# IETF 95 - TSVAREA Minutes

Chairs: Spencer Dawkins, Mirja Kühlewind, Martin Stiemerling
Notes: Brian Trammell
Jabber Scribe: Allison Mankin

## Administrivia

We welcomed Mirja as incoming TSV Area Director, replacing Martin.
We noted that the CONEX, PPSP, and STORM working groups have been concluded since IETF 94.
We noted that we have created a TSV Area Review Team, and a triage team to identify drafts that need special attention from TSV reviewers. This is to help TSV ADs do a better job of document review in less time.
We thanked Linda Dunbar for her service as TSVAREA Secretary.
Gonzalo Camarillo did a short commercial for a non-working group-forming BOF at this IETF meeting - ACCORD: “Alternatives to Content Classification for Operator Resource Deployment”.

## Nomcom

We noted that Spencer’s AD position is up for review in this Nomcom cycle.
We made changes to the TSV AD position for Nomcom in the previous Nomcom cycle, and do not plan to make more changes for this Nomcom cycle, but we are always happy to listen to suggestions. The current position description is at https://trac.tools.ietf.org/group/iesg/trac/wiki/TransportExpertise, and comments are welcome at tsv-area@ietf.org,

## Rekindling Network Protocol Innovation with User-Level Stacks

This presentation was given by Felipe Huici.

On slide 28

Jerry Chu: Numbers for TCP?
Felipe: These are for UDP. We have a minimal TCP stack, later in the slide.
Jerry: Could you run the Linux stack?
Felipe: In principle, yes.

This was followed by Q&A:

Nacho Solis: What happens when you run many different stacks in different apps?
Felipe: These were all the same stack but completely independent. Stacks are light.
Nacho: What is the cost of enabling multiple stacks?

Felipe: Minimal. Within 1% of the switch (?)
Nacho: But you have to context switch across multiple processes,
Felipe: Yes. A few percent performance loss. You can't beat the kernel.

Jerry: But you get something for the cost.
Jerry: How hard is it to take a full TCP stack and put it on top of this?
Felipe: Not a trivial port. It's a research paper but who would maintain it?
Jerry: We're interested in this area.
Felipe: Okay, so Google will maintain it.

Jerry: How much performance boost compared to just raw sockets? Does most of the boost come from NetMap?
Felipe: Not sure what would happen if we just put on top of DPDK. But this doesn't need hardware support.

Jana Iyengar: I love this work, read the paper, general direction in which the industry is moving.
Jana: This solution allows rapid deployment on the server side, but the block is on the client-side. The extensibility of TCP as a protocol is limited, the deployability on the client-side is limited.
Felipe: Yeah, that's an issue. We require a kernel module. You can't ask people to do that. In BSD you can turn it on, in Linux, it's not upstream.
Jana: That's like five clients. But I do have a solution here. Would love to see more work that looks at UDP performance. Lots of work in creating transports on top of UDP. You say the client-side is the block. But even windows 10 upgrades itself now, so it's the middleboxes. So I'd encourage work to look at increasing performance of UDP-based transports.

Christian Huitema: On the graph you showed on TCP options (from the Honda paper). The block is not the OS, it's the middleboxes. QUIC changes this somewhat by having encrypted packets, it makes you failureproof.
Felipe: In the paper we use the term "well-designed extensions" for things that get past middleboxes.
Christian: When you go to the client side, a user level stack is a library inside the app. So to update the stack you have to update the app. Lots of apps don't get updated very quickly. This is kind of an area for research -- how to manage this operationally.
Felipe: Don't know how you could modify the library automatically.
Christian: We see this with OpenSSL. Bugs in OpenSSL are therefore more persistent

Nacho: Having the stack built into the app defines the API underneath the app. At the same time in TAPS we're trying to work from a higher level above and create a new API there. And then you need to figure out how to get these new stacks to interoperate in the network. So here we've opened a fight on three fronts.

Felipe: Our work looks at only two: compatibility with sockets or speed. We picked speed, we need to change the apps. Maybe you can do an abstraction layer, but then people need to use it. Maybe you need to sacrifice performance to do that but it's more compatible than our stuff.
Nacho: You're providing weapons so we can start fighting here. Very useful work.

Jana: Have you looked at UDP vs. TCP performance in NetMap?
Felipe: ?
Jana: Need to figure out how to get TCP offload-like performance for UDP transports.

Brian Trammell: As I said a year ago, I still like this work. TCP vs UDP is more complicated than you think - they're both broken in middleboxes, just in different ways. Take a look at the discussion yesterday in MAPRG.

## Current and future hot topics in TSV / cross-area

This was an open mike session.

Mirja: just to start off, what we're already doing on stack evolution - TAPS WG, etc.

Al Morton: Speed measurement. In the words of Randy Bush at another workshop last year, "rathole". There's a relationship between what's being developed for the future of transport and what we'll be measuring in the future. I think we know how to do IP packet transfer, but when you go above that, it gets very complicated. We've been working on this for a while - Matt Mathis and others have been tackling this for years. We are on the verge of being where regulators are not going to leave us alone anymore. We need to work on methods of measurement that are useful, and relevant, and fair to the service providers, and we may have to do this more quickly than we would like. As a service provider, I'm dealing with this more often than I want to.

Christian Huitema: This is a new focus, on latency. Prior focus: fair sharing of resources or bandwidth. We didn't have a focus on latency. Protocols like TCP were not designed to play nicely here. Mechanisms like Slow Start on high-bandwidth networks cause huge spikes in latency. What we are doing now is AQM. But we have very little synchronization between work on AQM and transport evolution yet. Bob Briscoe is pointing out that deploying AQM can actually get pessimistic results when you deploy it on routers with a lot of background traffic.

Brian Trammell: Latency measurement, latency vs. bandwidth, ISOC and RIPE had a joint workshop on measuring latency a few years ago that went nowhere. We want to look at the interaction between latency and capacity. We've stopped thinking that more queues are better but we should be looking at more complex curves than just use less queue. Re AQM: in SPUD we're looking at lots of things that look like a transport expressing preferences about its AQM parameters. We're just at the beginning of this work.

Mirja: We should be looking at that workshop report. Post a link to the workshop report to the list?

Gorry Fairhurst: It's not really fairness, we're trying to avoid starvation and congestion collapse. Now we have better techniques, we just need to explore them.
Mirja: That's probably within scope of the ACCORD discussion - I hope we have more discussion there.

Mirja: Have people seen non-TSV considerations? It doesn't look (from a show of hands) like we're spending much time in non-TSV working groups!