Discrimination, Liberty, and Innovation

Some Thoughts on the Invariable Trade-offs of Normative Purposes and Technical Means in the Internet

Matthias Bärwolff
Technische Universität Berlin
matthias-at-baerwolff-dot-de

ABSTRACT

The Internet has been a loose federation of networks allowing a variety of local discriminations to persist, in order to set off the conceptual problems of moving vital networking functions into the end hosts. And yet, those discriminations have been largely without a loss of generality at the narrow common ground of the IP layer. Discriminations have been ranging from simple access restrictions, to legal usage restrictions, to more or less elaborate application and location discrimination. Today's "equilibrium of discrimination" is tilted towards the latter category, and this has prompted concerns about potentially adverse effects on second-order properties of the internet — most notably "innovation". We argue with reference to von Hayek [26] that a narrow conception of innovation has been well compatible with the market and technical non-neutral realities of the internet. Rather that fighting or prohibiting discrimination patterns, we should focus on furthering fairness and efficiency of the Internet — and thus its "rules of just conduct".

Categories and Subject Descriptors: K.4.1 [Computers and Society]: Public Policy Issues — Regulation; K.2 [History of Computing]: Theory; C.2.6 [Computer-Communication Networks]: Internetworking — Standards (e. g., TCP/IP)

General Terms: Theory, Design

Keywords: Design principles, architecture, internet, congestion, discrimination

1. INTRODUCTION

This paper argues that discrimination, its negative connotations notwithstanding, is a conceptual necessity of the Internet, particularly given its universal scope, decentralized governance, and shared nature. The global scope and thinness of the shared internet layer makes physical access control largely futile, and the decentralized governance beyond national jurisdictions makes it hard to impose legal rules and material sanctions. The sharing of resources in largely indeterminate ways thus necessitates very much a discrimination regarding the end hosts and their applications, precisely because they are so many, and because legal rules and social norms cannot succeed in the absence of feasible sanctioning mechanisms.

This is, of course, not a new insight. In fact, it is one that has remained largely stable over the last 40 years. Licklider and Vezza noted back in 1978, before TCP/IP and ASs, the architectural pillars of today's Internet, became commonplace:

> If we could look in on the future at, say, the year 2000, would we see a unity, a federation, or a fragmentation? [. . .] The middle alternative — the more or less coherent network of networks — appears to have a fair probability and also to be desirable. [16, p. 1342, my emphasis]

Licklider and Vezza have proven right in their prediction: neither have we got a ubiquitous homogenous network offering universal service and common end-to-end SLAs, nor have the PTTs succeeded in taking over public data networking and turning it into centrally controlled full scale services networks from virtual circuits all the way up to application control. Instead, we have a largely global network with a common shared internet layer and no performance guarantees in and of itself. Everything that happens upon this common ground is largely market based contractual arrangements entailing all sorts of different services and uses of the network based on local control and exclusion, or, in economic terms, the extertion of private property rights. The Internet is global, but it is neither fair nor neutral [20].

But, not only is such middle ground the result of the complex and largely spontaneous social interactions in the broader context of the Internet; it is an almost trival technical necessity. To quote from the 2002 version of the now classic *Tussles* paper: "There is no such thing as value neutral design" [6, p. 350].

From the intersection of those two dynamics — the spontaneous social interactions in an open society, and the technical impossibility of complete neutrality — comes the emerging discussion about "tussles in cyberspace" [7]. In an Internet that is subject to its stakeholder "tussling" (chiefly about the distribution of social surpluses) the question has thus been put forward as to how to best shape the "tussle spaces" technically, such that the dynamics of the variables of concern to the stakeholders involved remain isolated and do not unnecessarily spill over to other domains [7].

A prime concern of many in the broader discussion about tussles and their implications have been the potentially adverse effects of practices such as DPI and other more or

ReArch'09, December 1, 2009, Rome, Italy.

less severe violations of explicit or tacit assumptions about the Internet at large on its "second-order" property of "innovation" [14]. In this context, the notion of innovation has come to be equated with vague notions of end-to-end autonomy, sovereignty, and liberty. This indeterminacy is unfortunate because it blurs any informed reasoning about useful implications from such arguments. If it goes along with ignorance about the historical realities of technical discrimination, there is even less hope to arrive at any relevant conclusions about the need for normative principles or rules to accompany today's Internet.

This paper is a modest attempt to add to the discussion, and help draw in on the principles of the Internet which are in constant flux, and thus in need for equally constant rearticulation in writing.[1] Specifically, we are on about formulating a plausible basis on which to argue about normative "second-order" purposes of the Internet and conducive technical means to obtain these. The paper thus (1) sketches the historical realities of discrimination on the Internet and its immediate predecessors, (2) taxonomizes the thus emerging patterns, and (3) discusses the high level purposes of the Internet and whether and how the technically and economically sensible discriminations can be made orthogonal to those purposes. Such considerations will also add to the discussion of whether the notion of innovation as it stands right now is a sensible one to inform the discussion about future internet developments and the need for legal regulation of the Internet we have. The following sections are largely sequenced along this logical outline.

2. A HISTORY OF DISCRIMINATION

In this section we consider the various types of discriminations that have been present on the Internet along a broad classification ranging from blunt to subtle. The following subsections thus deal in turn with the history of access in the first place, formal policies regarding acceptable usage, as well as prioritization and throttling schemes. We close with a summary taxonomy.

2.1 Getting Connected in The First Place

At first, access to what was later to become the Internet was severely limited, not only because of the limited scope and capacity of the Arpanet, but also because in its initial design each Interface Message Processors (IMPs)[2] was tied to precisely one host computer. The limitations of this approach were soon realized, and BBN changed the IMP software so as to allow up to four host computers to connect directly to the IMP.[3] However, while the host computers were typically time sharing systems and thus allowed an even broader access to the network, it is fair to say that in the first two years of the Arpanet (1970 to 1971) only those with terminal access (both physical and modem dial-up) to a host computers linked to an IMP had effective access to the Arpanet.

Access broadened substantially once BBN added terminal handler software and Multi-Line Controllers (MLCs) to their IMPs, thus turning them into Terminal IMPs (TIPs), IMPs that could serve hundreds of terminals connected via modems and dial-up [22]. With the ever increasing scope of the subnetwork of IMPs, the broadening scope of access to those IMPs well beyond the initial host computers in immediate vicinity, and the rise of micro computers (personal computers), access to the Arpanet was soon obtainable for a much broader audience.

From the 1980s, following the adoption of TCP/IP on the Arpanet, more and more networks came to be connected to a collection of networks (inter-network) whose scope has been increasing ever since, and access to which can be considered virtually universal since the mid 1990s, if only by means of dial-up access via commercial ISPs. While this development has prompted many networks to police access to their networks by means of gateways,[4] the resulting Internet is, by large, a collection of hosts acting logically as peers.

In summary, the level of discrimination by restrictions on access has been very strong in the early days of the Arpanet.[5] Nowadays, however, the problem of getting access to the Internet in the first place is virtually nonexistent in Western industrial countries.

2.2 Do No Evil, Please

As the reach of the Internet grew beyond the confines of computer science research centers and military sites, so did the problem of regulating what the network was permitted to be used for. In the days of the Arpanet there was hardly any regulation of usage at all; once a site was connected to the network it was very much free to use it for all purposes it wanted to. After all, it was largely an experimental research facility, not a mission critical business or government network.[6] Also, there was not an awful lot that could

[1] There is a nice quote from von Hayek that is fit to mention here:

> Principles are often more effective guides for action when they appear as no more than an unreasoned prejudice, a general feeling that certain things simply "are not done"; while as soon as they are explicitly stated speculation begins about their correctness and their validity. [. . .] Once the instinctive certainty is lost, perhaps as a result of unsuccessful attempts to put into words principles that had been observed "intuitively", there is no way of regaining such guidance other than to search for a correct statement of what before was known implicitly. [26, p. 60]

[2] At a high level, the IMPs for the Arpanet are roughly what routers are for today's Internet; they formed what was called the subnetwork, the actual data network commissioned by ARPA and built by BBN. At a low, and actually more fitting level, the IMPs can be regarded as switches, the Arpanet being a self-contained network very much like an Ethernet or a WAN.

[3] Also the permissible distance from hosts to IMP was increased over time by introducing changes to the IMP-Host interfact (very distant host interface, [17]).

[4] DoD's Defense Communications Agency (DCA), then in charge of the Arpanet, split off the military host sites from the Arpanet right after the introduction of TCP/IP in 1983, thus forming a network called MILNET which was connected to the Arpanet only by means of special gateways controlling the traffic between the two.

[5] In the early 1980s, networks such as Usenet, the "poor man's Arpanet" were built for the very reason to compensate for the lack of broader access to the Arpanet.

[6] However, there had been instances of sanctions against private uses of the Arpanet [15], and there had been some debate on the limits to freedom of speech on the Arpanet (which, after all, was a government facility, not a private medium) [9, pp. 205 f.], but as a general matter there was no overall policy against certain reasonable uses of the network

be done with the network in the first place — the variety of applications and content was rather limited.

With the NSFNET assuming a central backbone function in the late 1980s, serving and interconnecting a growing number of regional networks with IP routers, it became necessary to articulate a set of rules about what types of usage of the network were permissible, or "accepted", that is, in line with the mandate of the National Science Foundation. Specifically, pursuant to the 1950 NSF Act, all traffic carried by NSFNET would have to be related to "research and education in the sciences and engineering". To this end, an Acceptable Use Policy (AUP) was spelled out that restricted acceptable use to educational and research purposes only.[7] It should be noted that enforcement of those rules by sanctioning violations never became an actual issue.[8] Despite a lack of actual efforts to monitor network traffic, the AUP sufficed to bar outright commercial uses [21, pp. 38 ff.].

Very soon after NSFNET had demonstrated the feasibility of interconnecting different networks by means of IP routers and a backbone network, commercial network providers started to cooperatively interconnect their networks, thus creating an internet infrastructure without restrictive AUPs, and eventually rendering the NSFNET redundant. Commercial networks, however, have been applying their own AUPs which regulate what users may do with and via the network they connect to. Such rules are typically part of the contractual relationship between network provider and end user, and include provisions against unsolicited bulk emails (spam), copyright infringements, illegal material, and malicious abuse of network resources. Such rules can form the basis for technical or legal measures directed against violations. In the extreme case, an end user may have their access completely removed, and have legal actions brought against them. More often, however, network providers will prefer to remain squarely in the role of passive conduits, and remedy any problems resulting from material imbalances in the shared use of their resources by subtle technical means, leaving the pursuit of copyright infringements and other illegal actions of end users to those parties harmed or the government.

as long as they were experimental but not malicious or illegal. Arguably, it was the very use of email *beyond* strict research purposes that elevated it to the network's first "killer app", in turn growing the scope of the network, first by email relay gateways, and later by actual IP interconnection.

[7]To quote from the AUP [21, appendix]:

> NSFNET Backbone services are provided to support open research and education [. . .] [and] open scholarly communication and research. Use for other purposes is not acceptable. [. . .] Extensive use for private or personal business [is unacceptable].

[8]To quote from the report [21]:

> We do not understand how NSF could enforce the AUP against an uncooperative end-user if, for example, the end-user's network also refused to cooperate, because we do not see how the AUP is legally binding on end-users. The AUP is not even part of the award conditions enforceable against Merit [NSFs prime contractor], and NSF generally has no direct relationship to other networks connected to NSFNET or to end-users that would facilitate enforcement. (p. 40)

Also, the rules that end users are subjected to by their network providers may contain clauses that restrict entirely reasonable and legal uses, but which may help the provider price discriminate between different classes of customers. Depending on the price tag they pay for a given service, they may use it in more elaborate and extensive ways. However, few of such rules have actually been successfully applied to different customer segments for an extended period of time. In many cases telling apart the different uses of the network in the first place has proven either too costly (if not impossible), or too invasive to be viable from a customer and public relations perspective. Much rather, many providers nowadays pursue strategies of bundling various offers into a package that renders the need for price discriminating between different customers with different preferences unnecessary.

2.3 Technical Rules for Shared Resources

Sharing the resources and capacity of a general purpose network stochastically between a large number of end users has proven to be technically feasible while being extremely cost efficient compared to a network of dedicated circuits that would have to provide the whole sum of peek capacity needed by all end points. However, this architectural premise necessitates the implementation of strategies for those cases in which the demand put on the network exceeds its capacity. There are two basic themes to this problem: avoiding congestion in the first place, and recovering from it should it have happened. The simplest strategies, marking the extremes in a continuum of solutions are (1) regulating the access of traffic into the network in the first place so as not to overwhelm its resources, and (2) allow all traffic to enter the network and discard it randomly in the network should resources become congested.

In more general terms, there are two crucial trade-offs pertaining to the problem of data networking in general, and whose resolution is a direct result of the design of the network: namely delay versus throughput, and reliability versus cost [18].[9] A reasonable balance of the trade-offs in

[9]The account of McQuillan and Walden [18] is still valid today and worth quoting at some length:

> There is a fundamental tradeoff between low delay and high throughput, as is readily apparent in considering some of the mechanisms used to accomplish each goal. [. . .] Therefore, the network may need to employ separate mechanisms if it is to provide low delay for some users and high throughput for others.
>
> To these two goals one must add two other equally important goals. [. . .] First, the network should be cost-effective. Individual message service should have a reasonable cost as measured in terms of utilization of network resources; further, the network facilities, primarily the node computers and the circuits, should be utilized in a cost-effective way. Secondly, the network should be reliable. Messages accepted by the network should be delivered to the destination with a high probability of success. And the network as a whole should be a robust computer communications service, fault-tolerant, and able to function in the face of node or circuit failures.
>
> In summary, we believe that delay, throughput, reliability, and cost are the four criteria upon which packet-switching network designs should be evaluated and compared. Further, it is the combined performance in all four areas which

shared data networks has been aimed at by various means over time. Most if not all them have been based on some notion of discrimination between different applications and traffic patterns, the different needs of them being the criterion for their being treated differently by the network.[10]

The Arpanet dealt with the above mentioned issues by implementing a strict virtual circuit approach, only permitting as many packets to the network as it was able to deal with at any given time.[11] Also, it employed two priority mechanisms. One was triggered by the choice of a second message number indicating to the subnetwork the desired priority of the messages sent over this channel [18, p. 284].[12] The second mechanism would give preference to one-packet messages (up to 1008 bits) over other messages (up to eight packets large). While the latter would have to wait at the source IMP (once permitted to it) for the destination IMP to acknowledge the reservation of the buffer space needed to house the packets to be sent, the former would simply be sent off to the destination IMP without first going through the buffer reservation procedure. If there was enough buffer at the destination for the one packet — fine; if not, the source IMP would wait for the destination IMP to acknowledge buffer space and then send off the packet yet again [18, p. 285].[13]

With the rise of TCP/IP [5] the balance of functions between hosts and network changed such that from the perspective of the hosts the network was in effect *assumed* to be unreliable, and any two hosts would communicate via a common transport layer protocol (initially mostly TCP or UDP), making them in effect part of the network in a number of important respects. With the hosts assuming vital functions such as reliable transport and reordering of packets, the network could not fail anymore — a reliable TCP connection would in principle function over any network and any concatenation of networks, no matter how dismal their performance.

However, the network would still have to deal with the problem of congestion, the more so since it could not reject traffic entering the network other than by dismissing it. And, recovering from congestion turned out to be a nontrivial problem, because TCP would send discarded packets over and over again at any rate it pleased, thus making the congestion problem worse rather than helping recover from it. The solution has been to bind the operation of TCP logically to those of the network concerned with congestion avoidance and recovery [13, 12], but the problems resulting from the hosts assuming vital network functions and thus putting them beyond effective control of the network have remained conceptually unresolved to this day [3].

The important point here is that the high-level tussling between the network and the end hosts makes *some* form of discrimination by the network and against the hosts inevitable. At least since the NSFNET saw its capacity overhead dwindling have we seen some outright discrimination of IP packets aimed at optimizing the use of network resources and imposing fairness rules upon hosts that are no part of IP and TCP in and of themselves [19].[14] While one may argue that 20 years ago packets were discriminated for more benign ends than today, the reasoning behind discrimination has remained, in fact, largely unchanged; and, arguably, most of the discrimination is actually intended to raise the value of the network for users rather than decrease it [1, 11].[15] In a sense, the discrimination functions logically above the IP layer that have come to be assumed by the network are a direct consequence of the architectural decision to make IP as irreducibly small a set of functions as possible — the famous "best effort" design. There is simply no way at the IP data plane for an end host to plausibly tell the network to prioritize certain outbound traffic. Also, there is no way in which an end can make the network filter traffic *from* certain locations [23].[16] The only way to even allow for negotiating about "mutually beneficial shapes of discrimination" between the hosts and the networks is by resorting to higher level protocols, none of which are a mandatory and thus ubiquitous part of the Internet.

In conclusion, there are a number of performance related reasons for discriminating between different applications and

> counts. For instance, poor delay and throughput characteristics may be too big a price to pay for "perfect" reliability. (p. 247)

[10] For the purpose of this paper we neglect here considerations of virtual private networks which may be sized, monitored, and governed so as to render the problems discussed here irrelevant.

[11] No more than eight messages could be in transit between any two end hosts at one time, and hosts would have to wait for Ready For Next Message (RFNM) messages by the destination IMP before they could send further messages [18, p. 284].

[12] The main idea behind different priority classes in the Arpanet is developed in 18, p. 274]. In order to maximize network performance, routing messages would have to have the highest priority, then came acknowledgments, then packet retransmissions, and only then normal packets.

[13] The objective behind this mechanism is straightforward: small messages could get out of the way much quicker than longer messages without excessive processing overhead and without imposing excessive costs upon other messages.

[14] Mills [19] reports on the Fuzzball routers used in the NSFNET backbone starting 1986 "helping" applications by determining their latency requirements from looking at the IP payload (determining whether a "datagram belongs to a TCP session involving the virtual-terminal TELNET protocol" and putting these in a priority queue) (p. 119). Bohn et al. [2] highlight the secrecy of the scheme:

> Because the backbone administrators did not have any way to provide an incentive to not use the highest priority, they did not publicize the priority-based treatment of traffic, and end users did thus not know it was possible to give high precedence to other applications [than certain interactive applications, specifically telnet] (p. 2 in technical report version)

The need for the prioritization scheme only disappeared when the NSFNET upgraded to T1 capacity leading to an "overabundance of bandwidth", and thus "the designers did not reintroduce the priority queuing for end-user traffic".

[15] While there may be (and have occasionally been) instances of malign discrimination against users and applications [28], the empirical evidence for material and sustained interferences that go against the interests of the end-points is rather slim. This is not surprising, for deviating in a material way from the behavior of a "normal" router *necessarily* makes those effect felt by the end-points, applications, or, ultimately, their users. Policies that go against the interests of an ISPs subscribers have thus far been unsustainable, see the Comcast BitTorrent blocking incident of late 2007 and its repercussions.

[16] The argument of Postel [23] applies to the Arpanet, but it is equally valid for today's Internet.

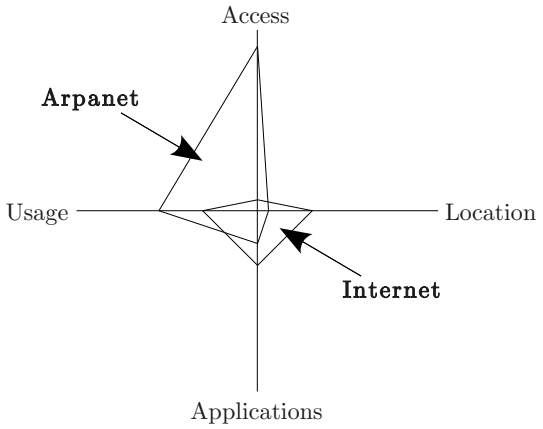

Figure 1: Shifting patterns of discriminations from Arpanet to Internet

locations, typically by employing different priority queues. The problem is to get end users to reveal their different priority valuations, match them with the capacity and resource constraints of the network, and price them such that transaction costs do not render the whole exercise futile. Throwing resources at the problem only ever goes so far. Thus the proliferation of middleboxes assuming various functions relating to the application layer — be it application level gateways designed to handle such issues on more or less explicit behalf of the end points,[17] or be it DPI boxes that aim at essentially the same purpose on explicit behalf of the network provider and often on implicit behalf of the end users.[18]

2.4 A Taxonomy of Discriminations

The taxonomy that emerges from the preceding sections may be summarized as follows: **Access restrictions** — Access to Arpanet was restricted mostly to ARPA funded universities and DoD research contractors in the US. Efforts like NSFNET gradually broadened the scope, and today access is largely non-discriminatory in most parts of the Internet. **Usage constraints** — Certain policies about the acceptable use of the network had been in place earlier, but is was chiefly the broadening range of the Internet that led to formal restrictions by acceptable use policies, their feasibility, however, being rather limited (NSFNET prohibiting commercial use, commercial network providers prohibiting certain uses in order to obtain effective price discrimination). **Application discrimination** — There has always been prioritization of latency sensitive interactive and real-time traffic, and with the Internet also the need to manage congestion (Arpanet's IMPs, NSFNET's Fuzzball routers, commercial network providers' DPI boxes). **Location discrimination** — And, at least since the rise of the Internet to a truly global affordable mass market network there have been efforts by network providers to curb malicious use of the network resources and attacks on other end users of the network by blocking certain traffic patterns from certain locations (e. g., port 25 blocking for home users triggered by excessive use indicating a "zombie" mail relay.).

[17]Think Akamai, a commercial Content Delivery Network (CDN), staging content close to the "eyeball networks".

[18]Think prioritization of voice or gaming traffic, and the "boosting" of small HTTP downloads.

From the early Arpanet to today's Internet the focus of discriminations has largely shifted from the former two to the latter two (Figure 1).

3. DISCRIMINATION AND INNOVATION

We are now turning to the issue of what those patterns of discriminations mean for higher-level purposes. But before doing so, we shall briefly elobarate on what normative higher-level purposes are desirable in the first place. In the introduction we have argued that there is some sloppiness about the notion of innovation in much of the current discussion that this paper pertains to. Considering the broader question of values in society at large is very instructive here to draw in on a sensible notion of innovation. According to von Hayek [26] there are two basic values that must be maintained in any sufficiently complex system characterized by spontaneous order through local knowledge and interactions: (1) upholding the purposes of the individual agents in the system, and (2) upholding the "rules of just conduct" that will maintain and further the overall order. Any other legislated rules or aims are necessarily futile at best, and destructive at worst, for they lack the local knowledge and the purposes of the individuals that make up the overall system. Generally, such purposes will also conflict with Rawls' conception of justice [24] which in principle "does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many" (p. 4).

Thus we have to be very careful not to conceive the notion of innovation too broadly. Innovation should primarily refer to the freedom of individuals to deploy and disseminate inventions [25]. It should not mean that dissemination is to be free of charge, nor should it assume away the roles of intermediaries, for innovation is *always* more than simply dropping an invention onto a preexisting and well-defined "infrastructure": Innovation means taking risks, driving integration so as to ease frictions, and thus shaping new structures, changing that which was before.[19]

Are the prevailing patterns of discrimination that we have considered in the preceding sections detrimental to innovation thus perceived? We argue that they are not, for two main reasons: (1) the Internet is robust enough to allow for various ways of diffusing innovations; and (2) the Internet is large enough so as to effectively render innovation an exogenous force relative to any overly discriminative local practices. Put in terms of our discrimination taxonomy (Figure 1), it is the largely non-discriminatory access and thus global reach of the Internet that renders all other discrimination practices largely irrelevant for the purpose of local innovations, as long as those local settings are large enough to get an innovation off the ground. There is nothing inherently idiosyncratic about the Internet that would preclude an innovation from diffusing along standard lines [25] despite the non-globalness of local circumstances [8]. Plus, if we assume the very reaonable position that the Internet extends beyond IP and to higher layer protocols that are increasingly becoming ubiquitous in their own right (e. g., HTML), then there is even less cause for concern.

[19]The example of HTML is instructive here, for HTTP 1.0 made such inefficient use of the TCP protocol, that it had to revised very quickly, so as to make more intelligent use of TCP (using fewer connections). The incident also highlighted the weaknesses of TCP's slow start algorithm.

The conclusion to this argument is that there is little point in driving up complexity and cost of the Internet at large aiming at minimizing all conceivable kinds of discriminations in an effort to ease the dissemination of innovations. Rather, we should aim at furthering fairness and cost efficiency in an effort to solidify the overall order of the Internet and thus strengthen its feasibility and global reach [4]. This is in all likelihood more important to drive innovations than efforts to stipulate a codified shape of an Internet "infrastructure" and drawing arbitrary lines between regulated common ground and fair game above, for according to von Hayek the agencies necessarily involved here simply cannot digest let alone make informed decisions based on all the dispersed local knowledge there is. Innovation at large is nothing that can be planned for, it can only emerge when private liberty and public order are in balance, with private liberty also pertaining to the right to abstain from innovating or having innovation imposed on one [27].

4. CONCLUSION

We have detailed the history of discrimination on the Internet, and argued that some discrimination is a sensible side effect of the shared nature of the Internet. As long as access is reasonably non-discriminatory and the overall system is large enough (both sideways, and upwards) innovation (as in *individuals* or groups of individuals pursuing and pushing their inventions) is at very low risk. Rather than trying to minimize application or location discrimination practices, it is more sensible to focus on how to get the Internet growing in scope and accessibility by furthering its overall order. Fairness and efficiency are thus more important than the eradication of discrimination per se.

Changing the fundamentals of the Internet is hard enough to begin with [10], and we may just not be as bad off with the Internet we have as some would have us believe. The various layers of interconnection, even if indirect and by name resolution rather than direct flat addressability, make for a robust federation of networks, and may in fact suffice to maintain the system of innovations from physical access, to protocols, to applications, and to all sorts of products and content that evolve at the host sides.

5. REFERENCES

[1] R. Beverly, S. Bauer, and A. Berger. The internet's not a big truck: Toward quantifying network neutrality. In *Proceedings of the 8th Passive and Active Measurement (PAM 2007) Conference, Louvain-la-neuve, Belgium*, 2007. http://www.mit.edu/~rbeverly/papers/truck-pam07.pdf.

[2] R. Bohn, H.-W. Braun, K. C. Claffy, and S. Wolff. Mitigating the coming internet crunch: Multiple service levels via precedence. *Journal of High Speed Networks*, 3(4):335–349, 1994. CAIDA working paper version at http://www.caida.org/publications/papers/1994/mcic/.

[3] B. Briscoe. A fairer, faster internet protocol. *IEEE Spectrum*, 45(12):42–47, 2008. http://www.spectrum.ieee.org/dec08/7027.

[4] B. Briscoe. Internet: Fairer is faster. Technical Report TR-CXR9-2009-001, 2009. http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/refb/FairerFasterWP.pdf.

[5] V. G. Cerf and E. Cain. The DoD internet architecture model. *Computer Networks*, 7(5):307–318, 1983.

[6] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden. Tussle in cyberspace: Defining tomorrow's internet. In *SIGCOMM '02: Proceedings of the 2002 conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 347–356, New York, NY, USA, 2002. ACM. http://www.sigcomm.org/sigcomm2002/papers/tussle.pdf.

[7] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden. Tussle in cyberspace: Defining tomorrow's internet. *IEEE/ACM Transactions on Networking (TON)*, 13(3):462–475, 2005.

[8] S. E. Gillett, W. H. Lehr, J. T. Wroclawski, and D. D. Clark. Do appliances threaten internet innovation? *IEEE Communications Magazine*, 39(10):46–51, 2001.

[9] K. Hafner and M. Lyon. *Where Wizards Stay up Late: The Origins of the Internet*. Touchstone, New York, NY, 1998.

[10] M. Handley. Why the internet only just works. *BT Technology Journal*, 24(3):119–129, 2006.

[11] M. Handley. Network neutrality and the IETF. http://www.ietf.org/proceedings/75/slides/plenaryt-4.pdf (presentation), ftp://videolab.uoregon.edu/pub/videolab/media/ietf75/ietf75-thur-tech-plenary.mp3 (audio recording, starting at 93 minutes into the mp3), 2009. Presentation at the 75th IETF Meeting, July 26-31, 2009, Stockholm, Sweden; Technical plenary, July 30, 2009.

[12] V. Jacobson. Congestion avoidance and control. In *SIGCOMM '88: Symposium proceedings on Communications architectures and protocols*, pages 314–329, New York, NY, USA, 1988. ACM. http://ee.lbl.gov/papers/congavoid.pdf (revised version).

[13] R. Jain, K. Ramakrishnan, and D. Chiu. Congestion avoidance in computer networks with a connectionless network layer. Technical Report DEC-TR-506, DEC, 1987. http://www.cs.wustl.edu/~jain/papers/ftp/cr5.pdf.

[14] J. Kempf, R. Austein, and IAB. The rise of the middle and the future of end-to-end: Reflections on the evolution of the internet architecture. RFC 3724 (Informational), 2004.

[15] L. Layten. File privacy. Email to msggroup@BRL, info-law@Sri-Csl at Wed, 12 Oct 83 6:34:41 EDT, 1983. http://www.dataswamp.net/computerhistory/archives/msggroup/020-jun83-dec83.txt.

[16] J. C. R. Licklider and A. Vezza. Applications of information networks. *Proceedings of the IEEE*, 66(11):1330–1346, 1978.

[17] A. McKenzie. "Very Distant" Host interface. RFC 263, 1971.

[18] J. M. McQuillan and D. C. Walden. The ARPA network design decisions. *Computer Networks*, 1(5):243–289, 1977. This paper is an extension of Crowther et al (1975), Issues in packet switching network design, presented at AFIPS '75, which in turn is based heavily on BBN Report No. 2918; http://www.walden-family.com/public/whole-paper.pdf.

[19] D. L. Mills. The Fuzzball. In *SIGCOMM '88: Symposium proceedings on Communications architectures and protocols*, pages 115–122, New York, NY, USA, 1988. ACM. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.8650 (reprint with different layout).

[20] M. Mueller, D. Cogburn, J. Mathiason, and J. Hofmann. Net neutrality as global principle for internet governance. Research paper, Internet Governance Project, School of Information Studies, Syracuse University Syracuse, NY USA, 2007. http://www.internetgovernance.org/pdf/NetNeutralityGlobalPrinciple.pdf.

[21] Office of Inspector General of the National Science Foundation. Review of NSFNET. Report, National Science Foundation, 1993. http://www.nsf.gov/pubs/stis1993/oig9301/oig9301.txt.

[22] S. M. Ornstein, F. E. Heart, W. R. Crowther, H. K. Rising, S. B. Russell, and A. Michel. The terminal IMP for the ARPA computer network. In *AFIPS '71 (Fall): Proceedings of the November 16-18, 1971, fall joint computer conference*, pages 243–254, New York, NY, USA, 1971. ACM.

[23] J. Postel. On the junk mail problem. RFC 706, 1975.

[24] J. Rawls. *A theory of justice*. Clarendon Press, Oxford, 1972.

[25] E. M. Rogers. *Diffusion of Innovation*. Simon & Schuster International, 5th edition, 2003.

[26] F. A. von Hayek. *Rules and Order*, volume 1 of *Law, Legislation and Liberty: A New Statement of the Liberal Principles of Justice and Political Economy*. University of Chicago Press, 1973.

[27] M. Walfish, J. Stribling, M. Krohn, H. Balakrishnan, R. Morris, and S. Shenker. Middleboxes no longer considered harmful. In *Proceedings of the 6th Usenix Symposium on Operating System Design and Implementation (OSDI 2004)*, San Francisco, CA, December 2004.

[28] J. Windhausen, Jr. Good fences make bad broadband: Preserving an open internet through net neutrality. Public Knowledge white paper, Public Knowledge, Washington, DC, 2006. http://www.publicknowledge.org/pdf/pk-net-neutrality-whitep-20060206.pdf.