

October 2014

Geoff Huston

How Big is That Network?

It's often observed that the Internet was a direct outcome of the progressive liberalization of national telecommunications markets in the late twentieth century. This allowed the entry of a wave of Internet entrepreneurs into various national telecommunications markets that were historically dominated by incumbent telephone monopolies. The resultant transformation of telecommunications over the past two decades is as much a testament to the transformational power of open markets as it is to the prodigious ability of the Internet's technology base to service the ever increasing demands being made of it. But it should be recognized that open highly competitive markets are not necessarily self-sustaining, and there is a continuing tension between the desires of incumbents who see their interests being served by increasing their market share, and those of new entrants, who often have a direct interest in disrupting the position of the incumbents.

There is a balance to be struck here. It's certainly the case that Internet access can be seen as an activity that benefits from economies of scale, and the case can be made that a smaller number of larger providers leads to a more efficient market outcome in terms of prices and service quality offered to consumers. However, market dominance by a small number of providers is often an indicator of a market state that is prone to distortion by the dominant incumbents, and one that can lead to price gouging and barriers to further competition.

With that in mind, the role of market regulators and policy makers is an important one. Too heavy a regulatory hand can stultify innovation, and often leads to inefficient outcomes that are detrimental to a highly efficient and competitive national economy. On the other hand a complete *laissez faire* regulatory position can lead to the formation of monopolies and cartels, with a similar risk of poor outcomes for both consumers and the national economy. This implies that there is a careful policy path to be followed that encourages continued investment and innovation in national telecommunications-related infrastructure and services, while at the time same time avoiding the formation of market distortions and inefficiencies. What helps in this regulatory process is clear information about the state of the industry itself.

One of those pieces of information concerns the market scope of the retail Internet Service Provider sector. To put it another way, how "big" is a particular network? How many customers does it serve? Is its market share increasing or falling?

While some network operators openly publish such numbers, other network operators and owners regard such numbers as commercially sensitive information. Can we use other data sources to estimate the client population served by each network provider?

There are a number of techniques that could be used to estimate the relative size of each Service Provider from public information sources, including the number of IP addresses that are announced by the network, the number of transit customers who use the network, and so on, but the widespread use of NATs in IPv4, the varying IPv6 address plans used by IPv6 service providers, and the varying

models of use of Autonomous Systems (ASes) by retail Service Providers add some considerable uncertainty to such indirect measurement exercises.

The approach described here uses the data generated from the use of an online advertisement placement system to provide the basic input to a measurement process. In this case Google's Online Advertisement delivery network has been used in a long running non-targeted advertisement placement program which is aimed at collecting a very large set of specific user capability over an extended period. Originally we had used this approach to measure the deployment of IPv6 (<http://www.potaroo.net/ispcol/2012-07/v6report.html>). Since then we've also taken up tracking the extent to which users use DNS resolvers that perform DNSSEC-validation (<http://www.potaroo.net/ispcol/2013-07/dnssec-google.html>), and other forms of measurement of the adoption of various technologies (<http://www.potaroo.net/ispcol/2013-09/dnstep.html>). Given the utility of this approach in measuring the adoption and use of various technologies, is it also possible to use the same approach as an indirect method of estimating the number of customers connected to each network?

Part of the actions of the ad impression is to request the fetch of a very simple web object from our servers. The source IP address used to perform that fetch can be inferred to be the IP address used by that client. We can relate that address to a network by using the routing table data from the BGP routing system, mapping each end user who received an ad impression to an originating AS number.

If the advertisement placement strategy were such that each part of the Internet was targeted uniformly for ad placement irrespective of location, then these counts of advertisement impressions per origin AS would be a good indicator of the relative size of each AS in terms of the population of customers served by each AS. However, this is not the case and we have observed that advertisements are placed with different levels of relative intensity in different countries. In order to compile a uniform estimate of the customer population served by each AS we need to compensate for this. The data set used to normalize the original ad impression numbers on a country-by-country basis is the estimate of Internet Users per country, published by the ITU-T (<http://www.itu.int/net4/itu-d/icteye/>). We assume that the ad placement process is uniformly distributed within each country, and furthermore we assume that the ITU-T estimates of user population are an acceptable estimate. We can then adjust the per-country ad placement numbers to a uniform level of ad placement per country per Internet user. This normalization then allows us to estimate the relative size of each AS in terms of the estimated population of users served by each AS.

It must be stressed that this is not a census, and it is an estimate based on indirect data. The approach makes a number of critical assumptions, and therefore has a number of weaknesses.

- This approach assumes that each AS is located within a single country, and its customers are also located in that same country. While this is the case for many ASes, there are a number of instances where large retail service providers span a number of countries with a single AS. This approach does not recognize such a multi-national distribution and incorrectly attributes the assumed customer population to a single country.
- This approach also does not use secure connections to the measurement server. While care has been taken with use of unique URLs in the measurement, it still admits the use of web proxy middleware. This is particularly a problem when the web proxy is located in a different AS than the end customers that are served by the web proxy, and in such cases the measurement approach is biased towards over counting in network the network that hosts the web proxy services, and undercounting in the served client networks.
- The instrumentation in the ad is not accessible in all varieties of mobile devices, and this approach tends to undercount the customers in those service networks with high populations of mobile users.

However, with these weaknesses acknowledged, this report on the estimated customer population served by the larger ISPs in the Internet provides some visibility in the scope of various enterprises in today's Internet.

The report illustrates the significant monolithic size of Internet access providers in China, where there are a small number of service providers which extremely large client populations. A similar picture of market dominance by a major access provider appears to be evident in Mexico, Indonesia, Peru, India and Turkey. There are also large scale market coverage by a small number of providers in France, Germany, Italy. At the other end of the market spectrum is the United States, where the largest provider appears to have a market share of some 20% of consumers, and a further 30 providers have an estimate customer population of a million or more.

The report is both one that looks at the largest Internet Service Providers at the level of the entire Internet, and also a breakdown to a per-country view, that looks at the Internet Service Providers who, by virtue of the country of registration of their ASN, operate in a given country.

This report is available at <http://stats.labs.apnic.net/aspop>.

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.