Centrality in the Internet

Geoff Huston
May 2021

It's just Economics

In classical public economics, one of the roles of government is to detect, and presumably rectify, situations where the conventional operation of a market has failed. Of course, a related concern is not just the failure of a market, but the situation where the market collapses and simply ceases to exist. Perhaps markets are more than enablers simple transactions between a buyer and a seller. Karl Marx was one of the first to think about the market economy as a global entity, and its role as an arbiter of resource allocation in society. When we take this view, and start looking for potential failure points, one of the signs is that of "choke points" where real investment levels fall, and the fall is masked by a patently obvious masquerade of non-truths taking the place of data and facts. Any study of an economy involves understanding the nature of these choke points. Telecommunication services are not an isolated case but can be seen as just another instance of a choke point in the larger economy. Failure to keep it functioning efficiently and effectively can have implications across many other areas of economic activity.

A continuing source of pressure on markets is that of innovation. New products and services can stimulate greater consumer spending or realise greater efficiencies in the production of the service. Either can lead to the opportunity of greater revenues for the enterprise.

Innovation in the Internet occurs in many ways, from changes in the components of a device through to changes in the network platform itself and deployment of arrays of systems. We've seen all of these occur in rapid succession in the brief history of the Internet, from the switch to mobility, the deployment of broadband access infrastructure, the emerging picture of so-called "smart cities", and of course the digitisation of production and distribution networks in the new wave of retailing enterprises.

Much of this innovation is an open process that operates without much in the way of deterministic predictability. Governance of the overall process of innovation and evolution cannot concentrate on the innovative mechanism per se, but necessarily needs to foster the process through the support of the underlying institutional processes of research and prototyping. Public funding through subsidies in basic research may be the only way to overcome prohibitively high transaction and adaptation costs in exploring innovative opportunities.

Innovation and Transformation

Business transformation is often challenging. What we have today with the rise of content and cloud providers into dominant positions in this industry is a more complex environment that is largely opaque to external observers. What matters for consumers is their service experience, and that depends increasingly on what happens inside these content distribution clouds. As these content data network (CDN) operators terminate their private distribution networks closer to the customer edge, the role of the traditional service providers, which used to provide the

connection between services and customers, is shrinking. But as their role shrinks then we also need to bear in mind that these carriage networks were the historical focal point of monitoring, measurement and regulation. As their role shrinks so does our visibility into this digital service environment.

It is a significant challenge to understand this content economy. What services are being used, what connections are being facilitated and what profile of content traffic are they generating, and just how valuable is it?

This brings into the forefront a venerable economic topic: Is big necessarily bad? There is little doubt that the digital environment is dominated by a small number of very big enterprises. The list of the largest public companies as determined by market capitalisation includes the US enterprises Alphabet, Amazon, Facebook, Microsoft and Facebook and the Chinese enterprises Alibaba and Tencent. Admittedly there are other metrics of size that includes metrics of revenues, profits, customers and the scope and impact of a corporate enterprise, but the considerable market capitalization of these seven companies place them in the global top ten, which makes them big. But are they bad? When is an enterprise so big that failure is untenable in terms of social stability?

The global financial crisis of 2008 explored the concept of "too big to fail" in the financial world. Do we have a similar situation with some or all of these digital service enterprises?

A Brief Historical Perspective

At the start of the twentieth century a member of the US Supreme Court, Louis Brandeis, argued that big business was too big to be managed effectively in all cases. He argued that the growth of these very large enterprises that were at the extreme end of the excesses of monopolies, and their behaviours harmed competition, harmed customers and harmed further innovation. He observed that the quality of their products tended to decline, and the prices of their products tended to rise. When large companies can shape their regulatory environment, take advantage of lax regulatory oversight to take on more risk than they can manage, and transfer downside losses onto the taxpayer, we should be very concerned.

It is hard to disagree with Brandeis if this outcome is an inevitable consequence of simply being big and given the experiences of the 2008/2009 financial meltdown we could even conclude that Brandeis' observations apply to the financial sector. But do these systemic abuses of public trust in the financial sector translate to concerns in the ICT sector?

Brandeis' views did not enjoy universal acclaim. Others at the time, including President Theodore Roosevelt, felt that there were areas where there were legitimate economies of scale, and that large enterprises could achieve higher efficiencies and lower prices to consumers in the production of good and services by virtue of the volume of production. The evolution of the auto manufacturing industry in the early twentieth century, and the electricity industry both took exotic and highly expensive products and applied massive scale to the production process. The results were products that affordable by many of not all, and the impact on society was truly transformational. The US administration of the day moved to implement regulatory oversight over these corporate behemoths, but not necessarily act to dismantle their monopoly position.

Regulation?

But if the only oversight mechanism is regulation, have we have allowed the major corporate actors in the digital service sector to become too big to regulate? Any company that can set its own rules and then behave in a seemingly reckless fashion is potentially damaging to the large economy and the stability of democracy. One need only mention Facebook and elections in the same sentence to illustrate this risk of apparently reckless behaviour.

To quote Brandeis again: "We believe that no methods of regulation ever have been or can be devised to remove the menace inherent in private monopoly and overwhelming commercial power."

But if we choose to reject Brandeis' view and believe that regulation can provide the necessary protection of public interest, then it is reasonable to advance the proposition that we need to understand the activity we are attempting to regulate. Such an understanding might be elusive. In the digital networking world, we are seeing more and more data traffic go 'dark'. Content service operators are using their own transmission systems or slicing out entire wavelengths from the physical cable plant. This withdrawal of traffic from the shared public communications platform is now not only commonplace, but the limited visibility we have into this activity suggests that even today the private network traffic vastly overwhelms the volume of traffic on the public Internet, and the growth trends in the private data realm also is far greater than growth rates in the public Internet.

How can we understand what might constitute various forms of market abuse, such as dumping, deliberate efforts to distort a market, or discriminatory service provision when we have no real visibility into these private networks? Yet these private networks are important. They are driving infrastructure investment, driving innovation and indirectly driving the residual public network service. Are we willing and able to make an adequate case to expose, through various mandatory public filings, reports and measurements, the forms of use of these privately owned and operated facilities and services? Do we have regulatory power to do so considering the size of the entities we are dealing with? We've seen in the past the many national regimes have attempted to avoid the test of relative power by handing the problem to another jurisdiction. The anti-trust action against Microsoft was undertaken in Europe and even then, the result was largely unsatisfactory. Even if we might believe that greater public exposure of the traffic carried by the dark networks might be in the public interest, we might simply not have the capability to compel these networks operators to undertake such public reporting in any case.

Consolidation?

The internet has been constructed using a number of discrete activity areas, and in each area appeared to operate within a framework of competitive discipline. Not only could no single actor claim to have dominate or overwhelming presence across the entire online environment, but even in each activity sector there was no clear monopoly position by any single actor.

Carriage providers did not provide platforms, and platform providers did not provide applications or content. The process of connecting a user to a service involved a number of discrete activities and different providers. The domain name being used can from a name registrar, the DNS lookup was an interaction between DNS resolver application and a DNS server host, the IP address of the service was provided by an address registry, the credentials used for the secured connection came from a domain name certification authority, the

connection path provided by a number of carriage providers, and the content was hosted on a content delivery network, used by the content provider. All of this was constructed using standard technologies, mostly, but not exclusively defined by the IETF.

This diversity of the elements of a service is by no means unique, and the telephone service also showed a similar level of diversity. The essential difference was that in telephony the orchestration of all of these elements was performed by the telephone service operator. In the Internet it appears that there is no overarching orchestration of the delivered composite service. It would be tempting to claim that the user is now in control, but this is perhaps overreaching. Orchestration happens through the operations of markets, and it would appear that the market is undertaking the role of resource allocation. However, the user does have a distinguished role, in that it is the users' collective preference for services that drives the entire supply side of this activity.

But this is changing, and not necessarily in a good way. Services offered without cost to the user (I hesitate to use the term "free" as this is a classic two-sided market instance where the user is in fact the goods being traded to advertisers) have a major effect on user preferences. However, there is also the issue of consolidation of infrastructure services.

As an example, Alphabet not only operates an online advertising platform, but also a search engine, a mail platform, a document store, a cloud service, a public DNS resolver service, a mobile device platform, a browser, mapping services to name just a few. It appears that in this case it is one enterprise with engagement in many discrete activities. The issue with consolidation is whether these activities remain discrete activities or whether they are being consolidated into a single service.

There are two recent technology examples where this is a likely concern.

The first is the recent specification of DNS resolution over HTTPS (DOH). The DNS is a widely abused service. Attackers often leverage the DNS to misdirect users to the wrong destination and then may attempt various forms of fraud and deception. National content control systems often rely on manipulating DNS responses to make it impossible, or more realistically mildly difficult, to reach certain named service points. The DNS is often used to understand what users are doing, as every Internet transaction starts with a resolution of a name to an address. Observing an individual user's DNS queries may well be enough to profile the user to a reasonably high degree of accuracy. The IETF had its moment of epiphany in the wake of the Snowden disclosures, and undertook a concerted effort to shore up its protocol to prevent casual or even quite determined attempts at eavesdropping. The DNS has been an integral part of this effort and we have seen the specification of DNS over TLS as a way of cloaking the content of DNS queries and responses from observation. DOH looks like a very small change from DNS over TLS, as they both use very similar formats on the wire. However, DOH treats the DNS response as a web object. It can be cached. It can be pre-fetched. Presumably it can be embedded in web pages. This creates the possibility of a browser defining its own DNS environment completely independent of the platform that runs the browser, independent of the local service provider and even independent of the DNS as we know it. If the browser can consolidate name resolutions functions into the operation of the browser itself then it need not rely on a distinct name resolution system, or even a distinct name system. The browser can consolidate names and name services into its own space. Given that some 80% of all user platforms use Chrome as their browser these days then that places a huge amount of unique market power in the hands of the Chrome browser and its provider, Alphabet. DOH

may make the DNS a secret to onlookers, but once it's a secret then its beyond conventional oversight and public purview, and whether the consequent deeds in this darkened space are good or bad are effectively impossible to determine.

The second is the use of the QUIC protocol. Applications have normally followed a conventional model of using the underlying operating system for common functions. There are operating system interfaces for working with the local file store, for various network services, such as the DNS and for network connections and the protocol to service the connection, such as TCP. TCP operates with its flow control parameters in the clear, so that network operators may deploy so-called middleware to override the TCP session behaviour and impose its own view of session throughput. It can be a very effective manner of allowing the network operator to discriminate across traffic types, selectively supressing the network demands from less preferred session flows and allowing other sessions to achieve preferred performance. QUIC, originally developed by Alphabet and implemented in Chrome browsers changes all that. Chrome includes its own implementation of an end-to-end flow control protocol within the browser and speaks to its counterpart at the remote end of the connection. The way it does this is to use the IP datagram service (UDP) from the host platform and use an inner encapsulation to support an end-to-end protocol in precisely the same way that TCP is supported within IP. QUIC also protects itself from observation and manipulation by encrypting its payload. In so doing the browser is consolidating the end-to-end flow control protocol into the browser and not permitting either the host platform’s operating system or the network to have any visibility into the flow state. Like DOH, QUIC drags the end-to-end protocol into a darkened state within the browser.

Both of these are examples of a deeper and perhaps more insidious form of consolidation in the Internet than we’ve seen to date with various corporate mergers and acquisitions. Here it’s not the individual actors that are consolidating and exercising larger market power, but the components within the environment that are consolidating. Much of this is well out of normal regulatory oversight, but the results are not dissimilar to the outcomes of corporate consolidation. The result in these two cases of application consolidation is that the browser provider attains significant gains in market power.