

- a method employing rendezvous-based traffic classifica-tion and robust statistics to determine and expose IPv6 and IPv4 performance phenomena for Internet services.
- an extensible mechanism for encapsulating the requisite rendezvous and traffic trace data prior to classification and for annotating IP traffic trace data afterward.
- a mechanism for transmitting streams of the encapsulated data to distributed framework components for online analysis in near real time.
- a serialized data file format for storing the encapsulated and annotated data for offline analysis, as in this study.

We utilize this method to assess the performance of traffic exchanged between hundreds of IPv6-capable campus hosts and Internet services with which these hosts rendezvous via the DNS.

A. Direct Labeling

Direct DNS rendezvous-based labeling is performed when TreeTop discovers that a given client end-host knows a peer remote IP address by a domain name as the result of a canonical "forward" DNS query to translate that name to an address. In this case, an nfdump record can be annotated because the client involved has used the DNS to resolve the name of its peer; we call this "CLIENT DNS NAMED". For instance, the sample nfdump record in Figure 1 has been

This direct labeling is the most reliable, but it requires the client host to use the same IP address as both the source of its DNS queries and as its local address when exchanging related application traffic. If that is not the case or if TreeTop does not observe a given client's DNS query response that contained a given peer IP address in an answer, we resort to "consensus" labeling.

B. Consensus Labeling

In our observations, the dual-stack hosts use just one of their IP addresses as the source of their DNS queries: a host's IPv4 address. Thus, for IPv6 flows in this work, we often can't perform direct labeling since the client host's IPv6 address is not usually the client address in the corresponding dnsqr messages. To label these flows' sources or destinations, we, instead, use a consensus-based approach based on the domain names resolved by other DNS clients in the population studied. If another host or hosts resolved a name to the peer address in question, it is generally agreed, by rough consensus of the population, that this host could also have used the DNS and named the peer.

C. Name Sampling

Whether direct or consensus labeling was performed, it is certainly possible that a given flow's source or destination may be known by more than one domain name. For instance A service with the name "www.example.com" might also be known as "login.example.com". In such a case, we would like to annotate an nfdump record with a single name for the peer, such as "*.example.com", but also with what caused the ambiguous label. To expose this information, we perform "name sampling."

D. Port-based Classification

To complement the aforementioned DNS rendezvous-based classifications, we employ traditional port-based application labels from an existing classifier that has been used in prior work. These are: "WWW," "P2P," "FTP," "Streaming," etc., and allow one to distinguish amongst multiple service types that happen to be identified by a single domain name or prefix, such as distinguishing IMAP from HTTPS traffic for Gmail.

The last part of our methodology to assess performance of IPv6 (and IPv4) flows is to calculate the bit rates based on fields already present in the nfdump records: ibyt (input bytes) and td (time, duration), which we do for all flows having a non-zero duration. We rely on the flow export implementation (in the commercial router) in that we assume sufficient granularity, range, and accuracy of these values for the distributions of rate values used in our performance analyses.

III. EMPIRICAL DATA SET

Since we are interested in assessing the performance of Internet services over IPv6 (as compared with IPv4), we select a campus population whose network and client hosts are IPv6-capable. On our campus, there are thousands of dual-stack hosts that reside within 22 IPv4 subnets and one IPv6 subnet and are mixed-use in campus offices and labs.

To gather the traffic traces and input data for this work, we monitor campus traffic at two observation points: (1) the campus clients' recursive name servers, and (2) a campus core router that forwards traffic between the client hosts and the commodity Internet. We perform full packet capture at the campus domain name servers, and collect non-packet-sampled NetFlow version 9 data at a campus core router. Thus, the payload of the DNS traffic is recorded, but the application traffic payload is neither needed nor recorded. Such monitoring of DNS traffic between the client end-hosts and their recursive DNS service and router-based flow export is feasible within the typical networks of large institutions, enterprises, or Internet service providers. Our interest is in the "canonical" DNS traffic, i.e., the standard DNS traffic expected to precede application traffic that consists of a query by FQDN and an answer containing one or more IP addresses associated with the query name. Because we assess performance using flow bit rates, we use non-packet-sampled flow export data that has complete byte and packet counts as well as start time and flow duration.

Both the DNS and flow export data were collected for the 24 hours of World IPv6 Day. We collected ˜14.2M DNS query responses for 2028 total IPv4 and 23 IPv6 client addresses; of these, ˜114,300 AAAA queries resulted in ˜6,200 NOERROR responses. The client hosts' total traffic was represented as ˜58.8 million IPv4 flows and ˜2.4 million IPv6 flows. The number of active IPv6 and IPv4 client hosts numbered in the the hundreds

IV. PERFORMANCE ASSESSMENT.

In this section we provide a sample assessment of the IPv6 and IPv4 performance for the World-Wide Web traffic (HTTP and HTTPS) involving two popular services, Facebook and Google Mail (Gmail), as observed during the 24 hours of World IPv6 Day (June 8, 2011). We selected these services due to the high number of active local client hosts that utilized them, thus providing a larger sample of hosts and their respective flows for each hour of the day. First we consider how the traffic was classified as being associated with each

service and the differences by IP protocol version, then the active clients, and finally, the flow bit rate as distributions in time series with hourly bins.

A. Service Domain Names

we perform our analyses with scripts that process an nmsg stream of nfdump messages annotated with the domain names that client hosts resolved to the source and/or destination addresses of each flow. The traffic is labeled by domain name (FQDN or domain suffix) and that label is the basis for classification, i.e., Facebook or Gmail.

The IPv4 Facebook traffic is that labeled with one of 950 FQDNs that have the suffix "facebook.com" , such as "www.facebook.com","developers.facebook.com","ssl.facebook.com","login.facebook.com", "upload.facebook.com", etc., including 867 different FQDNs matching "*.channel.facebook.com". 867different FQDNs matching "*.channel.facebook.com". The IPv6 Facebook traffic is that labeled with domains including: "www.facebook.com","developers.facebook.com","check6.facebook.com", and various others matching "*.facebook.com".

The IPv4 Gmail traffic is that labeled with the following domains: "gmail.com", "mail.google.com", and "www.gmail.com". The IPv6 Gmail traffic is that labeled with the following domains: "gmail.com", "mail.google.com", "www.gmail.google.com".

B. IPv4 and IPv6 Service Asymmetries

We observe that these services exhibit some asymmetry with respect to the specific DNS names resolved to access them over IPv4 versus IPv6. This is apparently due to differences in implementation of the IPv4 and IPv6 portions of the service. For Facebook, we see that the FQDNs matching "*.chan-nel.facebook.com" were not resolved by AAAA queries (but were resolved by A queries for IPv4 addresses), thus it's probable that Facebook Chat was not yet supported via IPv6 and may fall-back to IPv4 on a dual-stack host. Alternatively, it's possible that the Chat service via IPv6 was overloaded on another FQDN or that it used a non-DNS rendezvous mechanism, and thus may be structured differently (with respect to DNS names).

For Gmail, similarly, names such as "imap.gmail.com" and "smtp.gmail.com" were not resolved by AAAA queries; thus, we believe that these Google Mail features (IMAP and SMTP access) were not available via IPv6 at the time.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

To accommodate this in these results, we select only WWW traffic (by selecting flows with the port numbers for HTTP and HTTPS) so that IPv4 Gmail traffic involving IMAP and SMTP would not be mixed into the performance results for comparison (below).

Such asymmetries or differences in service implementation between IPv4 and IPv6 are a challenge to attempts to directly compare service performance between IPv4 and IPv6. The initial performance analysis presented here assumes that the IPv4 and IPv6 traffic classifications are equivalent for these two services, ignoring the Facebook Chat complication noted above.

C. Active Hosts

we plot the total number of active local host IP addresses, IPv4 and IPv6 for Facebook and Gmail, respectively. The horizontal axis is labeled with the hour of day in local time, five hours west of UTC; the lowest level of activity is at about 0600 and the highest (for these services) during the noon hour, with activity decreasing toward the end of the work day (after 1700 hours). Also, note that there are two regimes: roughly the first 12 hours of World IPv6 Day have low activity

and thus fewer flows for which we examine their bit rates; the latter 12 hours have high activity with many more hosts and flows being used to calculate bit rate distributions.

D. Flow Rates

Bit rate distributions for unidirectional flows were calculated for all non-zero duration flows, simply by dividing the number of bits by the flow duration (in seconds). Bit rate is labeled on the vertical axis in the remaining plots.

In these assessment we employed robust statistics to broadly compare IPv4 and IPv6 performance for two popular services. We find that the number of active hosts (observed via their flows) greatly influences the bit rate distributions. Second, we see evidence of wildly varying near peak (99th percentile) rates in flow export data for a given Internet service. Third, we see that there are regimes in which IPv6 rates are higher and others in which IPv4 rates are higher.

These assessments show that ostensibly the same services over IPv4 and IPv6 exhibit different performance as measured by the clients' sessions' flow bit rate distributions, meeting our objective to develop an analysis method and presentation by which one could expose performance phenomena and assess IPv6 performance. These observations motivate and guide future work including other visualizations and forensic tasks to determine the root causes of performance anomalies for services on both IPv4 and IPv6.

V. CONCLUSION

In this paper we present a method to examine the performance of Internet services on IPv6 and IPv4, with which clients rendezvous via the DNS. Our approach is a new application of TreeTop's traffic classification technique that doesn't perform active measurements, doesn't need "insider" knowledge about those services IP addresses, and doesn't require inspection of application traffic payloads that may be encrypted, obscured, or otherwise unavailable. Instead, it relies on low-volume DNS query/response traffic and easily-obtained application transport information from packet headers.

We demonstrate the feasibility of the approach by implementing our method in the TreeTop Framework and a set of assessment tools. We demonstrate its utility by analyzing DNS traces and flow export data gathered from a campus network with an advanced deployment of IPv6 via dual-stack hosts, focusing on their traffic on the World IPv6 Day. A large proportion of traffic involving services running IPv6 is arranged via the DNS, allowing the associated service (e.g., Facebook or Gmail) to be directly identified. While we find that dual-stack IP implementations complicate measurement, our method is able to infer service identities by a consensus of hosts in the monitored population.

REFERENCES

1. K. Cho, M. Luckie, and B. Huffaker. Identifying IPv6 Network Problems in the Dual-stack World. In Proceedings of the ACM SIGCOMM Work-shop on Network Troubleshooting: Research, Theory and Operations Practice Meet Malfunctioning Reality, New York, NY, 2004.
2. C. Labovitz. Six Months, Six Providers and IPv6. http://ddos.arbornetworks.com/2011/04/six-months-six-providers-and-ipv6/, April 2012.
3. K. Claffy. Tracking IPv6 Evolution: Data We Have and Data We Need. ACM SIGCOMM Computer Communications Review, 41(3), July 2011.
4. D. Plonka and P. Barford. Flexible Traffic and Host Profiling via DNS Rendezvous. In Proceedings of the Securing and Trusting Internet Names Workshop (SATIN 2011), Teddington, UK, April 2011.
5. CoralReef. http://www.caida.org/tools/measurement/coralreef/, 2008.
6. L. Colitti, S. Gunderson, E. Kline, and T. Fefice. Evaluating IPv6 Adoption in the Internet. In Proceedings of the Passive and Active Measurement Conference, Zurich, Switzerland, April 2010.
7. NFDUMP. http://nfdump.sourceforge.net/, 2012
8. Google. Protocol Buffers. https://developers.google.com/protocol-buffers/, 2012.
9. L. Colitti, S. Gunderson, E. Kline, and T. Fefice. Evaluating IPv6 Adoption in the Internet. In Proceedings of the Passive and Active Measurement Conference, Zurich, Switzerland, April 2010.

AUTHORS PROFILE

Gowtham. Mamidisetti, is an Assistant Professor in Information Technology at Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh

Tej Varma, is an AssistantProfessor in Information Technology at Shri Vishnu Engineering College for Women, Bhimavaram, West Godavari Dist, Andhra Pradesh, India