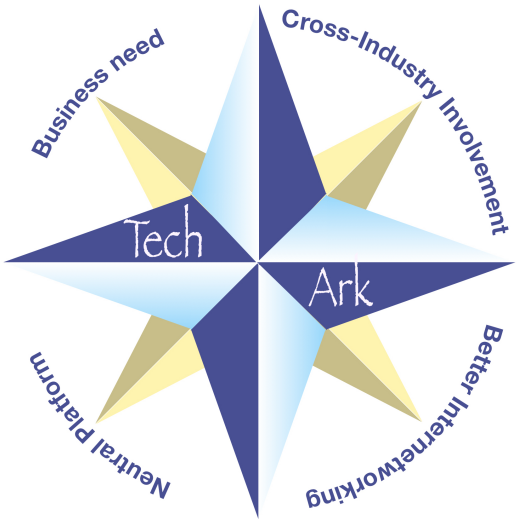
Internet Measurements Landscape (2016): Systems, Approaches and a Comparative Framework

***Leslie Daigle**, Thinking Cat Enterprises LLC[1]; **Phil Roberts**, Internet Society*

December 8, 2016.

This paper provides a high level overview of the many different Internet measurements activities that exist today, outlining a framework with which activities can be compared and contrasted, as they each have strengths and best-suited purposes. In standing back and looking at the framework, it becomes apparent that an important voice and perspective is missing – that of the network operator.

This paper is aimed at the general reader with an interest in the topic, including policy makers, measurement experts wishing to position their work in the landscape of such activities, and network operators seeking to understand available tools, services and practices with regard to measuring the Internet from their network's perspective.



[1] This paper was written as part of the "Network Operator Measurement Activity" of Thinking Cat Enterprise LLC's TechArk project (http://www.techark.org/noma), which is partially funded by the Internet Society.

TABLE OF CONTENTS

1 Introduction

The Internet has been measured and analyzed since the first connection was made between networks, resulting in the many different approaches and projects underway today that make up the landscape of "Internet measurement activities". While more comprehensive lists of all measurements projects are maintained[2], this paper provides a high level overview of activities, outlining a framework with which activities can be compared and contrasted, as they each have strengths and best-suited purposes.

Additionally, this paper highlights a perspective that is missing from publicly accessible Internet measurement: although they are often called on (or required) to provide data for external measurement activities, network operators are usually not the drivers or the designers of the measurements activities. A conclusion of this overview is that having the operator-driven perspective brought into a shared information repository would enhance the existing landscape of Internet measurements and could lead to development of independent "Internet health metrics".

This paper is aimed at the general reader with an interest in the topic, including policy makers, measurement experts wishing to position their work in the landscape of such activities, and network operators seeking to understand available tools, services and practices with regard to measuring the Internet from their network's perspective.

1.1 A Concrete Analog

Not all readers will be familiar with networks or measurement and analysis activities. Here is a concrete analog to provide some perspective into the challenges that always face the development of measurements of real systems.

Consider the task of setting up a home weather observatory – to measure rainfall, wind speed and current temperature. If observations are noted accurately and frequently, you can build up a picture of the weather at your home over time. But, placement of the sensors is critical. The anemometer has to be set up to be in the actual air movement (and not in an eddy generated by a fence or building or other obstacle). The rain gauge has to be unobstructed, and measured/emptied regularly (and quickly enough that rain doesn't evaporate). The thermometer, on the other hand, must be placed to read the air temperature, and not exposed to sunlight (because the thermometer will report the temperature of the probe, which will be unduly heated if left in open sunlight).

Professional meteorologists have understood this and set up equipment properly for over a hundred years. At the same time, the Internet has made it fashionable

[2] See, for example, http://www.caida.org/tools/taxonomy/measurement/

to collect and incorporate data from home weather stations as well, to give a more detailed coverage of areas that might otherwise be glossed over by professional measurement. The reliability of those measurements may be greater (a home enthusiast who understands the issues well and maintains the equipment more regularly than a remote outpost could be looked after) or less than professional meteorology readings (e.g., if the temperature probe is still reporting, but has not been checked in years, and is now encased in several layers of hornets' nest insulating it from the actual external air).

All these same things are true in measuring the Internet. Measurements reflect probes' surrounding conditions and also where they are put in the Internet, and care must be taken to ensure that the numbers recorded are an accurate reflection of the thing being measured. Whether in business operational networks or reporting from end users' home networks, the rates recorded by probes may reflect less on the condition of the network itself than the overall use of the network locally. For example, it may be difficult to distinguish, from a probe's report of round trip time to a well known website, whether the website was truly inaccessible to the end user or if the resident teenager was busy flattening the network with download torrents at the time the measurement was taken.

2 The Many Facets of Data Activities

This section discusses the different facets of activities that are undertaken in data activities, including measurements. In considering the life span of these activities – from activity design to data capture to immediate (or later) analysis – it is also important to look at the expertise of the entities responsible for the various parts of the lifecycle.

2.1 What data, and how

Generally speaking, measurements involve numbers, but beyond that there are many different ways to look at them. For the purposes of discussion of the different activities being discussed in this document, we will lay out our use of terms in this section.

When discussing data activities, there are three basic philosophies that shape the approach:

Collection is an activity that is a general gathering of data without an *a priori* hypothesis to be tested. The collecting of data may be systematic and wide-ranging, but determination of learnings is done after collection. "Big data" is about working through large quantities of accumulated data in order to find trends and correlations of interest.

Observation, for example of behavior in response to stimuli, is a useful tool in gathering data about the state of the Internet as a system. "Does this client support IPv6?" provides the opportunity to capture an observation.

Measurement generally starts with a particular hypothesis or question in mind, and collects data designed to prove or support the hypothesis, or answer the question. Measuring the round trip time of packets from one point to another in the network tells you something about the speed and effectiveness of the points' connection. If you wanted to know how the connection was doing over time, or at different times of day, you would make successive measurements accordingly with a view to determining progress.

Of course, these approaches are not mutually exclusive – a series of observations may be made and recorded as a collection, for future study and measurement. For the purposes of readability in this document, the term "measurement" is used generally to refer to all data activities through the rest of this paper.

When discussing Internet performance, two broad types of measurement activities are recognized (see RFC7799[3] for more detail, including discussion of hybrid approaches):

Active measurements: generate packet streams as part of the test procedure. This perturbs the general flow of traffic in the stream being studied.

Passive measurements: are based on observations of an actual packet stream on the network. Passive measurements do not change the nature of the flow being studied.

Each approach has strengths and weaknesses that define its utility for answering different types of questions.

Sometimes it is not possible to measure directly or completely the property of interest. In that case, two more concepts are important:

Metrics are a "stand in" of measurement for some other thing. You can say that page views are a metric of success of a particular piece of web content.

Sampling is used when it is inefficient or impractical to study every item of a kind. If there are thousands of units to study, a representative subset of a more manageable size may be examined in detail. The question of what is "representative" requires expertise both in terms of the subject of measurements and the statistical analysis of results.

[3] https://www.ietf.org/rfc/rfc7799.txt

2.2 Actors, Expertise and Uses

There are distinguishable roles in establishing, running and extracting results from any set of data about networks. In some cases, one entity will handle more than one role. In all cases, the access and expertise of the entity impacts how they carry out the role. That is, the operator of a network is in a better position to understand the network being measured (access), although they may not have as much experiment and analysis expertise as a non-affiliated researcher. The 3 key roles are:

Designer of the measurements/data collection and metrics

- Access: operator has better understanding of how the network fits together and potentially any "aberrations"; operator will have a better sense of how representative any "sampling" coverage is
- Expertise: network expertise is necessary in order to be able to frame measurements and identify metrics that work to achieve the desired answer

Capturer of the data

- Access: operator can see more detail; might collect and summarize/sanitize before sharing; nonetheless, some data is visible outside of networks and can be captured by external parties

Analyzer of the data

- Expertise: network expertise is essential for explaining the inevitable anomalies; data science expertise allows further inferences to be drawn for trending, expectations
- May be done by more than one entity at any given time, and they may have no relationship to the original data capturers
- May be done at the time of data capture, or some time later (days, months, or even years)

Since the analysis of data may occur at some considerable time later than its capture, there is always an additional element of variability in that the networks, operating practices, tools, and even the things that are being measured may change over time.

Finally, apart from the entities that cause the measurements/collection, there are users and uses of the measurements. Network operators may use measurements to identify issues in their network, researchers may use them to identify changes in the overall landscape of the Internet, and regulators may use them to create policies.

3 Issues – Internet data measurement

There are some particular challenges that need to be addressed when reviewing data collected from, or before forming any kind of measurement of the Internet.

3.1 What are the "endpoints"?

On the user's end, does the measurement start from the user's desktop computer, or the CPE[4]? While the smarts for the measurement may be running on the user's desktop, the reality is that the home network (between the desktop and the CPE) may factor negatively into any measurements. For example, a service provider might provide the network to deliver 75Mbps of data to the CPE, but the user's desktop may be connected to his home network by an old Ethernet cable – top speed 10Mbps. If the user runs a speed test from their desktop, they can't see anything faster than the 10Mbps wire delivers.

Similarly, providers are often interested in ensuring that "their network" is well-connected to popular sites, such as Facebook and YouTube. From a routing perspective, "their network" means routers and other network boxes that might be spread far and wide geographically, and have little to do with the "last mile" connection to the customer's premises. An ISP may have great connections to popular services, but if the customer is connected to the ISP by over-subscribed shared links, old copper, or other low grade links, the endpoint is not going to see advantage from that connectivity.

The same is true when talking about IPv4 and IPv6 connectivity – an ISP may support IPv6 in its core, but it takes a lot of work to update the hardware closest to the customers to ensure that each customer has IPv6 connectivity to their CPE. Then, what happens within the home network determines whether or not the desktop can actually connect to anything over IPv6.

On the server end, analogously, does the measurement reach a particular box on the network, or just one of several real or virtual servers that may be supporting a given service. For example, there is no single computer that "runs the Google website". Sometimes service instances can be distinguished by differing IP addresses, but even a single IP may support a large server farm behind the edge of the service network.

On the one hand, the user only cares about what they experience – which is everything from their desktop to the server providing the responses to their Internet activities. On the other hand, being able to break down performance by some logical "neighbourhoods" helps: separating out the home network performance from the performance within the access network, and subsequent hops to the network service.

---

[4] CPE is "customer premises equipment"; the box that connects to your ISP's access network.

3.2 What is "near"?

From Buenos Aires, Argentina to Cape Town, South Africa is 4,276 mi (6,881 km) across the globe. However, that's not how Internet traffic flows from Buenos Aires to Cape Town. Virtually (and, quite possibly, literally) all routes out of Buenos Aires to Cape Town go through Miami, US. To be quite clear, the distance from Buenos Aires to Miami is 4,405 mi (7,089 km) – already longer than the distance between the two endpoint cities – and then the distance from Miami to Cape Town is an additional 7,650 mi (12,312 km)[5].

That makes Seattle, US (2,732 mi (4,397 km) from Miami) closer to Buenos Aires in the network than Cape Town is, although that is not at all obvious from looking at a geographical map.

3.3 What is a "fixed point" on the Internet?

At a logical level, "the Google server" is a fixed point in the Internet. However, given the discussion of endpoints, above, it should be clear that there is no single Google server, or one single "Google fixed point". The same is true of other major global services. For some end users, Google and Amazon services may be "close" to each other, and for other end users that may not be true. The difference stems from the fact that each of Google and Amazon necessarily lay out their service CDN/duplication servers in ways that make sense to their own business, and not based on any global Internet service grid.

A "polestar" endpoint is one that is well known and fixed in the network – at a single IP address that is not anycast from multiple vantage points. This describes few major services today (anything popular is hosted by a CDN). Some NTP[6] servers, as general Internet infrastructure, fall into that category. Of course, services that are built out for the purpose of looking through the network towards fixed points can establish their own polestars.

3.4 Span and scope of measurements

With the variations outlined above, another challenge in setting up Internet measurements is ensuring appropriate span or scope of the measurements. For networks under your administrative control, you can manage and account for different factors, and you can install active or passive gatherers at any and all points as necessary. That gives you confidence in the measurements within your own network, but it doesn't help address the variability of any measurements that reach outside it (e.g., toward a "pole star" server). It also doesn't necessarily

---

[5] To make matters worse, most routes actually go from Miami to *some other network node*, in places such as Colorado, US or Paris, France, before connecting to Cape Town.

[6] Network Time Protocol – see https://tools.ietf.org/html/rfc5905

give information that is readily compared outside the scope of your own network.

To get global span, it is necessary to have some kind of reach into and/or through other networks, and diversity is important. The approaches discussed below outline how that has been addressed in projects to date.

3.5 Time

Time passes, things change – even if you measure the same thing from the same place, day after day, you may be getting different effects in the Internet. This is because:

- standard practices evolve and change, including things that impact basic measurement techniques (e.g., it has become common to drop or fail to respond to ICMP requests because they are a vector of denial of service attacks; previously, ICMP requests were a lightweight method of testing connectivity to an address)
- protocols change, new ones are deployed (e.g., IPv6 is carrying enough traffic that measuring IPv4 only will mean missing significant information)
- network layouts change – new routes come online, old ones get upgraded or go offline, equipment fails, etc

All of this makes it challenging to do longitudinal studies of network measurements if you don't know what may have changed in the underlying network.

4 Mechanics

This section outlines the actors and different approaches to measurements in and of the Internet. A closer examination of some examples follows in the next section.

4.1 Basic Considerations

Most of this paper discusses efforts focused on measurements that extend beyond the span of control of a single organizational administrator – i.e., across networks. It is useful to understand some of the practices and tools available for measurements within a network (with the same designer, capturer and interpreter). In the next section, some of the approaches used when the actors are different, and the span of measurement under consideration crosses boundaries into other networks (with separation of designer, capturer and

analysis). Under those circumstances, measurement activities are unable to control, or even have a detailed understanding of, the networks across which the measurements are made.

4.1.1 Collecting data

Passive data collection includes capturing and storing copies of traffic (packets) as well as logs from devices in the operator network. As computing hardware and storage scale, more data can be kept for longer time periods – e.g., all IP address assignments to end user customers for weeks or months. This is the sort of data that various governments have been accused of capturing, sifting through, and using as the basis for spying on network users[7].

For network operations purposes, these days it is important to log how much IPv6 usage there is within a given (IPv6-enabled) network, to understand uptake and usage of the protocol. More than one network operator that has enabled IPv6 is unable to provide this sort of concrete information about its success, and that impedes future development of IPv6 support.

4.1.2 Active Observation

Network operators will typically be interested in latency and other characteristics of the paths across their network. For access networks, this means from their customers' access toward well known services or egress routers through which customer traffic travels to the Internet at large. Therefore, they may engage in measurements activities that actively observe both network usage (traffic on the network) as well as network capabilities (capacity, latency, etc).

Because this is all within their own network, operators can interpret the data to understand where there are expected or unexpected issues with network connections, and address them accordingly. (E.g., detecting faulty equipment or configurations, improving connections to other networks or services, etc).

4.1.3 Sampling and points of measurement

When setting up measurements or a data collection framework, a design choice is whether to address all possible endpoints/connections or whether to focus on a subset of them -- sampling, as described above. For network operation purposes, it might be valuable to be able to review the state of *all* customers' experience of the network – from their CPE toward the access network, and through the network to an egress router. Practically speaking, however, it is difficult to get that kind of instrumentation onto CPEs for fixed-line networks. Also, measurement traffic has to be unobtrusive (lost in the noise of per-packet

[7] *N.B.*, governmental agencies are accused of using both network-operator collected data (as described here), and metadata "sniffed" from the middle of the network using their own tools.

based charging), and the impact of all of an access provider's customers "ping"ing a popular website might constitute a "denial of service" attack on the server!

Instead, measurements may be made toward a subset of possible services, and a reasonable point of measurement for the access provider may be from an aggregation point within the network – for example the DSLAM[8] or CMTS[9], each of which supports thousands or tens of thousands of customers.

4.2 Measuring across networks

Many of the same approaches are used for measuring across networks. However, without intimate knowledge of the makeup of intermediate networks, even measurements where both endpoints are well understood can yield inexplicable, or at least variable, results.

Nevertheless, the Internet remains an accessible network for measurements and observation (as compared to the telephone network, for example). Several approaches have been taken to get a measure of the global Internet.

4.2.1 Probes distributed in others' networks

Several approaches rely on distributing "probes" in remote networks, and using them to measure towards known points. Typically, one entity owns or manages all the probes, and is responsible for configuring them. Configuration includes establishing the types of measurements they perform, with what frequency, and towards which known fixed points. The owning entity is also responsible for causing the probes to be distributed around the network. Different strategies have been employed by existing probe-based systems.

Dyn Internet Intelligence[10] – has established measurement software within IXPs and access networks. From these "Vantage Points", the measurements platform can show information about Internet performance into particular markets of interest, including relative performance of different routes from the vantage point to the selected target.

RIPE Atlas[11] – uses hardware probes that are given to people around the globe to install in their own networks, so that RIPE's collecting servers can get views from as much of the globe as its ambassadors have reached. See Section 8.2 for more detail about RIPE's Atlas infrastructure.

---

[8] Digital Subscriber Line Access Multiplexer

[9] Cable Modem Termination System

[10] https://dii.dyn.com/dii/

[11] https://atlas.ripe.net/about

Netradar[12] – is mobile specific and software based. Mobile device owners are encouraged to download the software which then runs on their devices, takes measurements in the background and shares those with a centralized measurement server. See Section 8.1 for more information about Netradar.

The systems above have key characteristics in common: they are set up and managed by a single entity that is managing the distribution of the hardware or software probes and the centralized polestars or measurement servers.

The systems differ in terms of their strategies for getting probes into the network and the degree to which that is heavily controlled.

They also differ in intent – where RIPE Atlas provides basic data collection that can be used by many different analyzers for different purposes, Dyn Internet Intelligence is geared at giving companies a perspective on competitive performance of major interconnections, and Netradar's primary purpose is to build a visualization of mobile network effectiveness based on geography.

4.2.2 Opportunistic Point-measurement across the globe

A different approach, requiring less up-front deployment of infrastructure than distributing and managing hardware or software probes across the network, is to attract connections opportunistically.

APNIC Labs/Geoff Huston[13] – has developed a framework for global testing using scripts embedded in on-line ads. The script is invoked by web users' computers when the ad is displayed, and in the background they contact the APNIC test servers using URLs which illustrate various capabilities. The subsequent data analysis can then make some determinations about how many users have IPv6 capability and other such characteristics. See Section 8.3 for more detail about APNIC Labs' measurements

Speedtest.net (by OOKLA)[14] – one approach that OOKLA uses is to provide a service for users to test the speed of their connection, through a web interface. Users who have concerns are invited to this self-evident domain name's website to determine their download and upload speeds from the device they are using to access the website.

In both of these cases, the origin of measurement is determined opportunistically, by end users consciously (speedtest) or unconsciously (ad-based) connecting to the measurements server.

---

[12] https://www.netradar.org/

[13] https://labs.apnic.net/

[14] http://www.speedtest.net/ and https://www.ookla.com/

4.2.3 Specialized/preferred positioning

Some organizations are in a prime position to see particular aspects of Internet traffic and make measurements or collect data from their unique viewpoint, without having to distribute probes of any kind. Rather than having to distribute probes throughout remote networks, traffic naturally comes to them, or they can reach out to any number of remote targets.

Globally popular end-user services are in a position to see traffic from all over the globe and compare/contrast what comes to them from different networks. For example, Google, Facebook and Yahoo! were key participants in the Internet Society's World IPv6 Day and Launch[15] and provided information about the level of IPv6 traffic hitting their servers from participating access networks.

Akamai's State of the Internet report[16] – as part of its business model, Akamai's service is naturally embedded through many networks across the globe. They use these vantage points to collect and share connectivity and security "state of the Internet" reports on a quarterly basis. See Section 8.4 for more detail.

Google's IPv6 measurements[17] – by virtue of being the world's leading Internet search engine (and provider of casual video, through YouTube), Google's servers are used by large percentages of most access network customers across the globe. They have published their perspective on the growth of IPv6 deployment, in terms of what connections they see reaching their servers. See Section 8.5 for more detail.

In these cases, the measurement/data capturers have an intimate understanding of their end of the connection, and are able to leverage their position in the network's use to build pictures of some aspect of the Internet's functioning.

4.2.4 Centralized collection/contribution

Many of the examples above give specific measurements of aspects of Internet operation that are of interest operationally, or reasonably visible across the Internet.

Another class of Internet measurement starts from data analysis expertise and then seeks out actual data for study. That data may be internal to networks or services, and often requires formal agreement in order to get access to what can be considered sensitive information (for business and/or end-user privacy reasons).

---

[15] http://www.worldipv6launch.org

[16] https://www.akamai.com/us/en/our-thinking/state-of-the-internet-report/

[17] https://www.google.com/intl/en/ipv6/statistics.html

CAIDA[18] — the Center for Applied Internet Data Analysis has been the home of just such data analysis, in furtherance of general understanding of the Internet, for years. They have published many seminal studies of different aspects of the Internet, its growth and its use.

The ongoing challenge for these studies is to gain access to data necessary to fulfill a research study, when the organizing entity has ownership of none of the implicated endpoints.

4.2.5 Mandated measurement

As the Internet has grown to be a service upon which many people and businesses depend, commercial regulatory bodies around the globe have considered various benchmarks and requirements for making claims about network access and performance. To back these up, network operators have been required to support various forms of testing in and of their networks.

SamKnows[19] – SamKnows positions itself as the "global platform for internet measurement", and has been called on to provide accurate broadband performance data for consumers, governments and ISPs. When regulators have required it, access providers must support SamKnows' infrastructure to allow their customers to rate their access experience empirically.

Measuring Broadband America[20] -- United State's Federal Communications Commission mandated measurements program to study broadband performance (fixed and mobile). This work has been done in collaboration with SamKnows to implement the test infrastructure.

These approaches still suffer from many of the issues outlined in Section 3, not the least of which is because different parts of the network may be "near" or "far" to different access networks that might even be located in the same region geographically.

5 Framework Overview of Select Activities

This section looks at a number of existing measurement activities through the lens of a framework built on the concepts outlined above.

- *Activity Type:* how is the activity structured to deal with cross-network measurement.

[18] http://www.caida.org/home/

[19] https://www.samknows.com/

[20] https://www.fcc.gov/general/measuring-broadband-america

- *Design:* what entity does the design of the measurement activity/experiment
- *Capture:* what entity is in charge of the data capture (and storage)
- *Analysis:* what entity is in charge of the analysis of the captured data
- *Approach to Challenges:* how does this activity approach the challenges to Internet measurements outlined in the sections above
- *Particular Strengths and Applicability:* particular purposes for which this approach is suited, and strengths of the activity

5.1 Dyn Internet Intelligence

Activity Type: The Dyn Internet Intelligence activity is based on probes distributed (by agreement) in others' networks and network infrastructure sites – e.g., IXPs and network operator cores.

Design: The design of the network of probes, and the measurements carried out by those probes, is done by Dyn data researchers.

Capture: Data capture is orchestrated by Dyn.

Analysis: Primary analysis is done by Dyn data researchers. Access to the data system is available to other researchers by agreement with Dyn.

Approach to Challenges: Dyn has knowledge of where its probes are placed, and uses traceroute to track routes between (known) measurement points in edge networks. As such, the probes provide a sampling of network performance across edge networks (from the outside of the network, looking in).

Particular Strengths and Applicability: Dyn's probe network is globe-spanning and supports comparative analysis of IPv4 traffic patterns across time and network paths.

5.2 RIPE Atlas

RIPE Atlas is described in more detail in Section 8.2.

Activity Type: Hardware probes distributed, organically and by specific agreement, across the globe, with specific "Anchors" identified as fixed vantage points.

Design: RIPE NCC have designed the basic probes and their autonomous measurements.

Capture: RIPE NCC is responsible for capturing the data collected by probes in the RIPE Atlas network.

Analysis: RIPE NCC data scientists have designed some experiments based on the data collected by the network of Atlas probes. Individuals with "credits" in the RIPE Atlas system are also able to design their own experiments with the probes, and analyze the results themselves.

Approach to Challenges: RIPE NCC has control over the actions of the probes distributed in remote networks (within a very scoped range of options), and has identified fixed anchors against which to test from those probes.

Particular Strengths and Applicability: The Atlas probe network is globe-spanning and supports multiple forms of data collection across time and network paths. That individual probe owners can design and run their own experiments broadens the range of possible applications.

5.3 Netradar

Netradar is described in more detail in Section 8.1. The key feature of the Netradar project is that it relies on software probes that users elect to install and run on their smartphones.

Activity Type: Netradar is based on software probes, distributed through the network through voluntary download and use on users' devices (mobile smartphones).

Design: Netradar designs the capture of the data; the actual uptake and coverage evolves organically as different users elect to download and use the software.

Capture: The data is captured from passive observation of user activity and stored by Netradar.

Analysis: Analysis of the data is done by Netradar's data analysis experts, as part of the Netradar project.

Approach to Challenges: Netradar gets into remote network endpoints by encouraging end users to install their software, making the end users' devices the Netradar probes. This allows Netradar to expand beyond any limited set of operators with which they might establish agreements, but it gives them little control or direction over where their probes wind up.

Particular Strengths and Applicability: As noted above, Netradar works across all networks, and provides the point of view of the user experiencing the network.

5.4 APNIC Labs (Google Ads)

The APNIC Labs work is outlined in more depth in Section 8.3. At a high level, the work is based on adding a script to a conventional online ad campaign. The

script performs a number of pixel fetches, where the pixel is delivered from an APNIC test server. Each pixel fetch involves a conventional DNS resolution followed by a HTTP fetch, so it is possible to vary the characteristics of the DNS and the web server to elicit a particular response from the end user's browser.

Activity Type: Ads attract traffic from all over the globe to APNIC test servers, making this an "opportunistic point measurement" activity.

Design: APNIC researchers design the data capture framework.

Capture: APNIC carries out the data capture.

Analysis: APNIC researchers review the captured data and perform analyses to draw conclusions.

Approach to Challenges: APNIC provides the "fixed point" in the Internet, and analyzes the client's requests.

Particular Strengths and Applicability: These measurements reach all kinds of users in networks across the globe, without requiring specific deployment of hardware or software probes.

By leveraging the Google Ad Words infrastructure, this activity has a natural and built-in distribution mechanism, which is so ubiquitous at this point that it ensures the APNIC tests have fully global coverage of the "user perspective" of the DNS and web resources.

Tests and measurements are fairly circumscribed to things that can be determined by providing DNS responses (as the Ad Word link is resolved to the APNIC server) and basic HTML replies.

5.5 Speedtest.net / Ookla

Ookla provides the backend technology (servers) for speedtest.net, and related ISP-branded measurement services (Ookla's NetGauge service). Speedtest.net measures against fixed Ookla servers in the Internet at large, which implies crossing the access network and intervening networks to get to Speedtest.net servers. (NetGauge is an in-network version of this, and uses Ookla servers positioned as black boxes within the client network. See, for example, the University of Michigan notes to users[21]).

Activity Type: Opportunistic (soliciting users to test their network speed at an easily-remembered website)

Design: Ookla designs the nature of the tests and the layout of the polestars (ookla servers) for Speedtest.net.

[21] http://www.itcom.itd.umich.edu/backbone/netgauge/

Capture: Ookla captures the data, showing users the results from their individual tests.

Analysis: Ookla develops the user interface to show results of individual tests.

Approach to Challenges: Opportunistic point measurements – inviting end users to "test their connection". In some cases, Ookla has successfully gained regulatory requirement for ISPs to offer this testing to their customers.

Particular Strengths and Applicability: Ookla has built global recognition of their testing services through speedtest.net, becoming the *de facto* broadband capacity measurement tool in many regards.

5.6 Akamai's State of the Internet Report

The nature of Akamai's content delivery network services are such that they have a very broad linkage to many different corners of the network.

Activity Type: Akamai's measurements leverage their specialized positioning in the network. "Over the course of a quarter, we're analyzing between 100-200 Trillion content requests from over 200 countries/regions around the world."[22]

Design: Akamai determines what information to review and share from their network operations.

Capture: Akamai captures the data.

Analysis: Akamai data experts analyze the collected data to publish interesting and important results in the "State of the Internet Report"

Approach to Challenges: Preferential positioning – measurements are taken from requests to Akamai's internal platform (not their streaming infrastructure)

Particular Strengths and Applicability: Akamai's positioning enables the passive collection of content requests, at the same time as having a more intimate view of many different operators' network setups than is typical in the commercial Internet.

5.7 Google's IPv6 Measurements

Starting from its days as the world's dominant Internet search engine, Google has been in a position to see traffic from the vast majority of Internet users on a regular basis. Among the data they have collected is information about how much of the traffic they see comes to them over IPv6. More information on this activity is presented in Section 8.5.

---

[22] https://blogs.akamai.com/2015/02/state-of-the-internet-metrics-what-do-they-mean.html, October 6, 2016.

Activity Type: Data collection from a specialized position in the Internet.

Design: Google's engineers design the framework for collecting the data.

Capture: Google captures the data.

Analysis: Google staff analyze the data and produce the published reports – as part of their own IPv6 progress reports, and contributing some of the data for the World IPv6 Launch content provider view metrics.[23]

Approach to Challenges: Preferential positioning – Google sees traffic from all over the globe to its servers, and can make constructive observations comparing and contrasting aggregate traffic from different networks

Particular Strengths and Applicability: Google is well positioned to see traffic from a large swath of the Internet's users. In some regions, it will see traffic regularly from virtually every access network.

5.8 CAIDA

CAIDA is a fully-fledged data analysis institute with particular focus on, and expertise in, the Internet and the networks that constitute it. Its staff is made up of data analysis experts, including students and visiting researchers. It carries out funded research projects that variously make use of publicly-accessible data and systems, as well as working in conjunction with network operators to review detailed operational data (usually under required non-disclosure agreements).

Activity Type: Centralized study – CAIDA undertakes traditional expert data analysis studies, informed by network expertise

Design: CAIDA staff design the data research activities.

Capture: Varies – CAIDA may develop software to capture data for a study, or may make use of data captured elsewhere (e.g., by network operators)

Analysis: CAIDA staff perform the analysis/create the analysis programs.

Approach to Challenges: Rather than deploying networks of probes, CAIDA works with different organizations that are well-situated to collect and share data in order to ensure integrity of the results.

Particular Strengths and Applicability: CAIDA data analysis is the broadest, and among the most rigorously defined and executed, of the types of network measurements activities presented here.

[23] http://www.worldipv6launch.org/

5.9 SamKnows / Measuring Broadband America

SamKnows has developed hardware and software systems to measure broadband performance – for customers and for ISPs. Also, from their website, "The SamKnows internet measurement platform is actively testing in more than 30 countries on behalf of telecoms regulators."[24] The "Measuring Broadband America" program is one such effort that relies on SamKnows

Activity Type: SamKnows hardware distributed to consumers and within networks to provide fixed points against which measurements are run – e.g., through the mandated measurement project, "Measuring Broadband America"

Design: SamKnows designs the hardware and the measurements activities, potentially tuning them to the local regulatory requirements for reporting

Capture: SamKnows collects the data; reports are shared with regulatory bodies

Analysis: SamKnows experts develop the analysis tools from the data.

Approach to Challenges: Probes, distributed voluntarily and/or by regulatory requirement.

Particular Strengths and Applicability: This work is very much geared towards access network performance.

6 Application: looking at a measurement question through the framework lens

While the sections above aim to give a broader understanding of the range of worthwhile measurements activities that are underway today, it can nevertheless be confusing to the casual observer when multiple projects seem to be measuring "the same thing" and coming up with "different answers". And, if one has a specific question in mind, it can be difficult to determine where to take it or how best to approach an answer.

In this section, we'll look at one particular question as an illustration of the differences in the approaches, and how the framework can help make strengths and weaknesses clearer for a particular application.

[24] https://www.samknows.com/regulators

6.1 "How much IPv6 usage is there?"

A very timely question is "How much IPv6 usage is there?" The first thing anyone involved with measurements will say is: that's not well-formed. What do you mean by "usage"? And in what scope – globally, per country, per capita? And so begins the first stage of approximation – not asking the question that is actually on your mind, but rather asking a question for which you might be able to obtain an answer.

Setting aside those realities for a moment, and assuming that all of the projects described here support specific analysis of IPv6 traffic versus IPv4 (some are dedicated to it; some do not support IPv6 at this time), we can look broadly at how the different activities/approaches can offer parts of an answer to the question.

Activity type	What can be measured	Comment
Probes distributed in others' networks	Testing v6 connectivity from the probe to established anchors (known sites)	This would be, at best, sampling. For activities (like Dyn, and to a degree, Atlas) that have control over deployment of probes, some sense of the validity of the sampling could be achieved.
Opportunistic point measurement	Every connection can be tested to determine whether the client is responsive to responses in IPv6	This yields information about what percentage of respondents were supportive of IPv6, although it is necessarily focused on the medium of the test (e.g., http). It does not indicate whether other devces, being used for other purposes, might be able to use IPv6 from the same starting point
Specialized/preferred positioning	Can measure how much of the traffic they are privy to is over IPv6	This doesn't necessarily indicate whether there are local conditions that deter the client from using IPv6 – i.e., there is more capability than is measured
Centralized collection	Depends on what data is collected from where –	But it will not likely be global
Mandated measurement	Can mandate collection of information about IPv6	Still only as broad as the mandate

7 Conclusions

The sections above outline many different approaches to Internet measurements. It is common to start from making a virtue out of a necessity – operators need to understand and monitor connectivity and performance of links, access, and services, and then the numbers can be pulled together to provide insight into the operation of some aspect (region, service, technology) of the Internet.

As such, the different approaches to measurement and data collection are largely complementary, and the world benefits from having a broad range of measurements services and perspectives.

Nevertheless, there is still room for more measurement, better understanding of the underlying challenges and opportunities in measurement. The multi-network-spanning activities outlined above are generally started outside a given network and reaching to multiple networks. A different perspective would start from within a given network and reach across its extent to remote destinations and services. This is the operator perspective, and it can also be the Internet user's perspective.

Collaboration between network operators to gather and share data of their user's perspective would be a great step to provide coherent pictures of different aspects of the Internet's health and evolution.

8 Appendix – Deep Dive on Selected Activities

8.1 Netradar

Netradar is a smartphone application that performs a number of network performance measurements on demand, and in the background. It is available for a very wide range of smartphone platforms including iOS, Android, Windows Phone, MeeGo, Symbian, NokiaX, Jolla and Blackberry. Not all features are available on all platforms. Android provides the most flexible platform for these kind of measurement applications.

The most recent versions of Netradar do not impose any artificial measurement load on the network and instead derive all of their measurement results from passively observing user data as it is transmitted and received across the radio interface.

Netradar collects a large amount of data about the network including:

- Location of test from GPS, network or WLAN
- Download and upload speeds
- Latency
- Manufacturer, model, operating system and version
- Network and subscriber operator
- Signal strength
- Base station
- Mobile technology, such as UMTS, HSPA
- IP address and transport ports, both public and private
- Timestamp

Three measurement servers are located in Europe, North America and Asia. Although motivated by a desire to measure the capacity and coverage of mobile network operators, when the measurement client is connected via WiFi, measurements of the fixed network are generated.

By overlaying the results of many Netradar measurements on physical maps of the earth it is possible to obtain very accurate visualisations of the quality of mobile Internet access and the differences between operators by location. Detailed performance comparisons of different smartphone models have also been generated from the Netradar measurement database.

8.1.1 Limitations

The most fully featured versions of Netradar perform background measurements without requiring any user intervention and do not impose additional load on the network (potentially incurring financial penalties for the user). Some versions of the measurement application are limited by the need for the user to initiate a measurement. IPv6 is not supported on the server side, so active IPv6 measurements are not possible.

Measurements of specific regions are dependent on incentivising local smartphone users to install the Netradar application which means that measurement results are very numerous for some areas (Finland) and much less numerous for others.

8.1.2 How is it complementary to other efforts?

Netradar is focused on measurements of basic Internet throughput and latency using smartphones as a measurement client platform. It is principally motivated by the desire to develop geographic visualisations of Internet access availability and to provide comparisons between the quality and coverage of different mobile operator networks. There is no user control of the measurements run other than the timing of measurement sessions for some versions of the client. Netradar is complementary to similar initiatives that have better coverage in other regions (e.g. OpenSignal[25]) and measurement platforms that allow for user defined measurements or measurements of higher-layers of the stack.

[25] http://opensignal.com/

8.2 RIPE Atlas

RIPE Labs started the RIPE Atlas project in 2010 to build on their active measurements activities inside the RIPE region. The goal was to distribute probes throughout the RIPE region in order to have a vantage point in every AS in the RIPE service region and to every major city. The expectation is that this would involve about 50,000 probes (https://labs.ripe.net/Members/dfk/active-measurements-need-more-vantage-points). At the time of this writing, June 2016, the number of active probes is approaching 10k with a global footprint - there are probes in 182 different countries!

RIPE Atlas is a global network of probes that measure Internet connectivity and reachability. The RIPE Atlas probes are small hardware devices that are connected to Ethernet ports on a probe's host's router. The probes can conduct a number of different kinds of measurements: ping, traceroute, SSL, DNS, NTP, and http. They collect data from these measurements and relay it to the RIPE NCC where the data is aggregated with measurements from other RIPE probes.

In addition to these probes there is a smaller set (about 200) of larger probes with much larger measurement capability and that provide regional measurement targets (anchors) for various kinds of measurement activity.

The RIPE Atlas project is an excellent example of coordinating engineering, operations, and community building to build a toolset that is useful to an array of people interested in the healthy operations of the Internet, from network operators to researchers.

8.2.1 How does it work?

RIPE maintains an FAQ that describes in detail how the whole Atlas system works technically, operationally, and in terms of community interaction here:

https://atlas.ripe.net/about/faq/.

Probes run a number of measurements autonomously. These measurements are sent back to the RIPE NCC. The built-in measurements include the probe's own network configuration information (IPv4 prefix and AS#, IPv6 prefix and AS#), uptime history, RTT to the first and second hops, ping and traceroute measurements to a number of predetermined destinations, DNS queries to root DNS servers, SSL queries to a number of predetermined destinations, and few NTP and HTTP kind of queries.

In addition to these defined autonomous measurements, hosts can set up their own measurements. These measurements are limited to ping, traceroute, DNS, and SSL. These user-defined measurements have access to other probes in the RIPE Atlas for conducting these measurements. And a host's probes can also be used in experiments conducted by other RIPE Atlas hosts.

An owner of a probe, a host, can always see the measurements that the probe produces either autonomously or as part of someone else's test.

An owner of a probe builds up measurement credits, a fixed number of credits per probe each day. The hosts use these credits when they want to do their own user-defined measurements. RIPE maintains a schedule of credit cost for performing user-defined measurements.

8.2.2 Reach

As mentioned above, probes are located in 182 different countries around the world. Some countries have a limited number of probes, and in other countries, the number of deployed is quite extensive. In terms of networks covered, RIPE measures this and reports that there are probes in 6% of the IPv4 ASNs, and 11% of the IPv6 ASNs.

APNIC and RIPE produced an analysis of the coverage of the RIPE Atlas as of mid-year 2015[26]. The analysis showed not unsurprisingly that the Atlas network has really good coverage in North America and Europe, but less so elsewhere. There are therefore some very large networks in Asia for example, with very little coverage in the measurement network. This is not really a problem with the RIPE Atlas, it just raises an awareness of the population set that it covers, and gives folks who are interested in extending the network an idea about where to focus efforts to improve coverage.

8.2.3 What are some examples of what has been done with it so far

A quick search online will show the variety of studies that have been performed using the RIPE Atlas, and will show some of the diversity of approaches that have been used over time. RIPE has created a summary page of some of the academic papers written based on analysis using the RIPE Atlas network[27].

RIPE maintains a page of tools (10 different ones at the time of this writing) that can be used by anyone[28]. One example of these tools is the DNSMON tool[29]. It maintains an up-to-date view of the operations of the Internet's DNS servers - the root servers and some TLD servers. The tool provides an interface that allows a user to examine a number of different measurements (unanswered queries, response time, and relative response time), using either UDP or TCP, for a number of different DNS views (a view of all the root zones together, or singling out servers in a given root zone, or singling out a number of TLDs). The data is

---

[26] https://labs.ripe.net/Members/emileaben/improving-ripe-atlas-coverage-what-networks-are-missing

[27] https://labs.ripe.net/atlas/user-experiences/scientific-papers

[28] https://atlas.ripe.net/measurements-and-tools/tools/

[29] https://atlas.ripe.net/dnsmon/

presented over time so you can see how performance has changed over time by simply scrolling through the presented data.

Another example is the OpenIP map. The project is using crowd-sourced data to help improve GeoLocation information. The tool then allows users to see information about traceroute maps geographically. You can select the OpenIP map, then run a traceroute probe and see it depicted on a map. There are a number of uses of such a tool. One example was used to analyze how well the IXPs in Sweden are succeeding at keeping local traffic local[30].

[30] https://labs.ripe.net/Members/emileaben/measuring-ixps-with-ripe-atlas

8.3 APNIC Labs

Geoff Huston at APNIC labs set out to answer some specific questions related to the success of getting new technologies deployed in the Internet. The methodology has expanded, and he has used it to answer some related questions along the way as well[31].

The Internet Technical Community determined that IPv6 is the protocol that will succeed IPv4 and put the Internet on a stable path for growth as the address resources crucial to end-to-end communication in the Internet were exhausted for the IPv4 protocol. Geoff wanted to know how well the Internet Technical Community was doing at getting IPv6 really deployed. To answer this question, one needs to have a global footprint (like Google, see Section 8.5) or get a significantly large enough sample of end devices run your measurement code (think millions). His brilliant approach to getting a broad enough set of measurements to be useful was to embed a small set of measurement code in ads that are distributed throughout the globe and viewed by millions of Internet end users. By training the ad distribution software, he has managed to get around 7 million samples a day from a diversified population spread around the globe.

For measuring IPv6, the script in the ads gives the clients 4 unique URLs to load, a dual-stack object, an IPv4 only object, an IPv6 only object, and a URL for reporting results. By comparing the results between loading the IPv4 only and IPv6 only URLs, one can determine the number of devices that are IPv6 capable. By examining how many devices use IPv6 to retrieve the dual-stack object, one learns the number of devices that prefer IPv6 to IPv4. These results are then tabulated. The results are analyzed and presented graphically on a map of the world[32]. One can select a particular country to examine in more detail, and get stats down to the individual operator (AS) level.

For measuring DNSSEC resolution, much like for IPv6, the script gives the clients 4 unique URLs to load, a DNSSEC-validly signed DNS name, a DNSSEC-invalidly signed DNS name, an unsigned DNS name, and a URL for reporting results. Because of specific difficulties of measuring the DNS based on the diverse implementation of DNS resolution, this technique best measures whether end clients use a service that provides DNSSEC resolution rather than measuring anything about the individual resolvers that may be used.

One of the interesting things about the DNSSEC measurements is that it detects behavior that is not expected. For example, it is expected that this code from the ad runs once, but according to Geoff, in about 30% of the cases, the DNS queries are repeated at a later date. Part of the reason is resolver cache refresh, but also evident in the data is the increasing practice of DNS analysts gathering query

---

[31] IPv6: http://stats.labs.apnic.net/ipv6/; DNSSEC validation: http://stats.labs.apnic.net/dnssec

[32] http://stats.labs.apnic.net/ipv6/

logs and replaying them at a later date. It raises the larger question about the nature of DNS queries, and the to what extent the queries seen in the DNS are "authentic."

These approaches provide a clear picture of IPv6 deployment and of DNSSEC deployment around the globe. As these results have been made known others have shared that they use similar approaches on a private basis to measure things that are of interest to them for commercial reasons. One of the ongoing concerns with an approach like this is whether the providers of ads will decide that this kind of activity is not an appropriate use. Too much code running in ads starts to raise concerns about the similarity of appearance between the measurement activity and the distribution of malware. So far the measurements have been continued without issue and we can hope they continue to provide their perspective into the future.

It is interesting to contrast the results of these measurements with the measurements that Google produces of IPv6 as seen by their infrastructure, and we examine that at some length below.

8.4 Akamai State of the Internet

Akamai has a globally distributed content delivery platform that delivers more than 3 trillion Internet interactions each day. Through this content delivery platform they collect data about these interactions that they analyze to gather useful information on trends about a wide range of questions about Internet state and evolution, including broadband adoption, mobile usage, outages, attacks, and web security threats.

Akamai publishes this data in a number of forms. They publish a quarterly State of the Internet report that is available via download from their website. It is available to anyone who is interested and willing to provide an email contact address. 2016 marks the 9th year of the availability of this report, produced each quarter.

In 2014 Akamai began publishing some sets of the data usually reserved for this report through their website. They offer a nice graphical user interface to look at data they have collected about connectivity, IPv6 adoption, and client reputation. All of these are offered on a global basis, with the ability to look at data from individual economies, or regions, and trends over time.

8.4.1 The SotI Report

The Quarterly Akamai State of the Internet Report (SotI) consists of a pair of reports - one on connectivity and one on security. Both reports rely heavily on data connected from their distributed content delivery platform but the teams at Akamai also include relevant data and analysis from other sources as well. When the connectivity report discusses IPv4 address depletion information, for example, they rely on data and analysis done by Geoff Huston at APNIC. So the reports are a combination of raw data, analysis from Akamai staff, and synthesis of data and analysis from other sources.

A business purpose for Akamai's SotI reports is to extend their customer base, but there is nothing about the reports that give a hint that the data or analysis is manipulated to present a picture of the Internet that is different than what a dispassionate analysis would present. It simply happens to be the case that this kind of data and analysis aligns very well with Akamai's efforts to expand their business.

An interesting aspect of Akamai's analysis is that although it is based on Akamai's data, about their hosted content, it does really tell us something about the State of the Internet. For example, part of their analysis for IPv6 usage over time indicates that the lack of support for IPv6 in consumer electronics devices presents a barrier to the growth of adoption of IPv6 because those devices are consuming a relatively larger amount of Internet content over time. While this is based on Akamai's view, and specifics of alternate providers views might be

different, it is a large enough data set to have some applicability into the snapshot of the health of the Internet.

The Akamai SotI Connectivity report provides a summary of what Akamai sees in terms of issues around connectivity. Much of the report is devoted to observed broadband speeds, divided up by economies around the globe, and tracked year over year and quarter over quarter. There is a separate section on mobile connectivity, which also provides some interesting insight into browser use on mobile platforms. One of the more interesting sections describes outages they have observed. The cause of the outages can usually be identified.

The Akamai SotI Security report provides a summary of what Akamai sees in terms of attacks on its infrastructure in a quarter. Most of this is focused on DDoS attacks but there is a substantial section on web attacks as well. The end of the report is a summary of notifications Akamai has made to its customers about security threats throughout the past year.

The examples in this paragraph will draw on the web attack portion from the most recently available quarter at the time of this writing (Q1 '16)[33]. The report starts out with a description of the nine most relevant web attack vectors, a summary description of what each consists of, some data about source and destination geographies, and a lot of data about threats versus business sectors, all graphically presented. This data is almost always depicted as percentages, and certainly no specific customer data is revealed. In this particular report there are also a number of graphs showing how the various attacks occur by target industry sector. This section of the report concludes with a spotlight on account takeover campaigns, how some have looked in the past quarter, and an analysis of how they are implemented. So there is a variety of useful information based on Akamai's data collected from its platform, analyzed and presented for public consumption.

8.4.2 The Online Visualizations

Akamai provides a set of online visualizations of their data. All of the data is available on a by country basis, and the IPv6 visualization is also available on a by network basis.

The Connectivity Visualization provides graphically much of the same information that is published in the quarterly connectivity report. You can select whether to display average or peak connection speed, unique IPv4 addresses, or graph connectivity on the basis of 4, 10, or 15 Mbps connection speeds. The connectivity graph appears to reflect snapshots of the quarterly reports.

The IPv6 Visualization data is of a fairly continuous feed of data about IPv6 adoption. One can view IPv6 adoption on a per network basis. The networks

[33] https://www.akamai.com/us/en/our-thinking/state-of-the-internet-report/global-state-of-the-internet-security-ddos-attack-reports.jsp

are ranked by volume of requests, and the measurements are of the percentage of IPv6 compared to the total traffic from each network.

The Client Reputation Visualization graphically represents the data that Akamai keeps in its client reputation database, by individual IP address. These are then graphed on a world map. You can filter this data on a number of categories where each category shows a threat and then the source networks for those IP addresses are ranked. The categories include Web attackers, DOS attackers, Web Scanners, and Web Scrapers. When one of the categories is selected, it graphically represents the IP addresses known to have originated the attacks, and then ranks the networks that are home to those attacks with the most attack sources being at the top.

8.5 Google's IPv6 measurements

Operators of individual services measure all sorts of things that tell them something about how their network/service is working. Google has managed to share publicly something about their measurements of IPv6 from the point of view of their service. As Google began ramping up its deployment of IPv6 for all of its services, it started publishing information that its users might find useful in understanding the transition to IPv6. They included statistics about IPv6 adoption as they see it in the Google network.

Google publishes global statistics, and graphs the growth in adoption over time[34]. They also give a current snapshot of adoption on per country basis around the globe[35].

Engineers at Google documented their methodology in detail[36]. The Google methodology is much the same as that used by APNIC in their ad scripts (see Section 8.3). For a small, randomly selected set of search results, Google gives clients a small segment of javascript code to run that sends an HTTP request to particular URLs. At those URLs are servers that are part of Google's IPv6 measurement infrastructure. Those servers log the requests including the information particularly to the request that was embedded in the javascript (server, IP address, timestamp, and hash). Google collects and analyzes this data in a number of ways, and generates the graphs that we mentioned above.

[34] https://www.google.com/intl/en/ipv6/statistics.html#tab=ipv6-adoption&tab=ipv6-adoption

[35] https://www.google.com/intl/en/ipv6/statistics.html#tab=per-country-ipv6-adoption&tab=per-country-ipv6-adoption

[36] http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/36240.pdf