Net Neutrality: The Technical Side of the Debate: A White Paper

Jon Crowcroft
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK
jon.crowcroft@cl.cam.ac.uk

ABSTRACT

Network Neutrality is the subject of much current debate. In this white paper I try to find the signal in the noise by taking a largely technical look at various definitions of network neutrality and the feasibility and complexity of implementing systems that support those ideas.

First off, there are a lot of emotional terms used to describe various aspects of what makes the melting pot of the neutrality debate. For example, censorship or black-holing (where route filtering, fire-walling and port blocking might say what is happening in less insightful way); free-riding is often bandied about to describe the business of making money on the net (rather than overlay service provision); monopolistic tendencies, instead of the natural inclination of an organisation that owns a lot of kit that they've sunk capital into, to want to make revenue from it!

The paper describes the basic realities of the net, which has never been a level playing field for many accidental and some deliberate reasons, and then looks at the future evolution of IP (and lower level) services; the evolution of overlay services, and the evolution of the structure of the ISP business space (access, core and other); finally, I appeal to simple minded economic and regulatory arguments to ask whether there is any case at all for *special pleading* for the Internet as a special case, different from other services, or utilities.

Mutatis mutandis

Categories and Subject Descriptors

A [**.**]: 2—*Reference*, C.2.1 [Packet-switching networks],C.2.4 [Distributed applications], D1.3 [Distributed Programming], D4.1 [Scheduling], D4.4 [Network Communication], D.4.8 [Stochastic Analysis], E.1 [Data Structures], G.1.6 [Constrained Optimization]

General Terms

General Terms: Performance, Design

Keywords

Data Communications, Review

1. INTRODUCTION

Let me try to illustrate the complexity and subtlety of the debate with a few, real stories from the last ten years of Internet Experience, each of which is chosen because it captures several facets of the problem space at once.

Priority Rights Like many other people, I have 8Mbps Internet Access through an unbundled DSL broadband provider, which I share throughout my house using a \$50 router to provide 10/100 Ethernet and Wireless access to a server and the family's laptops and media centers. I don't secure the net with WEP keys and access control, since I use secure end systems with host firewalls and virus checkers etc etc, although the router runs some useful filters to lower the background nonsense. When my phone line went down for 3 weeks earlier this year, my kids found 3 neighbours with open WiFi access to their broadband lines (luckily all still working – indeed 1 cable, and 2 different DSL providers, one bundled, and one unbundled). We asked them if it was OK to use their net (this is a UK legal requirement since recent precedents in *unauthorised* access to *open* WiFi nets being deemed an offence under the Misuse of Computers Act). My neighbours said "sure", but tellingly also admitted that this was because our usage would have no impact on their usage since they all used routers which implemented priority queues (see, for example, `http://openwrt.org/`). While their nets were open, they had all independently discovered that it was possible to set higher forwarding priority for their own packets than everyone else, thus being socially friendly at the same time as not giving up any resource they paid for.

We can unpack at least three lessons from this tale: firstly, it is literally child's play to build community wireless networks; secondly, it doesn't take technical experts to deploy priority services; thirdly, cooperation and selfishness are not necessarily orthogonal.

Content Re-Distribution In the mid 1990s, the UK Academic Network provider, UKERNA, ran an experiment in usage charging for International Traffic. The goal was twofold: firstly, the charges might trickle down to real users and create a disincentive to misuse or similar carelessness in moving large amounts of data around unnecessarily; secondly, the goal was to raise more revenue to pay for upgrades to the International Links.

It is one important lesson that the second of these goals was far more successful than the first, however, another piece of the story is interesting. A number of national research networks in Europe provided very large web proxy caches to create a positive alternative (lower latency, potentially higher throughput etc). At the same time, UKERNA allowed free access to web caches (very sensibly). However, other European countries (who had *not* implemented the charging disincentive side of the story) soon noticed that users in some UK universities were using their caches (especially in well provisioned areas with good UK connectivity such as Scandinavia). They rapidly introduced first priorities (better access for IP sources in their own ASs), and eventually blocking of IP addresses outside their own networks.

Data and Digital Mobile Phones I have a cell phone. It uses around 14kbps to carry voice, and provides a global service which is extremely pervasive and affordable. Indeed, there are more cell phones than Internet Hosts (2.5 billion active mobile phone numbers in the world at the time of writing). My cell phone provides data (GPRS, EDGE and 3G as it happens). The 3G service runs at around 384kbps in the UK, and seems to have pretty low latency – I do not know the architecture of the backhaul network once the wireless segment of a route is terminated, but it seems to support pretty close to zero loss. I can run Skype or any vanilla VoIP system on this fairly easily. However, the volume and time tariff of the data service is set such that a normal pattern of voice calls made over it would cost more than the GSM service. This is fairly surreal (in fact, usually when I read my e-mail via my phone, I "dial-up" over GSM as it is cheaper), but you can see that there are powerful reasons for the cellular network providers to stay in this regime for a while, or else have to explain a massive loss of revenue to their shareholders. Maybe? If not, what is their replacement source of revenue?

They key lesson here is that *legacy* service providers resist the pressure to become merely bit pipes.

Digital TV and Fibre Another UK example is that, as part of the agreement its privatisation, BT was explicitly *not* allowed to carry broadcast TV. At one point fairly recently they offered to put fiber to every home and office in the UK if this rule was relaxed, but the government regulator rejected this offer. The argument was that it might create a monopoly of Internet TV, despite the fact that digital TV was until recently a near (non-IP) monopoly in the UK, without the benefit of fiber everywhere. Indeed, the UK's current near monopoly commercial digital TV provider owns the box in your house, the channel and a large part of the content too, whereas an Internet alternative would create a *pair* of vertical bundles which would compete, but would also allow arbitrary bandwidth access to all the other Internet TV content in the world.

There are several lessons to tease out here: infrastructure and bundles are incommensurable; secondly, the timescales for regulation may often be wrong (both too short and too long), and need constant revision, possibly requiring smart regulated markets rather than fixed franchises (as with pollution credit, and 3G spectrum resale arguments).

Preferential Treatment of Customers The (possibly anecdotal) story of why Strowger invented the automatic telephone exchange is famous, but worth repeating in this context. Strowger ran a funeral business. He suspected that a rival business in town was getting more customers because the telephone operator was the sister of the owner of the rival business, and when asked to connect a bereaved client to a funeral service, would, of course, choose her brother's phone line rather than Strowger's. He automated the bias out of the system.

One lesson here is that a biased service may be entirely innocent at one level, but cause problems at another.

Each of these stories illustrates a different aspect of the neutrality argument, whether it is the basic IP service, an overlay service, a provider who owns a last mile and a distinct end-to-end legacy service like voice, and the bundling of content and service and network facilities. All of these arguments ran before the Internet existed, but the generality of the Internet allows for *convergence*, and the debate we are seeing is really just another result of the fallout of the final reality of convergence[1].

In the rest of this paper I look at these aspects of the debate: the IP Service, including current and future evolution; Access Networks; and Content and Bundling. Finally, I discuss economic aspects briefly, and draw some conclusions.

2. IP SERVICE: HISTORY AND EVOLUTION

The Internet provides a Universal Service[2], just as the Public Switched Telephone System provides a Universal Service. However, the service that IP provides is merely connectivity at the network layer, whereas the PSTN (in analogue, digital and wireless forms) also specifies delay bounds, minimum capacity, and availability. Parts of the Internet are engineered to provide resources (link speeds) that exceed the needs of new applications, but it is not part of the service or protocol specification, nor is there a forum for agreeing what might be such a part of a service, since *services and protocols are dealt with by different communities*, unlike the combination offered by the ITU.

Recently, there have been a number of concerns about the future evolution of what I might term the Universal IP Service. It has become apparent that for a variety of reasons, the core *connectivity* service is not as Universal as once thought. This is usually to do with security concerns (e.g. about appropriate content or activities), where there are differences in different organisations about what is appropriate. Sometimes the IP level connectivity is there, but higher level mechanisms prevent access to sites (e.g. firewalls blocking transport ports). Sometimes, for performance reasons, such filtering is more easily done simply by blocking IP addresses. Sites sourcing much spam or DDoS attacks, or other malicious traffic may be black-holed by providers, despite the fact that the cause might involve exploits of an "innocent" users' vulnerable machines. This leads to a great deal of work in ISPs' call centers, handling requests from these users to be "re-connected". In the PSTN, it is much harder for a

[1] "Convergence" is a telecom term for the merging of telephony, television and data services onto a single infrastructure.

[2] "Universal Service" is an ITU term for the minimum set of functions that all public telephony providers must provide within and between their networks.

provider to disconnect a user or exchange unilaterally, without due warning, due to the legal obligations on them to provide, at least, emergency phone call services.

2.1 IP Service: History

The Internet has been around for 30 years, and the core best effort IP service interconnects around 1 billion devices in the world today. There are a large number of corporate, private and government Intranets, as well as a significant number of interconnected commercial Internet Service Providers, which inter-work at the lowest common denominator level, which is to say that connectionless Internet datagram packets can be routed from end system to end system through a collection of intra-domain and inter-domain routers (and associated firewalls, NATs and other devices).

From the very first, the optimisations in routers driven by the statelessness of the IP level has meant that it is hard to introduce enhancements to the core service model such as Quality of Service. The scale of the system is such that now that any alteration to the model must retain backwards compatibility for a significant period. Examples of enhancements such as multicast, Mobile IP and IPv6 succeed or fail on the ability of the new feature to work in parallel with the existing services.

The core service model supports a very simple definition of performance, which is to say that there is none. Instead, it is implicit in provisioning in each segment of the network (and varies with time, and with source/destination), what basic performance one might see. Over the evolution of the net, efforts concentrated on core connectivity, and a low cost method to allow applications to co-exist in a shared resource, based on congestion avoidance and control by end systems. The complexity of these overall developments generally shows up in two places:

1. Below IP, mapping packets onto various link technologies.

2. Above IP, in transport (TCP, RTP/UDP) and application (HTTP, P2P) protocols.

As well as this, a number of practical middle box services have appeared which intermediate network access. Historically, these go back to gateways between different protocol worlds. Now they are used to provide programmatic ways of controlling access between heterogeneous segments of the net for a variety of reasons.

Technology moves on, and as it does, it diffuses through the research, academic, and then commercial networks. This process of innovation is continuous, and has an impact on services. Prioritisation of new over old is not a common deployment technique.

One of the key areas of evolution in terms of differences between ISPs has been that of SLAs. Many ISPs offer statistical guarantees of performance (above and beyond a simple bland statement of "Best Effort"). For example, zero packet loss is offered by some tier-1 ISPs, while 95th percentile delay guarantees are given by others. Few offer this to traffic transiting to other ISPs, so already in the last 7 years or so, I see a variation. Inevitably, there is a tendency, under competition, to "level up" to the better offerings, as tools for provisioning and traffic engineering become more widely available, and as capacity prices have continued to fall, making the feasibility of pure statistical multiplexing based guarantees easier to achieve (even for VoIP traffic). However, there is little evidence of anyone using the same techniques to "level down" – indeed, the sheer numbers of ISPs (e.g. 300+ in the UK alone peering at the LINX) means that any such effort is doomed to lose customers to competitors quickly.

This type of economic dynamic (introduction of new services piecemeal, followed by widespread adoption) seems to have been missed by many commentators on net neutrality.

Next, I describe some of the current realities of the net, which has never been a level playing field for many accidental and some deliberate reasons.

The Internet was never really a level playing field. Recently, many areas of the Internet have tilted so far as to stress the system a little, but the idea that the network is innately fair (for whatever definition of fairness you wish to choose, whether proportional, max/min, or other), is fairly bogus. Some examples of accidental favouritism, effectively wired into the Internet Protocol Suite, include:

End-to-end service Most traditional Internet applications run on TCP. The throughput you get from TCP depends crucially on (at least) four constraints. Firstly, your bottleneck capacity may be your (or the far end's) link speed or system I/O capacity. Secondly, the throughput is limited by any other user's TCP flows traversing a shared bottleneck. Thirdly, your capacity is a function of advertised window size, MSS and so on. Finally, and most arbitrarily, your capacity is a function of the round trip time and packet loss probability on a link (the latter may simply be a function of the other users' load, but not always). The dependence on round trip time is inverse: so the further you are from a sender, the less capacity you get than other people.

Inter-domain Routing The Internet is rich in numbers of service providers. To reach a site on another service provider's net, your traffic must traverse at least 1 border router. This introduces additional delay, but also, if the path (as often is the case) traverses multiple ISPs, it maybe that the return path is not the same. This has a different effect on your traffic than others (e.g. users in the far end's domain, or at different intermediate ISPs). This is not directly intentional – it is a side effect of the business relationships of ISPs: they are not targeting *you* personally.

NATs We are all too well aware of the whole midbox debate, so I will not rehearse it here. However, I would say that anyone behind a NAT is not providing a service, so they are not on a level playing field.

Firewalls I guess the division of the Internet into those places reachable by a first TCP SYN packet to port X, and those not, is another balkanisation. Of course, the network can always route around damage, but the net-cost of having to implement the superset of damage-avoidance rules may make it infeasible for most mortal users.

Proxies Caches, as I explained above, are put there to distribute load, and improve users' experience in terms of download delay (in fact, simply a precursor to p2p and torrent ideas). However, caches (and many replication systems) implement rules to control the performance seen by the overall set of users. Indeed, many popular news and software distribution websites now implement *admission control* algorithms to control the perceived performance. The net effect is that users during a "slashdot" event, see messages that are analogous to call blocking in under-provisioned (or overloaded, e.g. during flash crowd) telephone networks.

What happens when favouritism, or differentiation, is made a network layer first class service? Let us look at that next.

2.2 IP Service: Evolution

The basic IP service has no real definition (well, there is a definition of Best Effort as part of the PHBs, but this doesn't define an end to end service. However, many ISPs and some Internet Exchange Points define Service Level Agreements (SLAs), which derive from related thinking in Telecom networks of yore. In circuit markets, you buy facilities to connect points with certain characteristics. For example:

Isolation My traffic is not impacted at all by yours.

Protection My circuit is backed up to the nth degree by failover paths.

Throughput I get the capacity I pay for, point-to-point (see later, end-to-end)

Delay Whatever pattern of packet timings I send with is preserved (c.f. jitter) at the far end, and I see non time-varying delay

The generality of the Internet has led away from a purely TCP (and associated Best Effort tolerant) based applications. Now we have a very significant and growing number of users of network applications such as VoIP, IPTV, videoconferencing and networked games. Note that each of these applications has user expectations associated both with performance, *and* with being charged. We are not averse to paying for phone calls, for watching some TV programmes, for being charged a lot for (legacy ISDN) videoconferencing, or for paying to be in a game (or even for objects in the game). Internet users now expect to see some of the properties of circuits.

A number of technologies have emerged to support services that look a bit like circuits in the Internet, although most are only deployed within a single ISP, and often, mainly for corporate customers so far.

Differentiation The IETF community has been struggling with a variety of concepts for introducing Quality of Service mechanisms to the Internet for 15 years or more[3]. Finally, we have a simple, but effective technique, which some ISPs have deployed, principally (as far as this writer is aware) to support the legacy services on IP such as VPNs and VoIP backbones for a national telephony service. However, these are good proofs of concept and there are plenty of customers for a more dynamic service enhancement.

Provisioning/TE Any technology for QoS assurance of any kind is deployed coupled with a detailed knowledge of the topology of the network, the workload and traffic matrix, and its variation over the day, and a detailed model of all source behaviours. These are then fed into some provisioning model which also contains the traffic engineering mechanisms that the ISP is deploying. This could be based on a tool such as network algebra (c.f. Cruz, le Boudec et al, work in this area) or an emulation or simulation that is used to compute whether a new user or service can be admitted. The timescales of this are rather different than what was used in traditional admission control for telephone calls, but that is because we have more headroom in today's networks, and we have better tools to comprehend aggregate behaviours in the core.

Note that in the previous discussion I used the word *"core"*. Of course, the Internet as a whole has no core. It is built out of many ASs by many ISPs. Each may have a core network and may use intra-domain provisioning, but the case for interdomain QoS has yet to be solved.

In the broader global scope, several proposals have been around for a while, including the old Internet 2 and Abilene idea of Brokers, extensions to BGP, and even the use of int-serv/RSVP to allocate inter-domain slices within which differentiation is done. All of this is subject of future work (or breakthroughs!). The inter-domain space is largely "valley free", which means that paths traverse up and down the ISPs in a hierarchy of tiers. So in some sense, one could imagine a "core" at the AS level – however, tools to reason about performance at this level are not yet available even in research.

In any case, not all customers are equal:

Horizontal relationships As has been observed by BGP experts, the inter-domain routing space has evolved to support a number of business models relating the ISPs either side of a border (and by implication, further afield). Usually, the dominant relations are termed: customer/provider and peering. There are other more complex ones, rarely published.

Vertical Relationships Application Service Providers and Content Service Providers may have a wholesale relationship with ISPs. For example, a typical content acceleration service has to acquire rack space in data centers, typically co-located with higher tier ISPs' POPs. The price for the rack space plus capacity (and other hidden benefits such as secure power supplies, reliable air conditioning, anti-DDoS systems etc etc) may be priced in some aggregate way. Indeed, buying redundant Internet access for reliability as well as performance (load balancing and lower latency access by having multiple sites around the world) may attract some bulk discount. However, such agreements are rarely, if ever, published.

Piecewise deployments can be seen as potentially applicable to other changes to the core IP service model, such as:

Security As hinted above, some ISPs provide firewall services in addition to NATs to protect users from unwanted access. In some cases this may go further

[3] The IETF has been steadily tracked by the research community working on better signalling, admission control, and fair queuing algorithms, as well as simplifications of models that allow for ideas such as core-stateless fair queueing, and measurement- and probe-based admission control. Fairly recently, the IETF also was directly trying to address provisioning of priority services for emergency use of the Internet.

and include black-holing of sources of SPAM, and of DDoS attacks (sometimes only on request). The idea of providing more sophisticated security services (e.g. signed/authenticated and approved system distribution for sites) is already common place in private networks, and one can imagine ISPs requiring (and providing) approved systems and system patches to remove vulnerabilities (especially ones demonstrated to allow other sites to 0wn and misuse a customer's machines).

Mobility The last few years has seen the emergence of Wireless ISPs (WISPs), offering pay-per-use wireless hotspots. Quite a few of these provide *roaming* arrangements, whereby credit on one service can be used on another.

Multicast IPTV is starting to take off with content problems being resolved, and net performance finally exceeding the threshold necessary to offer reasonable quality realtime TV. However, some live events may be of primary interest to large groups, and we may see pay-per-use IP multicast finally take off. On the other hand, P2P TV is also emerging as a model which doesn't stretch the ISP at all, but meets the requirements provided enough up-link capacity is available from participating customers. The ISP might in either case, broker the content and rights.

The key argument in the neutrality debate about differentiation lies in the question: does one level up or down? When offering a new service with higher performance, clearly any serious business will price and provision things so that the lower tariff attracts lower performance. But what is the trend? Is the additional income used to provide more capacity so that the "poor" do better, while the "rich" do even better? Or is the capacity shared in a different way, so the rich win at the "expense" of the poor? The jury is out, but you can bet your life it is a zero sum game at any instant.

3. ACCESS NETWORKS

An entirely different version of the net neutrality debate concerns the access network. Here, there is some evidence that we are re-playing the arguments that led to the divestiture of AT&T all those years ago, and that the competition in local loop in different parts of the world varies enormously, and so one has to be very careful whether this is really a general debate, or one that reflects lack of competition in the local loop. As I hinted in the introduction, this sort of debate can also be held concerning wide area wireless (cellular) access, and has been noted in the previous section, it could also apply to WiFi pay-per-use hotspots.

Legacy services with vertical bundles (PSTN, with phone line which happens also to be the last mile access for IP, same for cable TV) are crucial to many users of the Internet. The operators who own these local loops are quite heavily regulated in many parts of the world, in terms of telephony, and in terms of allowing competition access to the *exchange* (or *head end* in the cable case) end of the lines. Whether the line/access are bundled or unbundled is crucial.

The costs associated with maintaining 100s of millions of phone lines are quite high. The cost of deploying ever increasing speed DSL kit at the exchange ends is also high, and many incumbents would like to offset this by increasing charges. The cost of providing an alternative is also high, although fixed wireless broadband is a possibility looming on the horizon, as is the replacement of the entire access net with fiber in highly developed parts of the world such as Korea and Japan.

However, if the operator that owns the last mile also still owns significant long haul networks, and wishes to capitalise on both, there is a strong incentive to provide some modest level of *walled garden*, by offering improved access link speed, provided some bundle of higher levels is subscribed to. This is entirely familiar to telephone users, digital TV users, and cellular telephone subscribers[4].

The real question here is whether the last mile needs to be regulated, for example when there is a near monopoly *and* the provider behaves monopolistically. If that occurs, regulation can ensure performance and bundles are transparently measurable and priced, and alternatives (or potential alternatives) are evaluated on a level playing field by regulators and understood by consumers. This is one area where it seems to me the current regulatory frameworks (especially where this writer works, in the UK) have many of the right components, and there may not need to be any new definitions of neutrality. The Internet is just another service.

4. CONTENT AND BUNDLING - OVERLAY SERVICES

One of the grand challenges to net neutrality was the subject of many of the (US) companies representations to governments, and that was the threatened actions by some ISPs to block or lower performance to certain applications *en masse*. The statements made by some ISPs implied that overlay services that are crucial to many users such as VoIP and Web Search engines (specific examples of course being Skype and Google, but no doubt they were just the most visible examples) were *free riding*.

This emotive term was used almost certainly by marketing people, since it has connotations of illegal file sharing and piracy. However, most large scale overlay systems buy significant quantities of Internet access at very high speed, and (more importantly) buy it from many ISPs in data centers in POPs (as discussed in the previous section) so that they can offer a global application service. In other words, they are not "free-riding" for free at all. Nevertheless they make a lot of money, and ISPs that only offer IP packet transport are unsurprisingly jealous of that revenue.

Let us think about that for a bit because it is really quite amusing. An ISP is not forbidden from also being a content service provider (modulo certain special cases such as the BT TV example I mentioned earlier). An ISP that has data centers could build its own VoIP call-out service, and its own search engines. Indeed, it might be able to pinpoint "click-through" far more accurately than a search/lookup service at lower cost simply by monitoring network access patterns. However, what is the effect of "taxing" the profit from overlay service providers? Well there are two possible outcomes: firstly, the service cost is passed on to the consumer (and the net profit decreases); or the service provider leaves the network (analogous to Google not indexing Belgian Newspapers

[4]note that the situation is *very* different in Europe, the US and Asia with regards to joint versus separate ownership of access and core networks, which also leads to confusion in this part of the debate.

as per a recent event). The effect is to *damage* the ISPs core business. The point is that there *is already a value chain* between clients, web sites and search engines, and between broadband Internet clients and VoIP service providers and the ISP. The profit made by the overlays is *not independent* of the profit made by an ISP. Of course if the ISP is not making a profit, and the customers are, then the ISP should simply raise its prices transparently. Why would you want the market not to be free? The fact that they don't raise prices, and some ISPs don't make a profit speaks to some other problem.

A completely separate neutrality argument arises concerning the different kinds of content filtering, or *censorship* carried out at various levels (IP and above, e.g. by search engines) in different parts of the world. Technically, I do not feel competent to comment on this, but I would observe simply that the same rules are applied to postal service (e.g. for books, DVDs etc), and that the customer can work around those rules but takes the risk of breaking the law. Most cases I have read about in this area are merely reasonable observance of local variation in what is legal (e.g. pornography laws in the UK are more strict than most of the rest of Europe, holocaust denial is illegal in several, but not all countries in Europe, etc etc).

5. ECONOMICS AND NEUTRALITY

Many of the economists arguing about neutrality have observed that the Internet has been an engine for innovation unsurpassed by earlier playgrounds. They argue that this is a win-win for the consumer and the vendor; that innovation favours the brave, but has a high return on the riskier side of things for the investor (and I think the way the Internet business weathered the `.com` fiasco largely supports this); and that the consumer has seen a remarkable improvement in wealth of services, increase in performance *and* reduction in cost all at the same time.

The neutrality proponents argue that this is good, but the neutrality opponents argue that we are reaching the limits of this part of the Internet evolution. As with other industries (famously studied by economists is the last 100 years of the car tire industry), after a period of evolution in quality, once sees a shift to process engineering, where optimisation moves on to the details of how a service is operated, rather than the business of finding whole new service offerings to deploy.

This matches our experience in the lower layers, where the core IP service saw a fair amount of evolution in the 1990s and very early 1980s, but then the action moved on up the stack to TCP and RTP/UDP evolution, and eventually HTTP and Web Service evolution, and now on to multi-party application evolution (P2P, games, etc).

The neutrality opponents might argue that "We don't know how to do a flag day any more. If we want to modify IP to do something new, it will hurt some users, some of the time, in some places, to give more to other users, some of the time in those places (and possibly in other places)."

The difficulty of this part of the debate is that it is like comparing the proverbial apples and oranges. What is innovation worth? How much do customers care? What is the replacement of digital TV by IPTV worth to me, or VoIP instead of GSM? Or P2P movie distribution instead of netflix?

Part of the debate is about trying to define what the IP service is so that regulation (or even law) can be proposed in an agreed form. I believe that that it to narrow a remit, and (as I have outlined in this paper so far) that the neutrality debate ranges up and down the Internet Architecture. Any definition would have to capture this. Here is a strawman meta-definition of mine, aimed at seperating the various components of the problem space (note, this is not meant to be a prescription):

Connectivity Neutrality must be defined w.r.t end to end service at each and every layer.

Performance Neutrality must define rules for SLAs (existing ones, new ones with EF or other delay bounding services for IP TV), in a measurable, comprehensible and transparent fashion.

Service Neutrality must define rules for availability of new net services multihome, multicast, mobility etc, in a way that allows differences to exist until it is no longer reasonable.

Cross Layer Neutrality must define how combinations of services are built and how the consumer gets to choose between them.

However, having defined neutrality thus, I believe that these are Platonic ideals to which we might strive, but never attain. The system of innovation in the Internet community depends both technically and economically on differences, and the static models of neutrality fail to capture the essential living dynamics.

6. CONCLUSIONS

The net neutrality argument is a debate between radically different stakeholders, and one thing the reader must recall when reading any contributions, is that the goals of different stakeholders are very different. Libetarians and Liberals both argue in terms of welfare: perhaps one can say that the key argument of consumer value *and* service provider profit/margin will come from the ability to support X+Y (e.g. VoIP+EF) as a vertical bundle with a better SLA, but not to deny X with default horizontal class. However, providers often argue purely from their own business perspective, and always remember, legacy landscapes last longer than you think, and IP is now a legacy. A regulator strives for stability. This can be good, but in the current world, the system may evolve to a stagnant area, rather than one of continued innovation. A smart regulator in the 21st century might define the Internet service in a meta-description (It is that service which most of the Internet provides at the current time) allowing local fluctuations above that service to flourish, expand, and coalesce globally to the next phase (e.g. wireless access, multicast, differentiation, etc). The technical community need to rise to the challenge of Internet Service evolution in larger than incremental steps made by yet another BGP tweak. While greed (in the simple sense of maximising profit) motivates the providers and the innovators, in a sometimes holistic alliance, we should be very much afraid of fear which closes down the potential opportunities I have outlined above.

In conclusion then: we never had network neutrality in the past, and I do not believe we should engineer for it in the future either.

6.1 Reference material

There is a huge amount of literature in this area. Most of the papers I have read in preparing this note are about economics, and I found most of them to be naive in the extreme about the technical side of the Internet. Many were quite simplistic applications of market theories. The best single reference I can give which references much other (good) work in the economics, technical and legal/regulatory side is the current entry on Wikipedia, which my colleague, Tim Griffin pointed me to: `http://en.wikipedia.org/wiki/Net_neutrality`

Provisioning for EF for VoIP is very well described in RFC 3245 "An Expedited Forwarding PHB (Per-Hop Behavior)" by B. Davie et al. Mixing toll-quality voice and data affordably in the public Internet is not going to happen with pure over-provisioning everywhere just yet.

6.2 Acknowledgement

This paper benefited from a very close read by Richard Mortier at Microsoft Research, Mark Handley, Ken Carlberg and Richard Gold of UCL, and comments from the PhD students in SRG in Cambridge.