Dagstuhl Perspectives Workshop on End-to-End Protocols for the Future Internet

Jari Arkko
Ericsson Research
jari.arkko@ericsson.com

Bob Briscoe
BT Research
bob.briscoe@bt.com

Lars Eggert
Nokia Research Center
lars.eggert@nokia.com

Anja Feldmann
Technische Universitat Berlin
Deutsche Telekom Labs
anja.feldmann@telekom.de

Mark Handley
Dept. of Computer Science
University College London
m.handley@cs.ucl.ac.uk

This article is an editorial note submitted to CCR. It has NOT been peer reviewed.
Authors take full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

This article summarises the presentations and discussions during a workshop on end-to-end protocols for the future Internet in June 2008. The aim of the workshop was to establish a dialogue at the interface between two otherwise fairly distinct communities working on future Internet protocols: those developing internetworking functions and those developing end-to-end transport protocols. The discussion established near-consensus on some of the open issues, such as the preferred placement of traffic engineering functionality, whereas other questions remained controversial. New research agenda items were also identified.

Categories and Subject Descriptors: C.2.2.b [Computer Communication Networks]: Protocol Architecture; C.2.6 [Computer Comm'n Networks]: Internetworking

General Terms: Design

Keywords: Architecture, Control, Cross-Layer

1. INTRODUCTION

Critical momentum is building to evolve the current architecture of the Internet, particularly with respect to its control and management. Several substantial "Future Internet" initiatives have started in Europe, the US and Asia, and vendor and network operator communities have begun to discuss the topic. The expected impact of these developments is some time off, given that most architectural changes take at least five to ten years to deploy. Therefore they must have a useful life of perhaps 30 years beyond that.

However, a disconnect exists between the people designing the internetworking layer and those designing end-to-end data transport methods. This is problematic because the benefits visible to average users depend on robust inter-operation of end-to-end transports used by applications and the internetwork layer. For instance, identifier-locator separation may cause locators to be changed in a manner that causes TCP to treat associated effects as congestion.

A workshop [3] held in Jun 2008 at 'Schloß Dagstuhl' [1] addressed this concern. It brought together researchers and engineers experienced in current and next-generation internetworking architectures with those developing the Internet's end-to-end transport protocols. The primary goal was to begin a dialogue between these communities that will allow a future Internet to deliver real performance, cost and service-quality benefits to its users and services, while at the same time encouraging new applications to evolve.

The 34 workshop participants came from diverse backgrounds; with an even split between academia and industry (plus a few from both and neither). They also had diverse interests under the umbrella of new communication paradigms and network architectures: cross-disciplinary motivations for new architecture, integration of communications with processing and storage infrastructure, improving or enabling new applications, performance characterisation, availability and robustness, handling of unwanted traffic, transition and deployment issues and formalising design principles. Pure end-to-end functions with no interaction across traditional layer boundaries were ruled out of scope, as were issues specific to single applications or link technologies.

Beyond encouraging dialogue, more specific goals included: identifying erroneous assumptions and misunderstandings between the communities; to start synthesising ideas into a common thought framework; identifying the main architectural concepts and building blocks; reducing controversy over the placement of functions; identifying potential layer interactions (good and bad); identifying incomplete and new research issues; and discussing how the process of architectural change might be started.

2. INVITED PRESENTATIONS

The programme favoured discussion rather than conference-style presentations. But six short, invited presentations, briefly summarised here, set the scene. The full presentations are available [3].

2.1 Architectural Stresses & Attempted Fixes

Mark Handley pointed out that many stresses in the Internet have been relieved by point solutions. However, more systemic stresses remain; those that seem to require cross-layer solutions. The research community has proposed numerous solutions, but no major architectural changes have resulted for nearly fifteen years. The community seems not to be good at the holistic thinking necessary to ensure that organisations incurring the cost are also those that realise the gains. Mark took the workshop on a tour through the stresses building up for applications, for transports and for the network, describing how they are all reinforcing each other.

His first example was multimedia applications that desire continuous service despite an unpredictable network. Available capacity for an application changes as other flows come and go, as wireless links fade, as mobility makes the logical links over different radios come and go, and as routing adapts to topology changes. The problem is more about fast acceleration of competing flows than high bit-rate under constant conditions. Fast acceleration at flow start is also the main requirement of online applications with short flows such as Google Maps, games and the like. Faster acceleration requires a richer interface between the transport and the network. However, the research community has not been successful at even appraising the trade-off between faster connection start-up (*e.g.* VCP [20]) and lower queuing delays (*e.g.* XCP [11]).

A second example is applications that do not want complete network transparency, to avoid attacks from spam or distributed denial-of-service (DDoS) attacks. They want firewalls and network address translators (even with IPv6), because they provide some semblance of zero-configuration security. If applications want controlled transparency, they will need richer connection signaling which is in tension with faster flow start-up.

Mark's final example was the routing system, which is trying to maintain its scalability while customers try to improve reliability by multihoming. The most vaunted techniques (*e.g.* LISP [8]) propose encapsulation across the core of the Internet with on-demand mapping on entry to the core in order to find a good exit from the core. But it is likely that this extra level of hierarchy will lead to bigger and more frequent jumps or jitter in round trip time – between the first and subsequent packets, and every time a dog-leg is removed or the underlying topology changes. Additionally, the likelihood of not having a route at all increases. This worsens the already-difficult environment that transport protocol designers are trying to cope with when they try to improve acceleration and flow start. LISP-like approaches might solve routing scalability at the expense of application performance and reliability.

Mark ended by setting down a position that was widely discussed over the following days. He argued that multipath-capable transports with congestion-controlled sub-flows would move many of the stresses out of the routing system. At the same time they would give applications with potential for multiple paths greater reliability—for multihomed enterprise networks, for mobile individuals and for hosts using multiple interfaces [18], especially if built as standard into transport protocols. Even if only a small proportion of heavy traffic sources started employing multipath transmissions, the resulting traffic engineering benefits would affect the Internet as a whole—even traditional single-path flows. He pointed out that applications such as BitTorrent already have similar schemes with many similar advantages. Their use of multipath transmissions has highlighted that Internet service provider's (ISP's) pricing models are currently somewhat suboptimal, which unnecessarily leads them to wanting to suppress the trend, rather than turning it to their advantage – improving their customers' reliability and potentially even solving the routing scalability problem.

2.2 Evolution of the IP Model

Dave Thaler reminded everyone that the interface that IP offers to higher layers carries a lot more implicit baggage than many realise. He presented a comprehensive catalogue of how the assumed IP model has changed over time [14], sometimes consciously, often as an unnoticed side effect while all eyes were on some other deliberate change.

One of his many examples was that it has generally been assumed that packet loss and re-ordering are rare and probabilistic, not deterministic. However, load-balancing across link aggregations has introduced considerable persistent packet re-ordering, whereas route optimisations for mobility in IPv6 or for multicast have introduced deterministic loss and/or re-ordering at flow-start and other current proposals such as LISP continue this trend (§2.1).

As a second example, it used to be assumed that the received packet header should be the same as that sent, excepting certain well-documented fields such as time-to-live (TTL). However, network address translators (NATs) broke this assumption, and then NAT application layer gateways (ALGs) broke it further, in a misguided attempt to fix the problems NATs had introduced.

As a final example, it used to be assumed that host addresses persisted for long periods, so applications cached the addresses returned from name resolution with no notion of a lifetime. Dynamic host configuration and mobility have reduced addressing lifetimes, causing cached application state to be invalid more frequently.

Dave gave many more examples. Some seemed to have only local significance; for instance, whether a host with an address assigned to one interface should forward datagrams to or from that address via another interface. Although different operating systems take different views on this strong/weak host model, it may not remain only a local matter. For instance, the weak host model complicates attempts to validate source addresses, which some are proposing to mitigate DDoS attacks.

Dave warned that many proposed changes to the Internet break previous assumptions that were not well recognised or documented. Increasingly, this means that applications built on earlier assumptions do not work well. Also, the Internet is hampered by problematic assumptions, because no-one made it clear at the time that applications should not rely on them, so they did.

Discussion centred around how much benefit there would be to tying down assumptions, and whether an architecture that is merely a collection of piece-parts with fewer constraints on their semantics gives more evolvability at the expense of some predictability. (A good example arose later when Jon Crowcroft presented the sourceless network architecture—see §4.2.)

2.3 Router Assistance for Transports

Lachlan Andrew surveyed to which degree congestible resources could assist end-point congestion control. He focused on routers helping to determine the source's transmission rate, but acknowledged that router assistance comprised of more functions, such as encouraging sources to minimise congestion (*e.g.* congestion charging) and the network routing around congestion.

The spectrum of options ranges from first-in-first-out (FIFO) drop-tail queuing, which gives no assistance, to weighted fair queuing (WFQ) active queue management (AQM), which is biased towards drop or early drop. From there it continues to more explicit ways of giving feedback, such as explicit congestion notification (ECN) and similar schemes that give minimal explicit congestion signals, MaxNet-like [19] schemes that give precise congestion signals, XCP-like [11] schemes, where a queue calculates precise bit-rate for sources, to ATM-likepolicing, where network-side queuing takes over completely. He then posed the questions "which of these are necessary?" and "which are possible?".

On the economic front, some form of network assistance seems necessary to fix the free-riding problems that would otherwise result from a pure end-to-end approach. The lack of a principled approaches to fairness based on congestion seems to have created a vacuum that is fuelling net neutrality problems; instead ISPs are dipping into packet payloads to allocate flow rates.

On the technical front, network assistance seems necessary in wireless scenarios to disambiguate transmission losses from congestive loss, or, at minimum, to try to locally repair transmission losses with minimal additional jitter. More generally, there is a hard limit to the amount of signaling information each packet can carry under a pure end-to-end approach. With faster flow rates, performance becomes increasingly reliant on higher acceleration, both at flow start and in response to the arrivals and departures of other flows, link changes and mobility. The theory of Shannon and Bode indicates that the pure end-to-end approach is becoming the limiting factor to performance, irrespective of how much capacity is added. Also, the recent work of Jacobsson *et al.* on a new model for TCP's inner ACK clock control loop [10] shows that there will always be networks that suffer queue oscillations as long as we rely on queue build-up for control – a pathology that is only avoidable with the assistance of virtual queues in the network.

Lachlan also pointed to hazards ahead. If the interface between IP and transports changes, the result should not embed current views of fairness and resource allocation into routers long-term. Also, bottlenecks are more often in access and aggregation network boxes—often commodity items sold in large numbers. Hence, even slight increases in complexity or cost will be strongly resisted. Further, every forwarding element in the Internet cannot be updated overnight, so new schemes must be incrementally deployable.

Lachlan's main point was that the first step should be agreement on a common signal between queues and end-system transports. Specific algorithms can be tweaked later. However, to reach consensus on the required wire-protocol and semantics, an understanding of the fundamental limits of an "assistance free" TCP are important. Lachlan briefly put forward his own wire-protocol proposal ADPM [5], which uses side information in the IP header to increase the information carried by the ECN field.

The main new point on the subject of congestion accountability had been introduced in Mark Handley's earlier talk (see §2.1), where he also had advocated architectural support for some form of congestion pricing. He had made the interesting point that ISPs today limit each individual's bit-rate well below physical access capacity as the only way they know how to limit congestion. They conform to the deeply entrenched practice of carving up the bit-rate of their layer-2 access networks, as a crude attempt to limit congestion further into their network. If instead they could target congestion rather than bit-rate, they could make the aggregate access capacity available to all, not just an arbitrarily limited share of it. Both Lachlan and Mark also made the point that DDoS attacks can be considered as a disconnect between the ability to send traffic and being accountable for the congestion it causes.

2.4 Using Cross-Layer Information and Working with Wireless Networks

KK Ramakrishnan picked up from Lachlan's point about network and wireless link layers helping to disambiguate transmission losses from congestion. KK started by exploring the origins of the black-box (pure end-to-end) approach to congestion control. Loss was considered the only universal signal of congestion that all layers must eventually resort to. KK also echoed Dave Thaler's points (see §2.2) adding that the early Internet design relied on an assumption of consistently low loss rates, contrary to evidence that wireless links can exhibit high (transient) packet erasure rates even after attempted link-local repairs with forward error correction (FEC) or automatic repeat request (ARQ). A second assumption Internet designers made was that elephants not mice should be the benchmark for performance gains—still prevalent thinking today.

KK described loss-tolerant LT-TCP [15] to demonstrate the considerable performance potential of a protocol designed to cope with loss rates of the order of 50%. LT-TCP exemplified what can be done if the transport knows all queues support ECN, so it can assume packet losses are never due to congestion. Interestingly, this work took the position that loss repair can be achieved by both link-level ARQ or FEC and end-to-end mechanisms, without the two mechanisms explicitly co-operating or even being aware of each other. With high packet erasure rates, link layer repairs can still leave residual losses that are best repaired end-to-end rather than the link introducing arbitrarily long delays without knowing how important delay is to the application. LT-TCP then adds second and third repair mechanisms, both end-to-end. The second employs proactive FEC based on the long-run loss rate, while the third repairs the remaining gaps with reactive FEC based on acknowledgements.

Yet another interesting cross-layer aspect of LT-TCP is the decision to keep a minimum of segments in the window (using smaller segments) even when the window (in bytes) is small. With one bit of congestion signaling per packet, increasing the packet rate without increasing the bit-rate can increase the signaling information rate from the network to reduce the chances of timeouts (as long as congestion is not due to packet processing overload).

The final part of KK's talk switched to considering path diversity, using the Mplot scheme [13] as an example. The environment was assumed to be wireless (with a high packet erasure fraction), because wireless environments tend to offer more opportunities for multiple paths. Traditionally, when networks introduce multipath they are careful to avoid splitting flows over multiple paths to avoid severe re-ordering and consequent TCP timeouts. However, when the transport introduces multiple paths itself, it can minimise re-ordering. In the case of Mplot, it only maps a packet to a path when it knows the path can take it ("adaptive packet mapping").

There were no new cross-layer techniques in this part of the talk, but Mplot still assumed all congestion would be signaled by ECN. Proactive FEC was used similar to LT-TCP earlier. Unlike the optimised scheme of Kelly (see §2.5), each Mplot flow operates its own independent congestion control. KK showed that using multiple paths reduces the loss variance, which translates into higher goodput because losses on each path are only partially correlated. Also, Mplot sends its ACKs over all paths while still spreading the main data load over separate paths. So, where there is diversity of delay over paths, this ensures the ACK stream used the minimum delay path, thus reducing the return leg of the RTT to a minimum.

2.5 Load-Balancing over Routes as an End-to-End Function

Frank Kelly presented the research [12] that has led him (and Handley, see §2.1) to adopt the position that end-systems, rather than the internal nodes of a network, are in the best position to balance load across paths.

The model is a succinct maximisation of the total utility of all users of the Internet, using the general α-utility model of Mo and Walrand, in which the utility of each flow is parameterised by both a weight w and concavity α. Appropriate values of these two parameters model flows aiming for TCP-fairness ($\alpha = 2$ and $w = 1/T_r^2$, where T_r is each flow's RTT) or other forms of fairness, such as max-min, maximum flow, proportional fairness and weighted forms of each. The shadow price p_j of each resource j can be signaled to each source by the resource adding to the ECN marking of passing packets and the receiver feeding these back to the sender.

Traffic engineering is achieved by a source-destination pair splitting flows into sub-flows that traverse different paths through the network between them. The algorithm they use to congestion control the bit-rate $x_r(t)$ of each sub-flow on path r depends on the desired value of the parameter α. For proportional fairness ($\alpha = 1$), the source increases its bit-rate additively by a/T_r on each positive acknowledgement and decreases it multiplicatively by $b_r y_{s(r)}/T_r$ on each negative acknowledgement, where a & b_r are constants. Note that if the source is to control the amount of congestion its set of sub-flows causes, the multiplicative decrease of each sub-flow must be proportional to the sum of the bit-rates of all the sub-flows $y_{s(r)}$, not just to the sub-flow's own bit-rate.

The model shows that the system can be stable even though sources are continuously load-balancing the whole Internet by shifting load from one route to another within a round trip time of the resources signaling congestion to them – *i.e.* on the same time-scale as rate-control. Frank then gave the condition for local stability, which is that the additive-increase constant of a source must be constrained by $a < \pi/2(1+\beta)$. The value of β must at least be the sensitivity with which the active queue management of each resource increases its congestion signaling with regard to load, *i.e.* $\beta > xp'(x)/p(x)$, for example, if $p \propto x^\gamma$ then $\beta > \gamma$. It is interesting that the constraint on additive-increase has to be tighter if the transport splits into sub-flows over separate paths. The above $1/(1+\beta)$ term in the constraint on additive-increase a for multipath sub-flows can expand to $1/\beta$ for a single path.

From this analysis, Frank concluded that load-balancing, particularly across network domains, is more naturally a transport layer function, especially given that transports are already responsible for rate control in the Internet architecture.

2.6 The Need for Collaboration between ISPs and P2P

Anja Feldmann's talk consisted of two parts. First, how network operators might shift traffic between routes to reduce congestion (traffic engineering) and second, investigating co-operation between ISPs and peer-to-peer systems.

It is well known that ISPs can cause instability by shifting traffic within their network. Anja's hypothesis was that an ISP would be advised to only move flows every hour or so, which is feasible because traffic volumes per flow follow Zipf's law. Therefore, an ISP can occasionally shift the top n "elephant" flows off congested paths onto less congested paths to keep congestion balanced.

On the collaboration between ISPs and P2P systems, Anja and co-workers had investigated the feasibility of an ISP-operated oracle service [4]. This addresses the problem that peer selection tends to be either random or RTT-based, leading to inefficient usage of underlying network resources. In the proposed interaction model, a P2P client sends a list of potential peers it could use to download a particular item of content to the oracle, expecting the list to be returned in the ISP's preferred order. A brief discussion was also given of the pros and cons of the P2P network making better use of the underlying topology independent of the ISP.

Simulations had been conducted to establish sensitivity of the results to several different topologies (underlay and overlay) and to different patterns of user behaviour. The simulated oracle kept 55-88% of content within each ISP, relative to 10-35% using random peer selection. This was consistent with a field trial of a similar approach by Telefonica. From the viewpoint of P2P users, mean download times were reduced by 16-34%. RTT-based peer selection gave results closer to ISP-oracle-aided selection.

Given that one theme of the workshop was on placement of functionality between network and transport, discussion was heated on this topic. There were objections to the assumption that an ISP knows its network better than the hosts that are using it. Although it was generally agreed that an ISP would know its own topology better than its hosts, some argued that hosts were better placed to monitor and to take advantage of fast-changing congestion levels, given that the current architecture is designed to enable hosts to detect and manage congestion. However, the discussion was inconclusive, because different people seemed to rely on different terms of reference, *e.g.* whether congestion over multiple domains was more relevant, and few had much evidence to back up their arguments, *e.g.* whether P2P networks would be able to take advantage of brief troughs in congestion.

3. PARALLEL GROUP WORK

Groups were formed to work in an informal productive setting.

3.1 Multipath Routing vs. Multiflow Transport Protocols

On the premise that multipath routing is desirable (§2.1), the group identified four potential locations to deploy it: i) In the routing system; ii) In IP on the host or a shim above; iii) in the transport layer; iv) in the application.

Below the transport layer, information is too coarse to control shifts of traffic between paths, beyond single networks slowly shifting a few 'elephants' away from hot spots (§2.6). Swarming download technology already does multipath in the application (*e.g.* BitTorrent), where it knows how best to chunk bulk data transfers. Given applications know their sensitivity to delay, they are best placed to do multipath routing. But it would be better if an application could use a generic multipath service in the transport, with an interface to express its delay sensitivity. The transport layer has all the information and machinery to control multiple paths.

It seems the cost of non-network multipath routing would be borne by the same parties that benefit most. Operating system vendors would bear the costs in return for direct gains by their end-users—gains in robustness, fewer bandwidth constraints and more opportunities to minimise delay. Leaf networks would want their users to deploy transport-based multipath to balance multihomed traffic with greater speed and robustness than network-based load balancing. Whether ISPs would prefer to control their traffic engineering by manipulating congestion notification, rather than by advertising route preferences remains an open question. Multihomed users of transport-based multipath would no longer *have* to to use provider-independent address blocks. It would then be in the interests of ISPs to reduce routing table sizes by encouraging such behaviour through their address pricing.

The outstanding technical issue with transport-based multipath is how to distinguish flows to ensure their routes diversify as soon as they enter the internetwork. Networks already split traffic over multiple paths within their domain (equal cost multipath), sometimes at the granularity of flow IDs (or the IPv6 flow label), but otherwise by IP address pair and Diffserv codepoint. In addition, there are likely to be sufficient hosts with multiple interfaces to solve this problem by using all their addresses, or at least creating virtual interfaces.

3.2 Is P2P a Solution or a Problem?

Whether peer-to-peer networks are a problem or a solution, they are popular and unlikely to go away. One side effect is that they might force ISPs to switch to more rational pricing (*i.e.* usage based; now there is a mismatch between skewed usage and flat pricing). For the interaction between ISPs and peer-to-peer networks, there is a distinction between ISP-served and user-served content (and storage). The former can be provided with or without an explicit contract with a content provider. The latter can provide users with peer selection hints to accelerate download and to help the ISP control traffic.

P2P networks offer interfaces to storage and enable community networks and thus a democratisation of content. They offer service creation at the edge (*c.f.* end-to-end principle), fast deployment, and they can drive competition. Moreover, they enable rapid experimentation with new services. But there are also significant scalability limitations, due to user churn and limited traffic control pos-

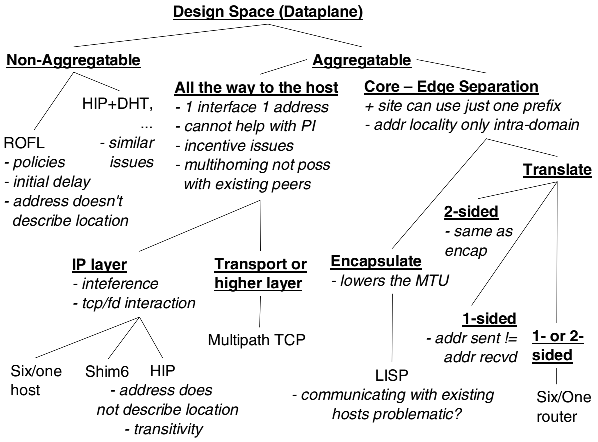

Figure 1: Routing Scalability Design Space

sibilities. As such, an ISP can help by providing either resources, help with bootstrapping and/or hints using its topology knowledge.

It remains open how to combine server-based infrastructure with a decentralised P2P approach, *e.g.* to push content to cities then let a P2P network handle further re-distribution. Given some P2P systems have a copyright infringement reputation, legal issues feature strongly in setting the appropriate level of interface between ISPs and P2P systems. ISPs are concerned to keep their "common carrier" status, while P2P systems want to avoid activity logs held by an ISP being open to subpoena by the courts.

3.3 Designing Internetworking Protocols to Minimize Impact on Higher Layers

Poor internetworking layer design may cause significant harm to upper layers. As a case study, this group focused on recent proposals to improve scalability of the routing system—a problem caused by demand for provider independence, traffic engineering and resilience through multihoming. The group produced an initial analysis of the issues to be addressed in order not to harm higher layers. As the design space of proposed solutions is wide and diverse, the group produced maps of the space (Fig 1 is the main one) to visualise clustering of issues.

The majority of the group were from the internetworking community, so discussion kept veering away from interactions with higher layers towards the purely internetworking pros and cons of each solution, evidenced by most of the $+/-$ annotation in the figure. Ultimately, the market will choose by weighing up the routing and transport aspects of each proposal anyway.

Jari Arkko outlined the IETF's work on "Performance Implications of Link Characteristics (PILC)" in 1999–2003. It published a set of RFCs [9] advising designers of new subnetwork technologies on interactions with higher layer functions. That advice applies equally to the numerous recently proposed internetworking extensions (HIP, Mobile IP, SHIM6 and those addressing routing scalability, *etc.*), which often create similar issues in higher layers.

The main problem this break-out group identified was with proposals that route (or 'map') the first packet of a flow reactively, rather than proactively. This contravenes much of the advice given in RFC3819, the main output of PILC. It advises that network designs should treat all packets with the same flow ID similarly. If that's not possible, designs should avoid delaying some packets more than others, particularly avoiding deterministic drop or delay of specific packet types. Increased re-ordering of the first packets of a flow would also be highly confusing to many transports.

The group agreed that short flows tend to deliver orders of magnitude more value per bit than long [17]. Therefore systematic extra latency for the first packets of flows would compromise the characteristic feature of the Internet that enabled it to win out against connection-oriented alternatives in the first place. Also, transports already take far too long to detect the characteristics of their path (§2.1). The last thing they need is for the first packet to silently confuse their attempts to characterise path delay, congestion and max transmission unit.

4. LIGHTNING TALKS

In this session, attendees had a chance to share recent insights.

4.1 The IP Address

Lixia Zhang talked on "The changing nature of IP Address and its definition in the changing Internet architecture." Her point was that the original IP address design was the best trade-off given their predominant usage at the time. Although IP addresses were intended to locate interface attachment points, higher layer protocols used them as host identifiers too, exploiting their global uniqueness. This dual use suited the common case of a stable 1:1 mapping between the two, avoiding the need for an ID to locator mapping function. But as the Internet and its technologies have evolved (§2.2), a 1:1 mapping is no longer the common case. Most hosts have multiple wireless interfaces and multiple logical VLAN or VPN addresses per physical interface are common. Without a separate host identifier, other corresponding hosts have nothing to tie together the set of addresses that could be used to reach the other host, which can cause problems if a host changes its attachment point by roaming or offers multiple attachments by multihoming. Lixia urged the community to clarify the definition of the IP address.

4.2 Sourceless Network Architecture

Jon Crowcroft presented the sourceless network architecture (SNA), joint work with Marcelo Bagnulo Braun and also related to [6]. The premise of SNA is that the source of a packet is above the IP layer, so the source field in the IP header is merely a convenience that would be redundant if the transport layer specified its own return identifier, much as shim6, six-one, HIP already do. Jon reckoned that these transport changes could be quickly implemented. Transports that split IDs from locators could then use their own mapping function to find the locator from the source ID, using cookies for a secure binding (*c.f.* SCTP). Jon offered several solutions to two legacy uses of the source IP address: i) the network wanting to notify the source of an error; ii) network ingress filters checking source addresses match expectations.

Affinity with Jon's argument was evident in the ensuing lively discussion. A third legacy use of source addresses was identified: to associate packets with flows in quality of service systems. For all three legacy uses of the source address, an architectural insight was offered: if a network node needs the source of a packet, it must be performing a transport function, so it should implement processing of transport IDs.

4.3 Stateless NAT

Christian Vogt proposed stateless NAT (aka. Six/One Router [16]) to allow sites to maintain provider-independent address configuration—an underlying cause of the routing scalability problem. Stateless NAT can be deployed at a site on one side of the Internet without any corresponding remote function. By sacrificing end host transparency, this makes stateless NAT incrementally deployable.

4.4 Influencing Outcomes in the Real World

Greg Minshall said 1988 was the last time anyone really changed the Internet (TCP congestion control). He argued we will never make significant changes as long as we limit ourselves to insignificant ones—on deployability grounds. Some objected that we still see significant changes, *e.g.* BitTorrent. Others agreed that aiming for strategic goals was hard and important research, but objected that working out the next step towards the goal is also valid and challenging research, not 'just' engineering. This in turn raised the objection that we should focus on how the Internet adapts to new pressures and innovations, not strategic goals as if we know the future. Nonetheless, there was general agreement that funding of even *basic* research is now far too constrained by short-term industrial relevance requirements. And worse, researchers gain industrial relevance merits by constraining themselves to apparently short-term solutions without needing any real industrial engagement.

4.5 Scoping and Layering

Lou Burness gave her insights from John Day's recent book on network architecture [7]. There is not just one network layer with an end-to-end transport layer above it. A network layer extends across a scope, such as one operator's network. Then, rather than this network layer being directly encapsulated within a network layer covering a wider scope (*e.g.* internetwork), each scoped network sits beneath a transport layer spanning the same scope. This transport might be so rudimentary (*e.g.* just tail-drop in queues and propagation of drops to the wider scope) that it is often not recognisable as such. We must learn to cater for these alternating network and transport layers over widening scopes if we are to accommodate such functions as ARQ or traffic policing in the architecture.

5. CONCLUSIONS & NEXT STEPS

The seminar achieved its baseline goal of dialogue between distinct communities. Enough individuals crossed between groups to counter the tendency of some to retreat into their tribes.

A straw poll on placement of the traffic engineering function revealed near-consensus that the best location would be the transport layer. One saw no problem with the network also doing TE slowly, and a minority of one preferred TE in the network alone. One other argued TE should be placed above the transport socket.

Numerous next steps were identified: A) A large group promised their own further research into the multipath ideas. Subsequently, Mark Handley has presented "Multipath TCP and the Resource Pooling Principle" to the IETF Transport Area and a position paper has already been published [18]). This was also proposed as a topic for a cross-layer joint meeting between the IRTF research groups on routing (RRG) and on congestion control (ICCRG). Three of the partners in the Trilogy project are also coding these ideas. B) Work will continue in the Internet Architecture Board on 'The Evolution of the IP Model'. C) Guidelines are needed on extending inter-networking without breaking higher layers. D) Some were interested in more diversity for where hosts get locators to access content. Subsequently an IETF BoF (ALTO) was organised. A research forum on this topic is also needed, or perhaps on cross-layer issues more generally. E) Another seminar will be organised (c.2010) to revisit the issues raised in this one. F) We need to address the IETF's inability to change its own architecture.

During the workshop, new research agenda items were identified: i) We will need to extend socket APIs to give applications (optional) control over multipath transports; ii) instead of assuming there needs to be consensus over whether functions (TE, resource control & QoS, loss repair, routing, *etc.*) should be placed in the network or on endpoints, we should research what aspects of these functions work in *both* locations, rather than assuming they will conflict. iii) The idea that every network encapsulation includes its own rudimentary transport layer needs development.

Acknowledgments

This workshop was held at International Conference and Research Center for Computer Science "Schloß Dagstuhl" in Wadern, Germany, supported by German federal and state funds. It was sponsored by Trilogy [2], a research project supported by the European Commission under its Seventh Framework Program. The organisers would like to thank the participants for their contributions: B. Ahlgren, L. Andrew, M. Bagnulo, S. Buchegger, L. Burness, D. M. Chiu, C. Courcoubetis, J. Crowcroft, E. Davies, P. Eardley, K. Fall, A. Gurtov, F. Kelly, P. Key, G. Minshall, W. Mühlbauer, D. O'Mahony, J. Ott, K.K. Ramakrishnan, M. Scharf, B. Strulo, M. Särelä, D. Thaler, C. Vogt, K. Wehrle, M. Westerlund, D. Wischik, A. Wundsam and L. Zhang. B. Ford reviewed this report but could not attend.

6. REFERENCES

[1] Schloß Dagstuhl. `http://www.dagstuhl.de/`.

[2] Trilogy Project. `http://trilogy-project.org/`.

[3] Perspectives Workshop: End-to-End Protocols for the Future Internet. `http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=2008242`, June 2008.

[4] V. Aggarwal, A. Feldmann, and C. Scheideler. Can ISPs and P2P users cooperate for improved performance? *SIGCOMM Comput. Commun. Rev.*, 37(3):29–40, 2007.

[5] L. L. Andrew, S. V. Hanly, S. Chan, and T. Cui. Adaptive Deterministic Packet Marking. *IEEE Comm. Letters*, 10(11):790–792, Nov. 2006.

[6] C. Candolin and P. Nikander. IPv6 Source Addresses Considered Harmful. Tech Report IMM-TR-2001-14, Tech Uni Denmark, Lyngby, Denmark, November 2002.

[7] J. Day. *Patterns in Network Architecture: A Return to Fundamentals*. Prentice-Hall, 2007.

[8] D. Farinacci, V. Fuller, D. Oran, D. Meyer, and S. Brim. Locator/ID Separation Protocol (LISP). Internet-Draft draft-farinacci-lisp-10, Internet Engineering Task Force, Nov. 2008. Work in progress.

[9] IESG. Performance implications of link characteristics (pilc) charter. URL: `http://www.ietf.org/html.charters/OLD/pilc-charter.html`, 2003. (Archived).

[10] K. Jacobsson et al. ACK-Clocking Dynamics: Modelling the Interaction Between Windows and the Network. In *Proc. IEEE Conference on Computer Communications (Infocom'08)*, Apr. 2008.

[11] D. Katabi, M. Handley, and C. Rohrs. Congestion Control for High Bandwidth-Delay Product Networks. *Proc. ACM SIGCOMM'02, Computer Communication Review*, 32(4):89–102, Oct. 2002.

[12] F. Kelly and T. Voice. Stability of end-to-end algorithms for joint routing and rate control. *ACM SIGCOMM Computer Communication Review*, 35(2):5–12, Apr. 2005.

[13] V. Sharma et al. Mplot: A transport protocol exploiting multipath diversity using erasure codes. In *Proc. IEEE Conference on Computer Communications (Infocom'08)*. IEEE, 2008.

[14] D. Thaler. Evolution of the IP Model. Internet Draft draft-iab-ip-model-evolution-01.txt, IETF, 2008. (Work in progress).

[15] O. Tickoo, V. Subramanian, S. Kalyanaraman, and K. Ramakrishnan. LT-TCP: End-to-End Framework to Improve TCP Performance over Networks with Lossy Channels. In *Proc. IWQoS'05*, June 2005.

[16] C. Vogt. Six/one router: a scalable and backwards compatible solution for provider-independent addressing. In *Proc MobiArch'08*, pages 13–18. ACM, 2008.

[17] D. Wischik. Short messages. In *Proc. Workshop on Networks: Modelling and Control*. Royal Society, Sept. 2007.

[18] D. Wischik, M. Handley, and M. Bagnulo Braun. The Resource Pooling Principle. *ACM CCR*, 38(5):47–52, Oct. 2008.

[19] B. P. Wydrowski, L. L. Andrew, and I. M. Mareels. MaxNet: Faster Flow Control Convergence. In *Conf. Networking 2004, Springer LNCS 3042*, pages 588–599,. IFIP, 2004.

[20] Y. Xia, L. Subramanian, I. Stoica, and S. Kalyanaraman. One more bit is enough. *Proc. ACM SIGCOMM'05*, 35(4):37–48, 2005.