Winter December 24, 2009

Towards Extending the Equilibrium of Discrimination in the Internet to its Fringes

Matthias Bärwolff

Towards Extending the Equilibrium of Discrimination in the Internet to its Fringes*

Matthias Bärwolff

December 24, 2009

Abstract

The core insight to the history of the dynamics of limitations, restrictions, and discriminations on the Internet is that they are largely a corollary to the futility of obtaining an entirely 'neutral' design of the Internet that is both robust and scalable. The question then is how to arrange the necessary volume of discriminations so as to obtain a design as value free as possible, for only such a design can be in line with the premise of stipulating no values other than (1) good order at large, and (2) the appreciation of the pursuit of local purposes by individuals. We argue that the best way to achieve such an outcome is by not only allowing "tussling" of the stakeholders in the Internet, but also giving all of the parties the proper tools to discriminate against each and everybody else. Thus a market-like structure obtains, along with the according virtues and benefits. While this is a result that has very much held true 'inside' the Internet – in the peer relationships among different networks –, at the 'fringes' discrimination is largely a one-way street – from ISPs to end users. The end hosts' ability to discriminate on their part is largely confined to blunt host level tools such as firewalls and NATs. However, instead of prohibiting ISPs from discriminating against end users (and containing the technical consequences and costs of doing so) we argue that the more effective solution is to increase the set of tools for end hosts to discriminate against networks, too; so that a fair and efficient 'equilibrium of discrimination' comes to incorporate the hosts at the edges of the Internet, too.

*This paper is a development of the paper that I had submitted to the ReArch workshop 2009, and which is available from the proceedings (online at portal.acm.org). I strongly recommend the present paper over the earlier conference workshop submission.

1 Introduction

This paper deals with the topic of discrimination on the Internet, both on the inside (between networks) and at the fringes (where end users and their host computers connect to a network). On the one hand, we consider how the equilibrium of discrimination inside the Internet has been a result of central technical design decisions and has helped grow the Internet to the global scope it commands today. On the other hand, we discuss some of the aspects of discrimination at the fringes of the Internet, from ISPs towards end users. We argue that instead of limiting the ability of ISPs to discriminate against end users and applications, the more sensible strategy to pursue is to help end users get up to par with network owners by giving them the tools to access various networks at once, each host sharing their network connection with nearby hosts. Thus hosts can implement the equivalent of network owners' 'routing strategies' in their peer relationships to other networks.

However, this paper is largely an attempt to frame the problem in a useful way to begin with, and is thus more of a problem statement than a major step towards a full-blown solution. First, we look at the inside structure of the Internet; second, we consider how the rules that have come to govern the relationships among networks inside the Internet relate to some of the theoretical notions of system stability and purposes put forward by (von Hayek 1973); and, third, we discuss the problem of discrimination imbalance on the fringes of the Internet and briefly consider some of the remedies that could help increase the ability of end hosts to discriminate against networks.

2 Global Rules and Local Discrimination

Elaborating the internal structure of the Internet – the relations between the participants to the Internet and the mechanisms for managing those relationships – is a highly instructive exercise: the internal balance of discriminations offers a number of important insights into the desirable workings of the Internet at large that may help frame some of the more contentious issues about fringe discrimination, too. While (potential) ISP discriminations against end users and higher-level applications typically receive the bulk of attention (and blame) in policy and technology discus-

sions, we argue that any informed discussions on the problems at issue here can only be understood upon appreciating the role of discrimination as a general artefact of the Internet as manifested on its 'inside'.

The general pattern of discrimination inside the Internet is best understood before the background of the global ground rules of the Internet: The most vital global agreement on the Internet pertains to the protocol defining how data is to be forwarded from the source towards its ultimate destination – the famous IP protocol. This protocol prescribes a header format to be added to packets of arbitrary payload data. However, it does not define the service level to be performed by an intermediary tasked with forwarding such IP packet; a router may garble the packet, drop it, duplicate it, and send it to wherever it pleases. The important consequence of this omission of service guarantees, or, indeed, any reliability requirements, is that the Internet can never fail: no matter how dismal its performance, the Internet is still within its specified range of acceptable performance – and this is precisely the typical assumption made by higher-layer transport protocols about the service of the Internet: TCP assumes no explicit lower bound on IP performance, and neither does any other protocol for that matter.[1]

The second global agreement pertains to the sovereignty of individual networks participating in the Internet, and the way in which they exchange routing information. Individual networks are regarded as Autonomous Systems (ASs) which are free to behave internally in whichever way they see fit. Those ASs then exchange routing information via a common protocol (today: BGP version 4), that does little more than announce from any one AS to its neighboring ASs the networks it can reach and the number of networks in between itself and the destination networks. An AS is free to apply any possible policy on its routing decisions based on cost and other business considerations. E. g., a network owner may decide not to forward any traffic originating from youtube, or not to forward any traffic destined for Russian networks, or forward traffic from A destined for D only via B not C, etc. Or, a university multi-homed to different upstream ISPs may decide not to forward any packets originating from outside its

[1] Of course, higher-layer protocols will generally perform better on a more reliable and sane IP layer rather than a less reliable and insane one. And, particularly the congestion control aspects of TCP make the implicit assumption that the underlying IP service is reasonably reliable and sane. Note, though, that while those assumptions have a strong impact on the performance of TCP, they are not vital to its correct functioning.

network to other outside networks at all. The decision how to arrange its relationships with other networks is entirely a private decision of each and every network owner.

Thus the whole system of the Internet is very much based on (1) the notion of a common ground defining a lowest common denominator of what the Internet is, and which is acceptable to most everyone, if only because it contains no fixed lower bound to acceptable performance, and thus entails no minimum requirements for participation in the Internet; and (2) the notion that every participant to the Internet is free to exercise complete freedom over its relationships to any other participants based on idiosyncratic and private purposes, typically business or political considerations of some sort. Seen in this way, discrimination is actually a core feature of the Internet: Everyone may discriminate against everybody else, on the basis of entirely private considerations.

3 Global Order and Local Purposes

The success of the basic architecture of the Internet as briefly elaborated in the previous section is hard to understate by any measure. There has never in human history been any general purpose global communications infrastructure as potent as the Internet. To understand why the Internet based on the two principles considered above – (1) minimum common ground and no lower bounds on any service definition, and (2) maximum local freedom to pursue private purposes by discriminating at will against other participants to the global Internet – has been so successful, it is instructive to turn to von Hayek's 1973 consideration of values in large and complex societies. The basic point of von Hayek is that a large and intricate society can only develop on the basis of two normative values: first, common informal rules that prescribe the "just conduct" necessary to maintain a large order; and, second, the appreciation of local purposes of the individuals making up the system, and the appreciation of those purposes *only*. Any mandated normative purposes beyond the immediate individual ones are most likely to inflict more harm upon the system than do good, for in their abstraction they necessarily lack the local knowledge of circumstances and preferences that is dispersed in a society.

Von Hayek's theory maps well to the Internet system as described in the preceding section: there is (1) a set of rules (mechanisms and protocols)

to uphold and further the global order and reach of the Internet, and (2) it allows individual participants to pursue whichever purposes and thus policies they see fit. In this context the Internet has been growing ever since private networks were allowed to make up a global network of network – today's Internet. The discrimination that is inherent to the system has arguably been vital to its very success in integrating as large a number as possible of individual networks into a coherent federation based on a minimum set of global rules. True, in the early days of commercial inter-networking many ISPs agreed to simply forward one another's traffic without any compensation involved. But, this was more of a pragmatic market solution to the problem of monitoring and accounting for Internet traffic – and which turned out to be only one of many feasible solutions to the general problem of interconnecting different networks, and an increasingly rare one at that (Faratin et al. 2007). The fears that discriminate interconnection would pose problems for the Internet at large have not come to pass, and it is very likely that the Internet has grown as huge as it has precisely *because* network owners were free to pursue any policy of interconnection they wanted to.

Given that the core of the Internet has apparently been thriving on discrimination between what are essentially peers, the interesting question is how the fringe structure of the Internet compares to this. To begin with, the redundancy of interconnection and the number of possible paths between any two nodes is typically much lower at the fringes of a randomly partially connected mesh network than it is at the core (Baran 1960). If a transit network inside the Internet goes down, the impact on overall end-to-end connectivity is typically negligible. If, on the other hand, an access network that a given end point is connected to goes down, then setting up alternative paths to this end point is a far more complex operation involving all sorts of offline efforts ('humans in the control loop'), say, when restoring Internet access through acquiring alternative mobile access solutions such as UMTS.

4 Discrimination at the Fringes of the Internet

End points to a network are fundamentally different from nodes inside the network not only by their intrinsically lower redundancy of direct connections to other nodes, but also because they on behalf of end users host

the applications that are the ultimate sources and sinks of Internet traffic in the first place. The line between applications and network transport of application data is a vexed one, however: even though applications and transport are separated by a stringent logical line in a vertical stack of protocols, there are various interactions between the two broad 'layers' that are, in fact, very hard to properly separate.[2] The main problem here is that the network traffic of different applications is being multiplexed to a shared use of the network, with different applications having sometimes markedly different requirements on the network. Those differences in service requirements as desired by the applications or their users are, however, transparent to the Internet transport system – a result of the original design of the Internet (Clark 1988), and a situation which has since not substantially changed at the heart of it. Fairness between applications is thus a matter that is not addressed at the common shared Internet layer, for it simply offers no useful interface to manage those issues.

It is, of course, possible to keep 'bad' applications, or applications which impose unacceptable negative externalities on other applications, out of the network by mandating certain rules offline by legal or other normative rules, or even simply by restricting the set of people with access to the network to those considered trustworthy or controllable by means outside the network. E. g., access to the early Arpanet was limited to people employed by universities or research firms subject to ARPA funding and contractual obligations. Nowadays, China has become a notorious example of a regime trying to control access to the Internet in the first place by means of an elaborate registration and monitoring system (Deibert et al. 2008). Almost universally, ISPs, both commercial and non-commercial, restrict certain uses of the network by contractual terms (commonly referred to as Acceptable Use Policies, or AUPs). Thus it is typically prohibited to insert unsolicited bulk emails (spam) into the network, abuse or outrightly attack network resources, and disseminate content that falls under ordinary offline exceptions to the free speech doctrine (hate, libel, child pornography, etc.). And, last but not least, we are all normatively asked

[2]We simplify here from the four layer IETF Internet model and adopt a sloppy wording, much without loss to the validity of our argument. When speaking of Internet transport we mean the service provided by the IP layer, and when speaking of the "application layer" we mean everything that happens above IP and at the host computers, including, of course, TCP. In this framing, TCP may be considered an agent under direct control and discretion of the applications invoking its services.

to refrain from using the Internet in "TCP unfriendly" ways, meaning that one should back down in the face of congestion as signalled by packet loss, and generally not game the system of congestion control.

However, the most profound effect of the problem of cleanly separating between individual applications and shared transport has been that the Internet's transport system would try to infer from inspecting an application's traffic the likely requirements of that application, the possible value of that application to an end user, and, potentially, the legality of the application in the first place – all without having to rely on the actual interface between applications and network transport as specified in the relevant common standards. Such an 'inspection' is particularly useful at times of heavy network usage when indiscriminate uniform serving of all applications would impose equal performance penalties on all applications regardless of their sensitivity to such performance discounts. A network can thus raise the value of the network by discriminating between applications and serving them closer to their service requirements, particularly when the network is very loaded.

In fact, technical discriminations have become the most important of all means of obtaining discrimination on the Internet. Whereas norms and laws have an only indirect effect on the actions performed using the Internet (that is, in the worst case: *no* effect on a user's actions), and complete denial of access for a user forecloses *any* actions of that user (an equally insensitive effect); technical means allow the pursuit of very fine-grained policies aimed at certain sources or locations of traffic, certain applications, and certain application or content patterns. Typically, interactive low-volume traffic would best be prioritized over latency-insensitive bulk data traffic, and this general policy has indeed been implemented in networks from the Arpanet to today's Internet (Bärwolff 2009). Also, it is generally agreed to be reasonable to block or degrade traffic from locations that have been found to abuse network resources or inflict harm on other network users.

To return to von Hayek (1973), the purpose of discriminating against certain users and uses of the Internet at its fringes is just as much about maintaining the global order of the Internet as the 'inside' discriminations detailed in the previous sections. The potential for morally objectionable abuses of the ability to discriminate against users or applications notwithstanding, it is fair to say that most of the discriminations on the fringes Internet have in fact been about maintaining the robustness and overall

performance of the network, not least in the face of potentially malign end users which easily evade both access control and legal sanctions. However, the interesting question is: How do end users enter into this game of mutual discrimination based on private preferences and decisions?

5 Extending the Equilibrium of Discrimination to the Fringes

At the border between what may be considered the inner Internet and the end hosts the tussles between different stakeholders of the Internet spill across its otherwise neatly contained *layers*: the end hosts mediate between applications and the lower-level Internet, with applications using the transport services that the Internet module at that host offers to them. Typically, however, the availability of options at this very crossing is rather limited – much more limited than to nodes 'inside' the Internet: historically, a host computer would generally connect to but one point of access to the Internet.[3] Add to this the historical propensity of host level transport protocols such as TCP to assume exactly one network connection, not a multitude of them – and the design of IP addressing of the Internet has left multi-homing to be very much a non-standard way of connecting host computers and their applications to the Internet.

End hosts are thus not at par with the inner nodes of the Internet: they are subject to the discriminations pointed to in the preceding section, but they have a very limited choice of alternatives that they can call upon to best match their idiosyncratic preferences the way networks manage their routing relationships to other peer networks. Thus many argue that legislators and regulators should limit the scope of discrimination that ISPs may subject their customers to. And, as for the residuum of technical discrimination that cannot be done away with, it should be the users, not the ISPs, that control how the volume of discrimination distributes across the various applications they are using, for only the individual user knows their preferences, and any proxying from application patterns or other

[3] In fact, to connect one logical host to various networks (that is, to 'multi-home') requires the host to effectively act as a full-blown Autonomous System (AS), for addresses on the Internet do not point to host computers, but rather to points of network access. A computer connecting to multiple networks thus has to maintain an IP address for each of them.

observable traffic variables is only ever a vague approximation to the true user preferences (e. g., sometimes one may prefer a fast file transfer over a low-jitter VoIP call).

However, the actual policy implications of such reasoning are somewhat limited; in fact, it is probably next to impossible to (1) regulate the technical discriminations of ISPs such that the adverse effects of regulations do not outweigh the benefits from controlling ISP behavior (Kahn 1971*a*; Kahn 1971*b*), and (2) give users the proper tools to signal to the ISPs their discrimination preferences, especially given the required coupling of functions across stakeholders and the amount of cooperation and coordination across stakeholders thus necessitated. The Internet is ossified to an extent that even schemes that would not even require elaborate monitoring and accounting functions (Podlesny and Gorinsky 2008) are hard to introduce; the incentive for an individual ISP is almost universally much lower than that required to make an uncoordinated effort worthwhile.

How then to remedy the power imbalance at the fringes of the Internet, with respect to the tussling between users and application needs on the one hand, and the network and the need to manage the sharing of the lower-level Internet infrastructure on the other? Various research efforts come to mind here, such as multipath TCP (Han et al. 2006) and content centric networking (Jacobson et al. 2009). The important high-level conceptual point we are adding here is that a core objective should be to increase the level of redundancy across multiple networks that a given host can access in a decentralized, ad-hoc fashion. Ideally, a host computer should be able to discriminate between various networks, preferably without the overhead of having to pay for the full access to each of them. It is not the discrimination on part of the network owners that we should aim at curtailing, it is the lack of means of discrimination for end users that we should address.

For example, given the ubiquity of end hosts with wireless networking capabilities, it might be a sensible effort to try and find ways to share the network connection of any one host with other nearby hosts, thus making for a system of multiplexing various different network accesses to each of the hosts participating in such a system. A host could then set up various 'routing' policies, such as "for voice calls always use the fastest available connection", or "before downloading files larger than 3 MB and low priority via a GPRS connection, wait 24 hours and see if a fast DSL connection becomes available", etc. Given the proper technical means,

a host (or rather: multiple hosts in cooperation) could gain access to a sufficient choice of alternative networks, and thus take part in the discrimination game that has been shaping the inner Internet forever – all without expending unreasonable resources overhead or introducing any complexity inside the network. The necessary management tools could all be implemented at the host level, thus rendering a conceptual involvement of network intermediaries unnecessary.

However, there are many issues that would entail from such a concept and that deserve further material elaboration which we will only allude to here. The most significant questions are about the role of each of the parties involved in broadening the level of redundant connectedness at the fringes: On what grounds is access granted to one's Internet connection for unknown third parties? What about monitoring, accounting of credits and debts, and, possibly, payments in money terms? Who shall be accountable for any wrongs committed by third parties via one's Internet connection, and how should identity of those third parties be managed?

Many of those questions have been touched on in the literature, case law, and legislation on liability and identity management for wireless hotspots and arbitrary third parties gaining Internet access through them. The legal situation with respect to wireless hotspots dispensing with any identity management is somewhat sketchy and inconclusive, and the state of the art of managing such hotspots has thus become one where (1) access to unknown and untrustworthy third parties is prevented by encryption and other security measures, and (2) wherever public access is granted it is based on proper identification of the ones using the hotspot.

While browser-based access (universal access method, or UAM) is a feasible solution for acquiring casual access to one hotspot, it is too difficult a solution for seamless roaming and multi-homing; however, some sort of automatic authentication may be achieved in a similar manner.[4] Plus, managing credits and debts from using others' services or providing services to others calls for some kind of global arrangement on how to account for and manage the due figures as a distributed and highly available service. It would be desirable to have a scheme that allows for individual users to accrue credits and debts within a certain range, so as to do away with the need to transfer any credits within the system or manage actual

[4]Possibly, some automated web based authentication system based on OpenID could be used here.

payments. The possibility for such transfers, however, may later be added as an out-of-band or add-on service to the system.

An alternative approach would be to conceive the Internet edge redundancy system discussed here as a variant of routing overlays providing strong privacy (like TOR) and thus avoid the legal issues of determining and dealing with liability for wrongs such as copyright infringements, etc. However, we feel that such an approach makes not only for possible problems concerning the legal feasibility of such a scheme, but also introduces sizeable overhead and complexity that greatly diminish the flexibility and usability of the system.[5]

6 Conclusion

We have here argued that the Internet has been a system characterized by two important features: (1) global agreement on minimum rules to uphold a loose federation of different networks and distributed administrative control; and (2) the appreciation of local purposes and policies to be implemented at the sole discretion of every network owner. These two features have – along the lines of von Hayek's theory of distributed knowledge and the role of informal laws as a result of evolutionary pressure – grown the Internet to an extremely successful and potent network that has given rise to all sorts of applications. However, the system of mutual discrimination based on individual preferences has largely been confined to the inner Internet, where technical means such as the ease of implementing and conforming to the IP protocol plus some very simple control plane functions related to routing have made for a robust competition and a market-like structure along with the virtuous results of such settings. At the 'fringes' of the Internet – the places where users attach to the network,

[5]The latency and robustness characteristics of TOR do not exactly make for a general purpose basis for the whole possible breadth of Internet applications – latency and throughput come at a considerable discount, and a TOR node going down invariably breaks all of the circuits that depend on it.

However, the edge redundancy provided by our scheme could, of course, be used together with TOR to address some of the problems posed by AS level adversaries on TOR connections when both entry and exit node of a TOR circuit are in the same AS. (Increasing the size of the TOR network alone does not help much in mediating this thread, see Edman and Syverson 2009.) But, one may as well use much more mundane considerations such as speed and latency in choosing from the variety of Internet accesses provided by our scheme.

and where applications originate and terminate the actual Internet traffic – there is much less scope for useful negotiation of mutual relationships, and often end hosts are connected to exactly one network, making for next to no scope for elaborate 'routing policies' that would put them up to par with the parties inside the Internet.

However, rather than remedying this imbalance between end users and network owners by mandating discrimination prohibitions for ISPs vis à vis end users, we propose that it would be more useful to pursue strategies of multiplexing existing network connections to multiple hosts by having hosts sharing their network access with other nearby hosts in an ad-hoc manner, thus making for a system in which every host can resort to more elaborate 'routing policies' involving various network options, and thus extend the equilibrium of discriminations to the edges of the Internet, too.

Extending the system of 'peer discrimination' to encompass the host computers has for a long time not been a priority in Internet design, but it should be now; and it ought to before the disequilibrium of discrimination at the fringes of the Internet makes for a need to legally regulate a system that has thus far been working based on global informal rules and local purposes – precisely the two values that von Hayek argues are so instrumental in the success of large complex systems, and that would be at risk by imposing legal constraints on all our ability to discriminate against one another.

References

Baran, P. (1960). *Reliable Digital Communications Systems Using Unreliable Network Repeater Nodes*. The RAND Corporation Paper P-1995. RAND Corporation. URL: http://rand.org/pubs/papers/2008/P1995.pdf. See p. 5.

Bärwolff, M. (2009). "Discrimination, Liberty, and Innovation". In: *Second Workshop on Re-Architecting the Internet (Re-Arch 2008), part of the 5th International Conference on Emerging Networking Experiments and Technologies (ACM CoNEXT), Rome, Italy, December 1-4, 2009*. ACM, pp. 25–30. See p. 7.

Clark, D. D. (1988). "The Design Philosophy of the DARPA Internet Protocols". In: *ACM SIGCOMM Computer Communication Review* 18.4, pp. 106–114. ISSN: 0146-4833. DOI: http://doi.acm.org/10.1145/52325.

52336. URL: http://nms.csail.mit.edu/6829-papers/darpa-internet.pdf. See p. 6.

Deibert, R. J. et al., eds. (2008). *Access Denied: The Practice and Policy of Global Internet Filtering*. MIT Press. See p. 6.

Edman, M. and P. Syverson (2009). "AS-Awareness in Tor Path Selection". In: *CCS '09: Proceedings of the 16th ACM conference on Computer and Communications Security*. Chicago, Illinois, USA: ACM, pp. 380–389. ISBN: 978-1-60558-894-0. DOI: http://doi.acm.org/10.1145/1653662.1653708. See p. 11.

Faratin, P. et al. (2007). "Complexity of Internet Interconnections: Technology, Incentives and Implications for Policy". In: *35th Telecommunications and Communications Policy Research Conference (Papers online)*. ID: 797, pp. 1–31. URL: http://web.si.umich.edu/tprc/papers/2007/797/Clark%20Lehr%20Faratin%20Complexity%20Interconnection%20TPRC%202007.pdf. See p. 5.

Han, H. et al. (2006). "Multi-Path TCP: A Joint Congestion Control and Routing Scheme to Exploit Path Diversity in the Internet". In: *IEEE/ACM Transactions on Networking (TON)* 14.6, pp. 1260–1271. ISSN: 1063-6692. DOI: http://dx.doi.org/10.1109/TNET.2006.886738. See p. 9.

Jacobson, V. et al. (2009). "Networking Named Content". In: *CoNEXT '09: Proceedings of the 5th International Conference On Emerging Networking Experiments And Technologies*. Rome, Italy: ACM, pp. 1–12. ISBN: 978-1-60558-636-6. DOI: http://doi.acm.org/10.1145/1658939.1658941. See p. 9.

Kahn, A. E. (1971*a*). "Economic Principles". In: *The Economics of Regulation: Principles and Institutions*. Vol. 1. 2 vols. John Wiley & Sons, Inc. See p. 9.

— (1971*b*). "Institutional Issues". In: *The Economics of Regulation: Principles and Institutions*. Vol. 2. 2 vols. John Wiley & Sons, Inc. See p. 9.

Podlesny, M. and S. Gorinsky (2008). "RD Network Services: Differentiation through Performance Incentives". In: *ACM SIGCOMM Computer Communication Review* 38.4, pp. 255–266. ISSN: 0146-4833. DOI: http://doi.acm.org/10.1145/1402946.1402988. URL: http://www.arl.wustl.edu/~gorinsky/pdf/RD_Services_SIGCOMM_2008.pdf. See p. 9.

Von Hayek, F. A. (1973). "Rules and Order". In: *Law, Legislation and Liberty: A New Statement of the Liberal Principles of Justice and Political Economy*. Vol. 1. 3 vols. University of Chicago Press. See pp. 2, 4, 7, 11, 12.