Spring May 5, 2009

Application Layer End-to-End Arguments: From Ends to Means, and Beyond Network Neutrality

Matthias Bärwolff

Application Layer End-to-End Arguments: From Ends to Means, and Beyond Network Neutrality

Matthias Bärwolff*

May 5, 2009

Abstract

This paper observes that the ultimate objectives of the end-to-end arguments do not necessarily entail a preference for having functions with the end hosts rather than with the network. The horizontal connotations of the end-to-end metaphor collapse when it is elevated to an application layer argument featuring strong "second order" objectives such as those pursued by network neutrality adherents. An internet that serves those ends will have to allow for ISPs to tussle with end users over economic value and surplus considerations, and acknowledge the beneficial rôle of the resulting transactions between the internet's stakeholders.

1 Introduction

There is no doubt that the internet has become the nexus of computing, featuring killer applications like email and WWW that have had a lasting impact upon society at large. In turn, the internet has become a prime target of economic and legal debate. However, quite often those discourses are divorced from a due appreciation of technological realities and constraints. Add to this an agenda that is political rather than academic in nature, and the scene is set for false analogies, overinterpretations, and unrealistic generalisations.

The discussion on network neutrality has fallen victim to precisely those effects. Crowcroft — a computer scientist by trade — remarks aptly: "Much

*PhD student at Technische Universität Berlin; currently as a visiting student at MIT's CSAIL

of what I have read on the subject of net neutrality by economists is technically naïve and simplistic" (2007, p. 567), a conclusion we readily agree with. The objectives of those advocating network neutrality are generally agreeable and laudable: a default status in favour of unfettered innovation and consumer choice is preferable to one in which they are controlled by centralised agents that are bound to introduce sizeable economic inefficiencies. However, instead of focusing on the high level problems and trying to solve them at a level of abstraction at least as high, very early on an analogy was drawn between end-to-end arguments (Saltzer, Reed, and Clark 1984) — a loose set of arguments in favour of putting functions with the end points of a network rather than in between those end points — and the objectives of network neutrality (Lessig 1999). And, while there have been serious attempts to extend the scope of the end-to-end arguments such that they encompass the objectives of network neutrality (van Schewick 2004), the suitability of applying low level end-to-end arguments to application level issues has remained questionable (Clark and Blumenthal 2007, p. 15). There may just not be a convenient shortcut of mapping the implications of the end-to-end arguments onto the problem set of network neutrality.

It is thus prudent to reconsider the core of the problem afresh: How, if at all, do we marry the objectives of network neutrality with the means of end-to-end arguments? This paper shall attempt to integrate the notions of end-to-end arguments and network neutrality by considering the meaning of the original end-to-end argument, the implications of a broader interpretation of those arguments, and how they relate to the objectives of network neutrality. To this end we proceed as follows. First, we shall reflect upon the meaning and scope of the original end-to-end arguments. Then we discuss the inadequacy of mapping the low level "vertical" implications of the end-to-end arguments to an application and user level "horizontal" version. Building upon an elaboration of the vital rôle of tussles in the internet we argue that higher level objectives are best served by promoting precisely these tussles. We close with the observation that objectives such as end user choice and innovation are *not* at risk, for the IP layer is extraordinarily resiliant to changes, basic interconnection is largely nondiscriminatory, and there are no structural externalities from placing application layer functions in between ultimate end points.

2 The Original End-to-End Arguments

It is a very common misperception in legal circles that end-to-end arguments and network neutrality are effectively the same — meaning a network that is neutral and thus void of any functions that could discriminate between applications and end points.[1] It has also been argued that the internet in the late 1970s was neutral in a broader sense because it followed end-to-end design principles (van Schewick 2004).

However, neither do the end-to-end arguments in any way imply a neutral network, nor has the actual internet been enforcing neutrality. The aim of the original paper on end-to-end arguments (Saltzer et al. 1984) was not to impose neutrality of the internet layer, but to discuss the engineering trade-off flowing from the reality of unreliable networks — an issue that had been addressed comprehensively by Pouzin (1974).[2] The more general conclusion of Saltzer et al. (1984) was that functions should generally not sit at the shared IP layer unless (a) they were common amongst functions, or (b) they increased performance significantly for some applications without imposing negative side effects upon other applications. Thus there is, in fact, ample scope for functions at the shared layer.[3] Put in economic parlance, the prime objective of the end-to-end arguments was to minimise negative externalities from optimising the shared module (the IP layer) in favour of certain applications. One result of this is that in the vertical protocol stack the IP layer is rather neutral among different application layer protocols resulting in the now famous "hourglass" architecture.

Yet nothing whatsoever is being implied about who may interfere with application layer protocols and to which ends. While the internet may, at a low level, be neutral, it is ignorant of what it is being used for. The notion of neutrality is thus a poor conceptual fit to the internet at large. Mueller et al. (2007) note:

> [Some claim] that the Internet protocol itself somehow embodies an agreement to treat all packets equally. But this claim is not accurate. *The TCP/IP protocols [. . .] don't care whether someone reads what is inside the packet or makes a routing priority decision based on the header information or the payload along the way; TCP/IP continues to work as designed whether or not that happens.* (p. 4, emphasis added)

[1] See, e. g., Lessig (2008) who explicitly equates the two notions (in order to impose immediate causality), and Zittrain (2008) who simply merges the two notions into "end-to-end neutrality" (pp. 162 ff.) (albeit in order to roundly criticise it).

[2] The seminal rôle of the Cyclades data network as an intellectual foundation of the later internet with its structural separation of internetwork and host functions (IP and TCP) is rarely appreciated, and all too often relegated to footnotes.

[3] For example, on the one hand, reliability of data transfer is not required by all applications, and also it is prohibitively expensive to achieve perfect reliability at the IP layer. On the other hand, *some* level of reliability is common enough amongst most functions, and can be implemented at reasonable cost. An IP network that drops half of its packets is evidently not what Saltzer et al. had in mind with the canonical example of reliable file transfer elaborated in their 1984 paper (pp. 280 f.).

In a sense, then, it is precisely because of the very flexibility of the upper layers that flows from the minimality of the shared internet layer, that neutrality at layers above IP (whatever neutrality is to mean here) is, in fact, a most unlikely result.[4]

3 Application Layer End-to-End Arguments

In the early days of the internet the minimality of the IP layer coincided with a minimality of functions in the path between two given end points.[5] In such a model, where functions above IP are at the same time in the end systems, as opposed to "in the network", a "horizontal" interpretation along lines of vertical decomposition does not fundamentally change the results of the end-to-end arguments. In both cases the bulk of the functions are in the *end hosts' application layer*.[6]

However, the absence of middleboxes assuming functions beyond those of the IP protocol is *not* a necessary characteristic of the internet's spatial nature. There is no inevitable or per se normatively desirable causality in functions being placed at the application layer *and* functions being confined to the end hosts. As we have noted above, upper level mechanisms and policies do not affect the low level workings of the internet, no matter whether they are in the end systems or in between them. Thus the original end-to-end arguments do not — despite their horizontal connotation — speak to the spatial design of applications, nor do they speak to the tussles amongst different stakeholders and the issue of deep packet inspection (DPI):

> [W]hat the end-to-end argument asserts is that application-specific functions should be moved *up out of the communications subsystem and into "the rest" of the system*. But the argument, as stated, does not offer advice about how "the rest" should be

[4] An example for this effect are the problems caused by the very first version of the HTTP protocol which by opening a large number of very short TCP connections instead of fewer but longer TCP connections put considerable strain upon network resources and incidentally crowded out other applications that observed the unstated assumptions of the TCP congestion control mechanisms (Day 2008, p. 131).

[5] The routers (also called gateways) that would connect the heterogeneous networks to form an inter-network assumed no more functions than *routing* based on a common addressing scheme and *fragmentation* (in order to allow IP packets to be easily transfered through networks with very small PDUs), and apart from routers, there were initially no other boxes "in the network". In passing, we note that distributed routing is not, in fact, as stateless as it might appear — it is by distributing routing state as widely as possible that the complexity of the protocols is drastically reduced (Day 2008, pp. 91 ff.).

[6] For convenience, we regard the transport layer of the 5-layer IETF internet model here as a part of the application layer. This simplification does in no way change the results of our discussion.

> structured. [. . .]
>
> The original end-to-end paper, because it uses a simple two-part model of the communications subsystem and "the rest," does not directly speak to the situation where "the rest" has structure. (Clark and Blumenthal 2007, p. 2, emphasis added)

Thus the original end-to-end arguments have to be conceived as *vertical*, rather than *horizontal* in scope. The horizontal connotations of the end-to-end arguments hold only in a model in which the horizontal placement of the application layer functions *happens* to coincide with with them being confined to the ultimate end hosts.[7]

This assumption, however, does not even remotely correspond with the reality of the internet anymore (and may never have, for that matter). Today's internet is replete with middleboxes (Carpenter and Brim 2002) that reach well above the IP layer in order to unwrap and inspect packet

[7] For the sceptical minds, a brief digressions on our vertical interpretation of the end-to-end arguments: It could be argued that the original end-to-end arguments had a horizontal notion in mind; after all, there is no explicit mentioning of verticality in the original paper (Saltzer et al. 1984), and the metaphor surely connotates such notions. However, given the direct historical contexts, a vertical interpretation is more plausible.

Even before the Arpanet came into being in late 1969, it was obvious to those concerned with host application level functions, that the guarantees of the "subnetwork" delivered by Arpa contractor BBN (BBN 1976) could not be trusted:

> [E]rror checking at major software interfaces is always a good thing. [. . .] [Thus] we would like to see some HOST to HOST checking. Besides checking the software interface, it would also check the HOST-IMP transmission hardware. (BB&N claims the HOST-IMP hardware will be as reliable as the internal registers of the HOST. We believe them, but we still want the error checking. (Crocker 1969, pp. 5 f.)

And, in 1974, Cerf remarks along the same lines:

> Along with Host retransmission, it is necessary to introduce some kind of end-to-end positive acknowledgment. The RFNM [request for next message — the flow control mechanism of the Arpanet] is currently sent by the *destination IMP* to the *source Host* and is taken to mean that a message has been successfully delivered to the *destination Host* (for multipacket messages, the RFNM is sent after the first packet has been delivered). It seems sensible to arrange a *Host level* acknowledgment which confirms delivery. In this case, the RFNM could also be eliminated. (p. 12, emphasis added)

Even though initially IMPs were physically separated from the end hosts, the important conclusion is that error control should be implemented in a layer that is logically above the network layer. Clark (1982) makes the explicit point that those two layers should be separated even if they reside on the very same machine, giving additional credence to our vertical interpretation:

> It must be remembered that things other than TCP are expected to run on top of IP. The IP interface must be made accessible, even if TCP sits on top of it inside the kernel. (p. 12)

payloads, upon which they make all sorts of decisions relating to forwarding, prioritisation, and management issues at large. Plus, more and more applications are designed with the explicit expectation of peers along the path of traffic, rather than plain IP forwarding boxes — an early example of this being email, where application level messages would be relayed on behalf of end users across multiple application level hops.

The assertion that middleboxes are in violation of the end-to-end arguments is thus largely untenable:

> There is no reason to believe that the original reasoning about an unreliable communications subsystem makes any sense at the application level. (Clark and Blumenthal 2007, p. 15)

In fact, a horizontal interpretation of the end-to-end arguments at layers above IP informed by concerns over "second order properties" such as end user choice and empowerment (Kempf et al. 2004) may yield results that markedly depart from those of the original end-to-end arguments revolving around the notion of vertical protocol layering. Building on earlier notions of "trust modulated transparency" (Clark et al. 2003, pp. 40 ff.) Clark and Blumenthal (2007) restate the end-to-end arguments in terms of *trust*:

> The original paper said: "The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system." The generalization would be to say: "The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at a point where it can be trusted to do its job in a reliable and trustworthy fashion." Trust, in this context, is determined by the ultimate end points — the principals that use the application to fulfill their purposes. Because the locus of trust is naturally at the ends, where the various principals are found, "trust-to-trust" is preferable to "end-to-end" from the point of view of the principals. (p. 10)

Also, they argue that functions are likely to be better put with third parties in between two ultimate end points, not only for reasons of convenience, but because end users cannot trust their very own computers (p. 16).[8] Thus

[8] While being a bit of an academic point only, I daresay that only a handful of people worldwide have a completely self-contained trust relation with their computer. Even with free software we are left to trust third party developers that the software we use is, in fact, trustworthy, unless we build the *whole* system including *all* system tools from bottom-up. Thompson (1984) notes:

an application level design of the internet needs to take delegation of functions into explicit account — it should be designed with delegation in mind (pp. 17 ff.).

The "trust-to-trust argument" may be further generalised along the lines of economic efficiency considerations. We can reduce the relevant trade-offs in placing functions at the application layer to a simple utility function of the ends that use the application, rather than a trust metric only. The general rule would thus be: *Functions that are relevant for the end user may best be delegated to third parties in between the ultimate end points to a transaction if the expected net present value of doing so is positive.* This generalisation allows us to incorporate transactions that involve compensation, and where the amount of compensation makes up for the potential losses in end user utility related to the untrustworthiness of intermediaries. The benefits to end users in such scenarios may still outweigh their costs.[9]

4 The Tussles over Functions and Value

In practice, delegating application layer functions to intermediaries "in the network" will entail either that they fully terminate the application layer protocols of the ultimate end hosts, or that they interfere "transparently" by means of DPI subject to out-of-band negotiations or implicit understandings between end users and network operators. The former is more robust and complete, allowing for high level complex functions, and the latter is less powerful, but also less costly in that it renders unnecessary any changes to the operation of and protocols used by the application at the end hosts. And, while the later often connotates malicious intention, it may, in fact,

> You can't trust code that you did not totally create yourself. [. . .] No amount of source-level verification or scrutiny will protect you from using untrusted code. [. . .] A well installed microcode bug will be almost impossible to detect. (p. 763)

A rather worrying example of this problem is provided by the recent Debian openssl vulnerability that rendered all cryptographic key material generated by openssl guessable and thus potentially futile. This highly critical security issue was only uncovered in May 2008, a whole 16 months (!) after it had accidentally been introduced (Weimer 2008). Incidentally, it is precisely for this reason that many system administrators prefer using flavours of Unix systems that have a solid corporate backing, such as Sun Solaris or Red Hat Enterprise Linux.

A similar argument applies to hardware, and it is by no means far-fetched. A vendor could quite easily build hard drives that will keep certain file system data for later forensics without the file system being able to do anything about that. Again, it is precisely for this kind of reasons that some people now put a lot less trust into Thinkpad machines than they used to.

[9] As an aside, it is not per se unreasonable to extend this framework to potential principals of the end users — a setting often found in corporate contexts.

well be that an end user is perfectly happy with an intermediary interfering with and augmenting his transactions even by means of "tacit" DPI. Plus, along with the ever larger capability of internet providers to employ DPI measures comes the capability for end hosts to employ ever stronger means of encryption to guard against DPI should they prefer not tho be subjected to such measures. The feasibility and scope of DPI is thus at virtually complete discretion of the end users and the applications they use. In fact, one may plausibly argue that there are few valid excuses for *not* employing end-to-end encryption and signatures as a default for any application level traffic.[10]

The important thing about tussles between ISPs and end users is that they flow straight from the private nature of the internet and the heterogeneity of its stakeholders, and that it is virtually impossible to perfectly contain them. Extending the framework of Parnas (1972), Clark et al. (2002) conclude that the logical modules in the internet should be designed so as to keep tussles (which we better assume take place rather than futilely assume or regulate away) from spreading across module boundaries, thus affecting parts of the system that are not actually part of the tussle at hand. The economic nature of tussles, not their favoured result, should affect the design of the internet's logical modules.

It follows, as a general rule, that ISPs should not be technically prohibited from employing DPI, and users should be allowed to circumvent it. At the same time, end users should be able to probe the network for DPI measures, and, by the same token, ISPs should be expected to try and conceal such measures. The likely result of this tussle is that DPI measures will be hard to be kept hidden from users determined to uncover them, for DPI that results in user traffic being actively influenced will by definition be impossible to be kept secret. However, in cases where the total surplus at stake is small enough, users are going to be rather indifferent about the ISPs employing DPI, e. g., in order to throttle (or "de-prioritise") lower value and time-insensitive in favour of other more time sensitive traffic.[11]

Second, end users should be free to push application level functions towards third parties in between the ultimate end points, and ISPs should be free to offer users compensation for assuming certain functions in the network. In fact, both sides should generally be able to strike whichever commercial agreements they see fit, resulting in functions being placed in

[10] We neglect here the significant overhead of encryption that would affect much sought after resources as opposed to casual one-to-one communications.

[11] In fact, the surplus could be negative, e. g., in cases of malicious DDOS attacks that could easily be contained with DPI and active network based measures. In such cases, DPI makes both the network and the end users better off.

whichever way the parties consider economically useful and efficient. E. g., it could be that ISPs pay end users for the privilege of placing advertisements with their web traffic, or giving their traffic lesser than best effort service. It could also be that end users pay the ISPs for special prioritisation of their traffic.[12]

Allowing for the application layer tussles between users and network operators to be played out may seem disorderly. But, as Clark et al. (2002) have valuably pointed out, it is better to design for tussles to be pursued within logical module boundaries, rather than have them spread across module boundaries, thus introducing all sorts of adverse side effects. While prohibiting tussles between end users and ISPs over the value and surplus created by applications may on the face of it be sensible, drawing the boundaries in the system such that the players will respect them despite their tussles is bound to be an uphill battle — and one with potentially more casualties at that.

[12] A brief elaboration to the last few paragraphs: The tussle between end users and ISPs can generally be seen as part of a conventional bargaining game in which end users try to conceal or understate the value they derive from the internet, and ISPs try to estimate the true value that users derive in order to capture part of their surplus. End users may use end-to-end encryption, leaving ISPs to guess the true value that they generate on top of their IP services. Such a result may make both parties worse off relative to the Pareto boundary. One solution would be for end users to reveal their value by abstaining from encryption, thus allowing DPI. This is likely when the value end users derive is small enough and the application of non-sensitive nature. Absent such incentives, ISPs may try and offer services based on DPI or application level relays that increase the value of the network to user while making them reveal the value they generate. The conclusion would be that such a solution would result whenever the total value from employing network based application level services is large enough to cover its costs and the surplus that ISPs would leave to end users upon guessing their utility.

We submit that our tussle analysis is an extremely simplified property rights based economic blackboard exercise, and ignores vital considerations about societal trade-offs regarding inalienable privacy rights, and the problem of informed consent. Schwartz (2000) notes duly:

> [T]he failure in the privacy market at present is so extensive that a mere declaration of a property right in personal information is likely to make matters worse rather than better. [. . .] [P]rivacy-control ignores the constraints on choice found in the movement to take-it-or-leave it processing of personal data. Due to information asymmetries, collective action problems, bounded rationality, and limits on "exit," privacy is effectively being defined down on the Internet. As a result, legal identification of a property right will not alone create a functioning market. (p. 832, references omitted)

Also, we note that the actual industry structure often comes to favour third parties rather than ISPs as intermediaries in between logical end points. This evidence is not incompatible with our elaboration; however, it suggests that the cases in which an ISP can, in fact, increase the total value of its network by DPI, thus having an according value proposition, are rather limited.

5 Saving End User Choice and Innovation

In the preceding sections we have argued with ample references and some digressions that networks are, as a matter of fact, nowhere near neutral, nor should they, as a normative matter, be so. Adherents of network neutrality will thus wonder what is to become of end user choice and, most of all, innovation, if there is no platform with fixed functionality, nondiscriminatory access, and prices reflecting no more than marginal costs. If the network is free to interfere with application level protocols, are not both end user choice and innovation at risk?

In view of the damage that tussles crossing module boundaries may inflict upon otherwise unrelated parts of the system, we have argued above that such tussles are generally preferable to unsuccessfully containing them. There is no perfect solution to the tussles — nationalisation and monopolisation surely are no particularly attractive options, given the structural problems of linking private enterprise with government oversight (Thierer 1994). The counterintuitive conclusion is thus that variety and freedom at the application level may be better served by promoting the relevant tussles, rather than by fighting them. Foreclosing the option of tussling between the players making up the internet by imposing "neutrality" is bound to make for severe inefficiencies and side effects that likely outweigh the benefits from neutrality.

Beyond the explicit acknowledging and allowing of tussles, there are three principal reasons why end user choice and innovation are at little risk from "network neutrality violations":

resiliance of the IP layer (bordering ossification) The IP protocol itself is the common ground upon which the internet is build. While application layer gateways have traditionally had their rôle, too, IP has become the core standard for interconnecting heterogeneous networks. Crucially, since everyone depends on IP, and there is no central body exterting control over the internet, it is virtually impossible to change the functions of IP even in the most minuscular ways (Handley 2006).

highly robust interconnection and competition In the late 1990s there was substantial fear that the interconnection structure of ISPs making up the core of the internet would prove uncompetitive. However, while the terms of interconnections have changed from a predominance of informal peering to a rich continuum of commercially driven transit arrangements ranging from settlement free peering to full transit (Faratin et al. 2007), the resulting structure has proven extremely competitive and efficient. After all, any autonomous system (AS) that abides by the the IP protocol and industry standard routing protocols can easily connect with any of the existing ASs, thus joining the internet at large.

minimal structural externalities at the application layer The vertical notion of end-to-end arguments emphasises the minimality of the IP layer because of the externalities that optimisations in favour of one class of applications may impose on other classes of applications. However, as we have noted above, there is no similar argument in the realm of our horizontal application layer end-to-end argument (Clark and Blumenthal 2007). Restricting or otherwise managing application level uses, end user devices, or even complete subnetworks in parts of the internet is not going to affect the internet at large — innovation is exogeneous to singular limitations and specifications at layers other than the IP layer (Gillett et al. 2001b; Gillett et al. 2001a).

The metaphor of network neutrality is thus futile with respect to sensible implications of its objectives. Promoting end user choice and innovation at the application layer can only mean that end users and their applications control the extent to which intermediaries may interfere with their application layer protocols — either by choosing among various intermediaries, and by tussling over their performance. The corollary of this point is that users should be free to hand over control to parties in the network, be it ISPs or other third party intermediaries.

Ideally, those end users (and application developers on their behalf) that do care about the integrity of their protocols, should have maximum control over the communication paths they are exposed to by choosing among various competing low level network providers. Moreover, these users will rely heavily on end-to-end encryption and integrity checks to monitor any deviations from standard IP only service. On the other hand, those users that do not care will want to have the option of delegating as many functions of their applications as possible towards ISPs and other intermediaries — without them being subject to non-discrimination rules.

We are by no means dismissing the need for regulation of the internet entirely, though.[13] It is particularly unfortunate that end users are the ones among the stakeholders in internet applications who have to bear the residual risks that accrue in using the internet: it is their transactions that may go wrong because of unexpected application behaviour, and it is their private data that may in the absence of encryption be misappropriated by third parties with malicious intents. Both application developers and network operators can often disclaim liability in such matters, the former because they have thus far successfully fended off any attempts to introduce

[13] See note 12 supra about the limitations of our paper.

producer liability for software vendors, and the latter because they can often rely on their status as mere conduits which exempts them from liability.

How this would give rise to any regulatory implications is beyond the scope of this paper. Intuitively, shifting liability to those parties that can best minimise the costs of residual risks, one of which are the ISPs, is tempting (Lichtman and Posner 2004). However, because of the complex interrelations between internet infrastructure, applications, and end-points it is not trivial to judge the net effects of such liability regimes. It is probably fair to say that informed and "empowered" end-users would be in the best position to safeguard their privacy, and the integrity of their transactions. How to get there is the hard part. Anyhow, it is a more sensible pursuit than that of network neutrality.

6 Outlook

Currently, tussles and negotiations in the internet play out in various ad hoc ways. For example, it has become common practice to allow access to wireless networks upon agreeing to some terms of service presented via a website. The resulting workflow is (1) establishing a wireless IP connection to an IEEE 802.11 access point, typically automatically via DHCP, and (2) pointing a web browser to an arbitrary domain beyond the wireless network's access point in order to get redirected to an HTML web page that details the terms of use of the network in human readable form, offering payment options or just requiring to click an "accept" button. The crucial point here is that the whole transaction hinges upon a *tacit* agreement to use HTTP and HTML as protocols for negotiating access. No application other than using a browser with human involvement will even provide any hints as to the bargaining position of the network operator, let alone offer the means of completing the due negotiation. Email, remote access, VoIP clients — none of those applications can be used to negotiate access.

It would thus be useful to establish a value free focal point aimed at allowing negotiations along the lines of the economic tussles between the internet's stakeholders. We envisage a common control protocol with a human readable upper interface, very much like the session management primitives of the SIP protocol (Rosenberg et al. 2002), but with a broader scope of applicability. Designing such a protocol is beyond the scope of this paper and will surely make for interesting further projects. However, using HTTP and HTML may be considered quite a good solution for this control plane issue to start with. After all, it takes place at the user layer, and HTML allows for almost arbitrarily complex negotiations involving human interaction. It would be sensible, though, to arrive at a protocol specification that would allow automated negotiations absent direct human involvement.

7 Conclusion

In closing we want to reiterate our main counterintuitive arguments. The rather trivial point is that the internet is neutral only with regard to its core protocol IP. Any broader notion of the internet necessarily introduces ample scope for discrimination. An application layer horizontal end-to-end argument will thus imply a balance of functions that is markedly different from that of the original vertical end-to-end argument.

More interestingly, though, it is only by acknowledging the merit of the tussles between end users and network providers that we can hope to address the normative objectives of end user choice and innovation. Not only is leaving the tussles to play out economically and technologically sensible; there is also very little risk that higher level objectives based on a priori values will be harmed by departing from principles of neutrality. First, the foundation of the internet — the globally shared and thus extremely resiliant IP protocol — ensures the internet's sustained flexibility with respect to application level uses and types of networks and devices it encompasses. Second, IP interconnection is widespread enough to ensure a resonable level of competition at the IP layer. And, third, there are no significant externalities upon the application layer at large by placing certain application level functions in the internet. The biggest problem is thus not deviations from an ideal of network neutrality, but much rather a lack of standard application level means to negotiate the tussles of the internet.

References

BBN (1976). Interface message processor – Specifications for the interconnection of a host and an IMP. Report 1822, Bolt Beranek and Newman Inc. (BBN). http://www.bitsavers.org/pdf/bbn/imp/BBN1822_Jan1976.pdf (This is the January 1976 revision of the 1822 report. The 1822 protocol was developed along with the first IMPs in 1969 but has undergone some changes over the years. It appears that the 1976 revision is the only version available online. A note on the authorship of the report: Robert Kahn claimed authorship of the 1822 Report in an interview with Judy O'Neill on April 24, 1990, http://ia331308.us.archive.org/0/items/Inet95/ (p. 11) and http://special.lib.umn.edu/cbi/oh/pdf.phtml?id=167 (p. 8)).

Carpenter, B. E. and S. Brim (2002). Middleboxes: Taxonomy and issues. RFC 3234 (Informational). http://tools.ietf.org/html/rfc3234.

Cerf, V. (1974). An assessment of ARPANET protocols. RFC 635. http://tools.ietf.org/html/rfc635.

Clark, D., K. Sollins, J. Wroclawski, D. Katabi, J. Kulik, X. Yang, R. Braden, T. Faber, A. Falk, V. Pingali, M. Handley, and N. Chiappa (2003). New Arch: Future generation internet architecture. Final technical report. http://www.isi.edu/newarch/iDOCS/final.finalreport.pdf. Sponsored by Defense Advanced Research Projects Agency (DoD) Information Technology Office (ITO) Under Grants # F30602-00-2-0553 (MIT) and F30602-00-1-0540 (ISI). Issued by the Air Force Research Laboratory, Rome, NY.

Clark, D. D. (1982). Modularity and efficiency in protocol implementation. RFC 817. http://tools.ietf.org/html/rfc817.

Clark, D. D. and M. S. Blumenthal (2007). The end-to-end argument and application design: The role of trust. In *35th Telecommunications and Communications Policy Research Conference (Papers online)*, Number ID: 748, pp. 1–24. http://web.si.umich.edu/tprc/papers/2007/748/End%202%20end%20and%20trust%2010%20final%20TPRC.pdf.

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

Crocker, S. (1969). Host software. RFC 1. http://tools.ietf.org/html/rfc1.

Crowcroft, J. (2007). Net neutrality: The technical side of the debate—a white paper. *International Journal of Communication 1*, 567–579. http://ijoc.org/ojs/index.php/ijoc/article/view/159/84.

Day, J. (2008). *Patterns in Network Architecture: A Return to Fundamentals*. Prentice Hall.

Faratin, P., D. D. Clark, P. Gilmore, S. Bauer, A. Berger, and W. Lehr (2007). Complexity of internet interconnections: Technology, incentives and implications for policy. In *35th Telecommunications and Communications Policy Research Conference (Papers online)*, Number ID: 797, pp. 1–31. http://web.si.umich.edu/tprc/papers/2007/797/Clark%20Lehr%20Faratin%20Complexity%20Interconnection%20TPRC%202007.pdf, http://people.csail.mit.edu/wlehr/Lehr-Papers_files/Clark%

20Lehr%20Faratin%20Complexity%20Interconnection%20TPRC%202007.pdf.

Gillett, S. E., W. H. Lehr, J. T. Wroclawski, and D. D. Clark (2001a). The disruptive user — internet appliances and the management of complexity. *BT Technology Journal 19*(4), 40–45.

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

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

Kempf, J., R. Austein, and IAB (2004). The rise of the middle and the future of end-to-end: Reflections on the evolution of the internet architecture. RFC 3724 (Informational). http://tools.ietf.org/html/rfc3724.

Lessig, L. (1999). *Code and other laws of cyberspace*. New York: Basic Books.

Lessig, L. (2008). Official testimony at the Federal Communications Commission's second public en banc hearing on broadband network management practices at Stanford University, Stanford, CA on April 17, 2008. Presentation, http://www.lessig.org/blog/2008/04/testifying_fcc_stanford.html.

Lichtman, D. G. and E. A. Posner (2004). Holding internet service providers accountable. John M. Olin Working Paper 217, University of Chicago. http://www.law.uchicago.edu/Lawecon/WkngPprs_201-25/217-dgl-eap-isp.pdf.

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

Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM 15*(12), 1053–1058. http://www.cs.umd.edu/class/spring2003/cmsc838p/Design/criteria.pdf.

Pouzin, L. (1974). Cigale, the packet switching machine of the Cyclades computer network. In J. L. Rosenfeld (Ed.), *Information Processing 74, Proceedings of IFIP Congress 74, Stockholm, Sweden, August 5-10, 1974*, pp. 155–159. North-Holland. http://rogerdmoore.ca/PS/CIGALE/CIGALE.html.

Rosenberg, J., H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler (2002, June). SIP: Session

Initiation Protocol. RFC 3261 (Proposed Standard). Updated by RFCs 3265, 3853, 4320, 4916.

Saltzer, J. H., D. P. Reed, and D. D. Clark (1984). End-to-end arguments in system design. *ACM Transactions in Computer Systems 2*(4), 277–288. http://web.mit.edu/Saltzer/www/publications/endtoend/endtoend.pdf (the paper had previously been presented in April 1981 at the Second International Conference on Distributed Computing Systems in Paris, France).

Schwartz, P. M. (2000). Internet privacy and the state. *Connecticut Law Review 32*, 815–859. http://ssrn.com/abstract=229011.

Thierer, A. D. (1994). Unnatural monopoly: Critical moments in the development of the Bell System monopoly. *The Cato Journal 14*(2), 267–285. http://www.cato.org/pubs/journal/cj14n2/cj14n2-6.pdf, http://www.cato.org/pubs/journal/cjv14n2-6.html.

Thompson, K. (1984). Reflections on trusting trust. *Communications of the ACM 27*(8), 761–763. http://www.acm.org/classics/sep95/.

van Schewick, B. (2004). *Architecture and Innovation: The Role of the End-to-End Argument in the Original Internet*. Dissertation zum Doktor der Ingenieurswissenschaften (Dr. Ing.) (PhD Thesis), Technische Universität Berlin, Germany. (available from the university library of Technische Universität Berlin, http://www.ub.tu-berlin.de, signature 4TA3016).

Weimer, F. (2008, May). [SECURITY] [DSA 1571-1] new openssl packages fix predictable random number generator. Debian Security Advisory DSA-1571-1, Debian. http://lists.debian.org/debian-security-announce/2008/msg00152.html, http://www.debian.org/security/2008/dsa-1571 (the vulnerability was discovered by Luciano Bello).

Zittrain, J. (2008). *The Future of the Internet — And How to Stop It*. Yale University Press. http://ssrn.com/abstract=1125949.