

May 2023
Geoff Huston

The Internet as a Public Utility

I recently attended a workshop on the topic of Lessons Learned from 40 Years of the Internet, and the topic of the Internet as a Public Utility in the context of national regulatory frameworks came up. For me 40 years is just enough time to try and phrase an answer to the big policy question: Has the Internet been a success in the experiment of using market forces to act as an efficient distributor of a public good? Or has it raised more issues than it has addressed?

The term *public utility* is used to identify an organisation or entity that maintains the common infrastructure used to provide essential services to the public. Common examples of services that are distributed by public utilities are those of the distribution of water, gas, electricity, postal mail and the telephone. Conventional economic theory often distinguishes a *public utility* from a conventional supplier of goods and services into a market by distinguishing whether the good in question is a *natural monopoly*. A *natural monopoly* is a situation where the high costs of establishing the service gives an initial provider such an overwhelming natural advantage that the situation leads to the outcome of a monopoly or an oligopoly in the supply of the good or service. The conventional response to this situation is to either place the service provider into public hands, using some form of commission or similar form of public ownership, or to use of a regulatory framework that curtails the ability of the provider to exploit their monopoly position and also binds the service provision to conform to social outcomes such as equity of access, affordability or similar. As Theodore Vail, the individual behind AT&T's transformation into a national monopoly at the start of the twentieth century, put it: "For the protection of the community, of individual life and health, there are some necessities that should be provided for all at the expense of all, such as roads, [and] pure water [...] The determination between services that should be operated by the government and those which should be left to private enterprise under proper control should be governed by the degree of necessity to community as a whole as distinct from personal or individual advantage."

The late nineteenth and early twentieth centuries saw the rise of public utilities, operating in areas of public transit, postal services, water reticulation, electricity generation and transmission, and telephony services. In all cases the scale of the effort, the quantity of the capital investment required, and the extent of the investment risk appeared to exceed the capabilities of individual private sector actors. Where the sanctioning of a privately operated monopoly was insufficient incentive, then some form of public entity was called upon to operate the service. In the past fifty or so years the position of public sector utilities has been eroded, with progressive waves of deregulation and liberalisation of these public utility services, and the private sector has assumed a greater role as a result. In Australia, for example, the public entities that operated a retail bank, an airline, a serum laboratory, electricity generation and transmission, and the national telephone service operator were all passed across to the private sector. Many nations have a similar recent history.

There has been much to say about the effectiveness of such a move to pass such service provision roles from the public sector into the hands of the private sector. This shift is intended to free the public sector from the burden of raising capital to fund public infrastructure and is also intended to impose the discipline of competitive pressures on the perception of structural inefficiencies in the public utility sector. The extent to which these goals have been realised and the extent to which these moves have

resulted in an improvement in the quality of services for the public remains a subject of much debate. However, here I'd like to take a sharper focus and look at the transition of the public communications space and its former focus on telephony to the deployment of the Internet and its role as a public utility.

Deregulation of Telephony

In the 1970's there was an increasing level of frustration with the incumbent telephone company in many national communities. The internal engineering of the telephone network had been transformed using digitisation, and the telephone operator was able to reap the rewards of this internal transformation. The result was that the retail price of telephone services had drifted away from the cost of the service and were increasingly based on the capacity of the consumer to pay for the service of direct communication at a distance. The telephone company was often the largest enterprise in the national economy by both employee count and financial turnover, and by dint of charging in arrears was the nation's largest credit provider. Rather than being an efficient facilitator for other economic activities, the telephone network was becoming a drag on the economy through its imposition of excessive cost on communications. Not only were these telephone operator enterprises bloated and inefficient, but they also acted as impediments to technical innovation, and they were protected in so doing by doing of their monopoly over the provision of the service.

There was growing pressure to expose these national monopolies to the rigours of competition through deregulation of the telephone service. In the United States the sharp end of this pressure happened in the courts, starting with a private company's efforts (MCI) to install a microwave link between Chicago and St Louis to compete with AT&T's monopoly in 1966. The end result of this effort to expose chinks in AT&T's monopoly was an anti-trust settlement with AT&T where a consent decree created nine regional monopolies (the so-called "baby Bells", or Regional Bell Operating Companies (RBOCs)) and the residual long-distance carriage component was exposed to competitive entrants.

There were a number of outcomes of these actions to disassemble the regulatory framework of national telco monopolies.

- The RBOCs were highly lucrative, but they were unable to use their capital to expand into other regions within the US market in competition with any other RBOC. The natural outcome was to place pressure on other nations to deregulate their national telephone service and allow external entities to invest in the resultant competitive environment, creating a means for these RBOCs to leverage their capabilities as a telephone service provider and expand into other markets without breaking the constraints of the US consent decree. This push to deregulate national telecommunications was taken up as a pillar of economic reform, and a number of international bodies took up this agenda. (It reached ludicrous heights for me when the IMF placed pressure on Fiji, a nation of just 300,000 people, to permit a competitive phone company to enter the Fijian market.) The unintended outcome here was the appearance of US hegemony in the telecommunications sector, as these offshore investments by these regional Bell enterprises disrupted the old order of the carefully balanced positions of the national telco monopolies.
- The second consequence was the enabling of the entrance of new technologies into the communications realm at a rapid pace. Cellular radio is a good case in point where progressive refinements in the technologies required to support a hand-held mobile handset motivated new providers to enter the market. These new providers had no legacy infrastructure that they needed to protect, so they could throw themselves into aggressively marketing these new mobile services as part of their competitive differentiation. The rapid growth of the mobile sector, funded largely by private sector initiatives, appeared to confirm the view that the time for protected monopolies in telephony was over and the regulatory liberalisation of this sector enabled the rapid deployment of new technologies and services at a national level.
- The third major outcome of deregulation of the telecommunications industry was the opening up of the communications network to data networking. In its original form this was along the

lines of closed enterprise networks that used a digital interface into the internal digital switching environment of the telco to support wide-area enterprise computer networks. However, the enterprise customers of computer industry were increasingly disillusioned with the downsides of proprietary computer technologies and was increasingly motivated to use vendor-neutral solutions based on open technologies. The rise of the Unix operating system in all its various guises, and the rise of the Internet protocol are heavily rooted in this desire to move away from vendor-lock in in the computing industry and embrace open vendor-neutral technologies.

In looking at the impact of deregulation, is it useful to ask: What was the nature of the regulation and its enforcement mechanism in this area of communications services as a public utility? The common approach was to define a set of accepted behaviours and practices and define penalties to be imposed upon behaviours that were outside of this set. For example, only the telephone operator could operate telecommunications services across public land and interconnect different private properties. Another example of regulated practice was a prohibition in reselling of a service, such that when an entity leased a service from the telephone operator it could not resell access to this service to a diverse set of customers of its own. In contrast, the technical aspects of the service, particularly as they applied to the interconnection of various public carriers, were implemented through the common adoption of recommendations rather than specific regulation. This interconnection issue made sense in the context of an international connection between two national service operators, where it's unclear if either national regulatory regime has the ability to impose conditions on the other. The common adoption of recommendations in such areas neatly circumvented such thorny issues of international law. In a crude summary, it can be observed that behaviours and practices were subject to a regulatory code, while the technical nature of the provision of the service as it related to interoperability was subject to a self-imposed set of conditions intended to facilitate this interoperation between service operators.

Enter the Internet

However, it is not clear that this regulatory and operational framework has survived unscathed over the past 40 years and the evolution of the mainstream use of the public communication realm from voice to data. The early Internet was constructed using the margins of oversupply of transmission services used by voice. We used acoustic couplers so that data circuits were transformed into analogue voice signals to allow their transmission over the voice network. The trunk circuits of the Internet of the time were leased telephone transmission super-group circuits, which were essentially multiplexed bundles of individual digital voice circuits. Forty years later the transformation to an all-digital service platform is basically complete, the underlying transmission fabric is now a data plane, and voice is just one of many applications that sit on top of this common data substrate.

What is the nature of the original telephone public utility in this altered environment? Is this public utility still just a voice service, irrespective of the altered nature of the platform? Do we still need the regulatory framework of universal endpoint numbering? What about the treatment of emergency services, regulated inter-provider call handling, and all the rich set of provisions that have evolved over more than a hundred years of public telephone services? If voice is merely an application that operates above the digital substrate of the Internet, then why is this application set the subject of such an impressive set of regulatory provisions while all other Internet application frameworks, such as electronic mail, the domain name system, or the web itself, operate in an environment that appears to be largely free of regulatory impost? And while we are into these questions, it also may be useful to ask: What is the intent of a public utility regulatory framework? Is it the provision of public access to the service that is the subject of regulatory concern, or specific constraints on the nature of the service itself?

So, let's try and tackle these questions, but at the outset it should be noted that the public communications utility function has changed. In whatever form the service may be, whether its voice or a whether it's a broader digital service portfolio, the public communications service is still an essential service, but there is no single monopoly entity providing this service. It appears that the nature of this service is no longer a natural monopoly. In its place, a competitive regime operates in some form of other. In some ways this changed proposition reflects an increased level of confidence in the capabilities of markets to perform

public utility functions without introducing distortions and inefficiencies. It also reflects an increased capability in the private sector to overcome the traditional barriers to entry that would've placed the operation of such a function into the public sector or operated by a single entity as a monopoly provider. Instead of a public sector monopoly we have the operation of a market of competing private sector operators.

The nature of the regulatory framework has changed. Instead of relying on a set of service-specific constraints on the behaviour of individual service operators, we have drifted into more generic market regulation where the intent of the regulatory framework is intended to protect consumer's interests by curbing behaviours that would distort the operation of an open competitive market. Market-based regulations tend to address behaviour relating to market dominance, abuse of market power, dumping, selective pricing, and discriminatory practices in the service offering. This avoids the challenging task of attempting to regulate the nature of the service itself, while addressing an overall intent of the market's role in the provision of public services: namely the use of markets as an efficient and fair distributor of public goods and services.

In taking this path we are attempting to avoid the more vexed challenge of attempting to regulate the applications themselves. We are not trying to define the behaviours of search services, office suite tools, cloud services, or even voice services in such a generic market-based approach. We also appear to be moving away from using a special set of provisions to encompass public utilities, particularly in the communications realm. Instead, we are trying to ensure that the market for such services operates in the interests of its consumers, and that Adam Smith's *invisible hand* of competitive interests in such markets protects the public interest.

There is a residual issue lurking here which lies in the access networks we use in the wired system. While the mobile environment has been able to increase access speeds by orders of magnitude through the progressive deployment of 3G, 4G and then 5G base stations and client devices, the wired network has proved to be quite challenging in the suburban consumer sector. Rewiring the access network with fibre is a technically superior approach, but also the most expensive. A number of national and local public sectors have funded this rewiring through a broadband public utility, but the capital costs that the public sector has to bear in such a program are forbidding. Other national regimes have allowed the operation of temporary monopolies in the access network to allow the private sector provider to recoup part of the initial capital outlay incurred in lifting the speed of the last mile access network to some form of broadband access speed. The outcomes appears to be highly variable. Some national regimes have a residual public utility service in the area of broadband access, while other regimes have left this to the private sector in its entirety.

The issue of what constitutes a base universal service that should be extended to all users remains somewhat unclear in many national regimes. What is the basic rate of a broadband access system that can be useful? For example, are we now considering a base service that includes high-definition streaming video? Or even multiple such streams operating simultaneously? Is there a regulated price point for broadband access? Is any such base price point applied everywhere? Or does the cost to the user vary with the location (such as the difference between rural and remote locations and dense urban contexts)? Is every competitive access provider required to offer their access services to all users in all locales? Or is a more selective approach viable? In the latter case the question arises as to whether there is a residual universal service obligation where the universal service operator is subsidised for the provision of universal service access (often through by financial imposts levied from competitive access service operators who operate within chosen locales).

I'd like to look a little deeper at the issue of the regulatory framework and the public utility role in the broader content of the Internet as a public communications platform, so we'll leave the issues of the provision of broadband digital access networks here and return to the broader considerations of the Internet as a digital service platform for public use and the regulatory challenges this has created.

Is market-based regulation working as intended?

Rather than try to address all these tricky service-specific questions were we to try and apply a regulatory constraint on the behaviour of individual applications, we've evidently decided to sweep digital applications into the adoption of general regulatory measures that apply to the operations of markets in the provision of public services. This appears to be an easier, and hopefully more effective path. But has it worked as intended?

At first glance, it seems to have worked well. Very well. In today's Internet, voice is just another application. Moreover, it's an application that can be used without paying an incremental use fee. Telephone operator companies still exist, but their revenue base does not lie in charging by the minute for voice calls any more. For revenue they have largely shifted their focus to their holdings in exclusive use radio spectrum licenses, and they charge a premium for mobile data services. Even in that area competitive pressures are eroding margins for mobile operators, so the longer-term privileged position of the telephone monopoly operators of forty years ago is all but over. Data services have continued a downward movement in retail cost, both in the mobile and broadband sectors.

The driver behind this continual cost reduction can be attributed to the continued operation of Moore's Law in silicon chip manufacture, where there continues to be an increasing yield of chips in term of gate density and an associated drop in the unit cost of processing and memory. The improved processing capacity increases the capability of digital signal processors which, in turn, increases the carrying capacity of fibre cables. For perhaps the first time in the entire history of the telecommunications sector we are operating in an environment of resource abundance rather than scarcity. We are no longer forced to use price as a rationing mechanism to distribute access to a common resource across overwhelming competing demands. For example, we are currently capable of reacting to hostile traffic attacks by simply using ever increasing capacity in service provisioning and simply absorbing the attack.

The computer industry has operated in a largely deregulated mode since its commercialisation in the 1950s. The most striking aspect of this industry lies in its ability to take technical innovation and bring it to bear on products and services at extraordinary speed. The transition from large scale mainframe computers to personal computers with the corresponding need to massively scale up production was mirrored by the speed of the second wave from personal computers to pocket devices, which again required a scaling up of the production capacity by orders of magnitude. One of the major positive aspects of deregulation of the communications realm has been to bring the computer industry's remarkable agility to bear on what had been an historically staid communications service realm.

In such an environment of agility, efficiency, and abundance, are public utility measures of any value at all to the public communications sector? Should we proclaim that the Internet and the underlying computing industry are doing their job effectively and we no longer need to apply special public utility measures to this activity?

Such a move would be entirely premature in my opinion. The rise of the Internet and its permeation into many other aspects of social and economic activity has not been associated with a rise in the level of vibrant competition in the digital service roles in meeting users' needs. Quite the opposite. The private capital venture markets have funded the expansion of the digital world with a number of quite disturbing mantras to guide their actions. Competition, they say, is for losers. "If you want to create and capture lasting value, look to build a monopoly", writes Peter Thiel, a long-term venture capital funder of digital ventures. A successful digital enterprise folds into its own scope of operations everything for which it is critically dependent, and where that is not possible it attempts to push the external dependency into a commodity service function to destroy any intrinsic value in the role. The ephemeral nature of digital goods and services has allowed monopolies to form and flourish without triggering any clear warning signals along the way as the accumulation of digital heft in a market is largely invisible while the accumulation is underway.

The accumulation of wealth by these digital giants, and more importantly the accumulation of data, create an over-arching level of dominance across much of the digital space. Their industrial age forebears in the gilded age in the late nineteenth century managed to not only dominate their time but extend that domination well into the future. Many of these guided age giants, including Standard Oil, General Electric and JP Morgan are still major corporate entities today, more than a century later. No doubt similar ambitions can be found to today's digital behemoths.

Is this lopsided digital environment, where a handful of enterprises are extremely dominant, an instance of a failure of the regulatory framework? It very well may be, but the real question is: What regulatory measures could have been adopted back in the day that would've curbed the formation of these digital oligarchs? The collection of user profile data to aid in the selling of product had been a feature of the media industry for many decades, and the mantra of "know your customer" had been extended into the retail sector and then into the advertising sector. It required an uncommon level of prescience to put the pieces together and forecast that the digital world would take this collection of user data and scale it up by many orders of magnitude, and at the same time increase the acuity of this profile data to the level of individual consumers. We had regarded personal privacy as a natural attribute of our society and simply had not seen that the application of industrial scale computing technology would be deployed on profile data gathering and analysis. By the time we wanted to react to this emerging issue, we had found that regulatory framework regarding each individual's rights to personal privacy was just inadequate and the political lobbying by these data gatherers so effective that the prospects of a stronger set of regulations in this space was a forlorn hope. We have ended up relying on the same data gatherers to devise their own codes of practice in this space, which is about the same as asking the local foxes to look after the chickens in the hen house!

What have we learned?

What have we really learned from the past 40 years in the area of regulatory governance of the public communications space?

We have certainly learned that using an historical regulatory framework that suppresses innovation and service evolution is largely a wasted effort.

So how should we regulate this Internet space? In considering this question, we can go back more than a hundred years and look at the same debate in the context of the Gilded Age giants of the late nineteenth century in the United States. At that one of the themes of the debate was whether it was the anti-competitive actions of the entity that should be constrained by regulation or whether the root cause was the over-arching size of the entity in the market, in which case the entity should be disassembled to remedy the problem. Rephrasing this dichotomy into current terms the question becomes whether these digital giants are abusing their dominant positions in many markets, or whether the overbearing size of the digital giants sustains the accumulation of vast data repositories, giving them an unparalleled view of both consumers and the other actors in their area of activity. One response is to impose codes of behaviour and greater levels of external scrutiny and oversight, and greater penalties for infringements of such codes (The European GDPR measures come to mind as an example of such a response). The other option is to respond to the accretion of activities, resources and data by these giants cleaving them apart through anti-trust actions.

Of course, this raises the concern that either interventionist action would have its own set of unintended consequences. In attempting to curb the worst excesses of this surveillance economy by taking direct action against implicit re-selling of user profile data to advertisers, there is the risk of pushing these entities to adopt a more transactional view of the provision of services to their user base and motivate the service providers to take a more predatory position with respect to their users.

If both extremes in this range of potential regulatory responses are considered to be infeasible in practical terms, then is doing absolutely nothing at this point in time a sensible response? If we simply wait out this period, then we are anticipating that the next cycle of technology will push today's digital giants out

of the spotlight. It has happened a number of times already in this space, with Microsoft shouldering aside IBM's dominance of the computer market, and then Google's direct assault on Microsoft's dominant position in office productivity software. Of course, such a level of faith in the future to disrupt the present-day dominant enterprises could be misplaced, and there is a distinct risk that we would be replacing today's digital giants with even larger digital behemoths with even more disturbing implications. At this point I could invoke the spectre of the evolution of AI models that combine highly capable mimicry and synthesis with deep reasoning as our next big challenge, after we've done with the perils of comprehensive digital surveillance. But perhaps that's a scary story for a different time!

Disclaimer

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

Author

Geoff Huston AM, B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net