TRANSPORT & COMPLEXITY

Russ White
November 2024

russ@riw.us

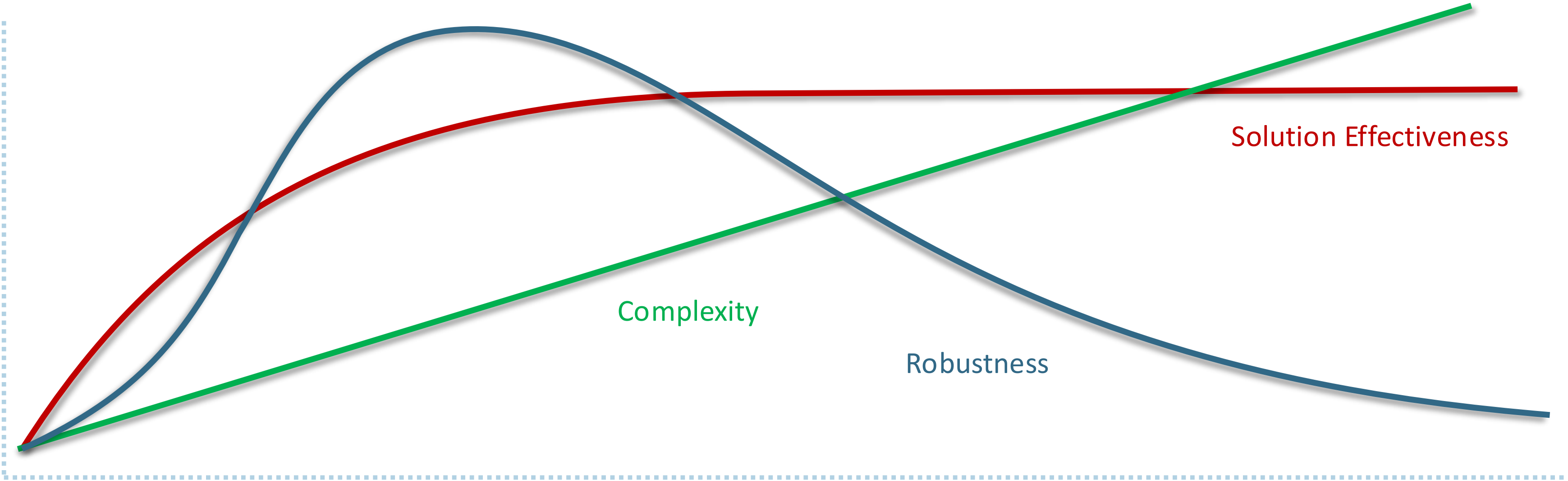
What really matters when moving stuff around?

What role does complexity play in moving stuff around?

QUESTIONS

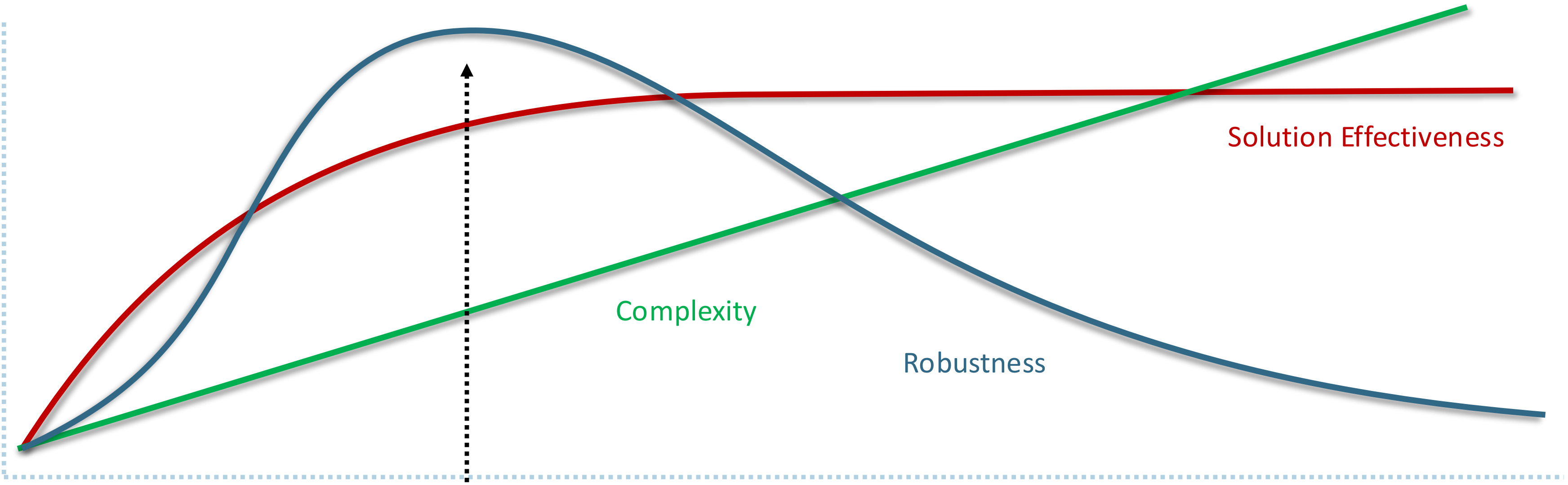
…complexity is most succinctly discussed in terms of functionality and its robustness. Specifically, we argue that ***complexity in highly organized systems arises primarily from design strategies intended to create robustness to uncertainty*** *in their environments and component parts.*

Alderson, D. and J. Doyle, "Contrasting Views of Complexity and Their Implications For Network--Centric Infrastructures", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 40, NO. 4, JULY 2010



COMPLEXITY

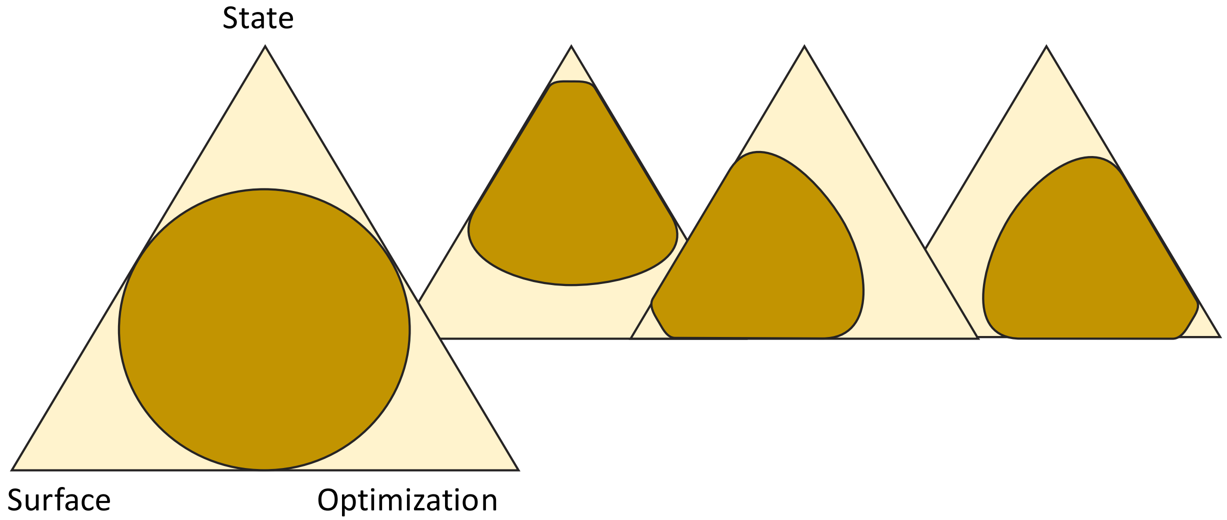
COMPLEXITY



How do we get here?

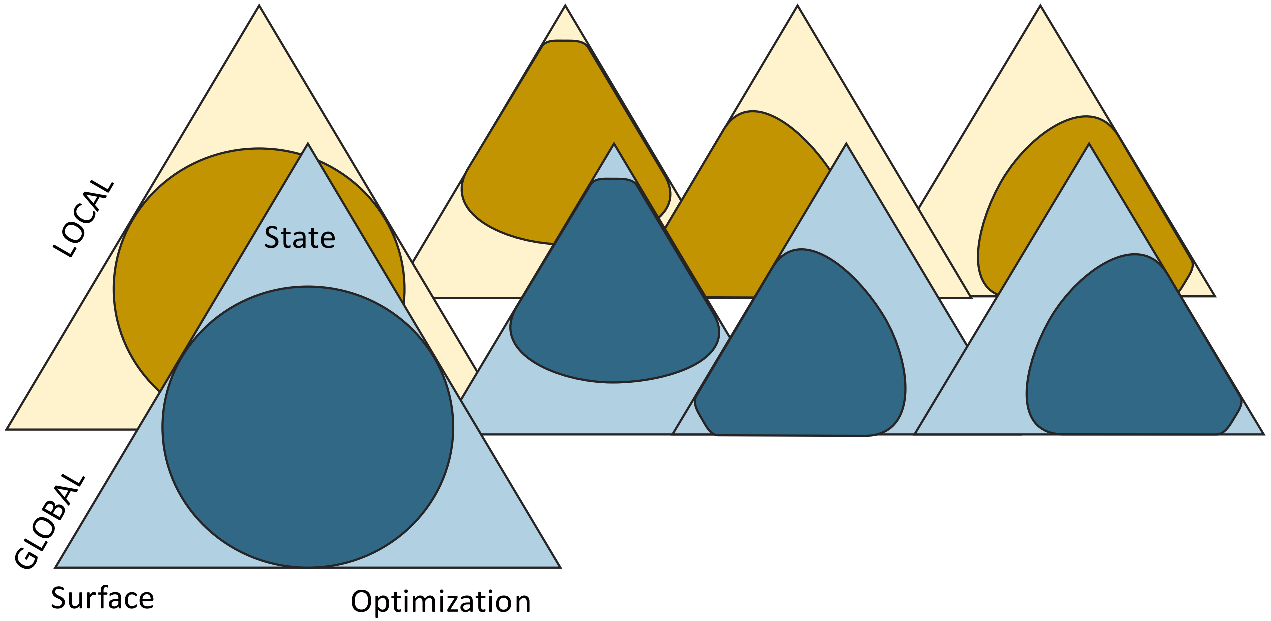
abstraction
layering
protocol
topological
virtualization

locality
local optimization
global optimization



... abstractions introduce tradeoffs ...

TRADEOFFS



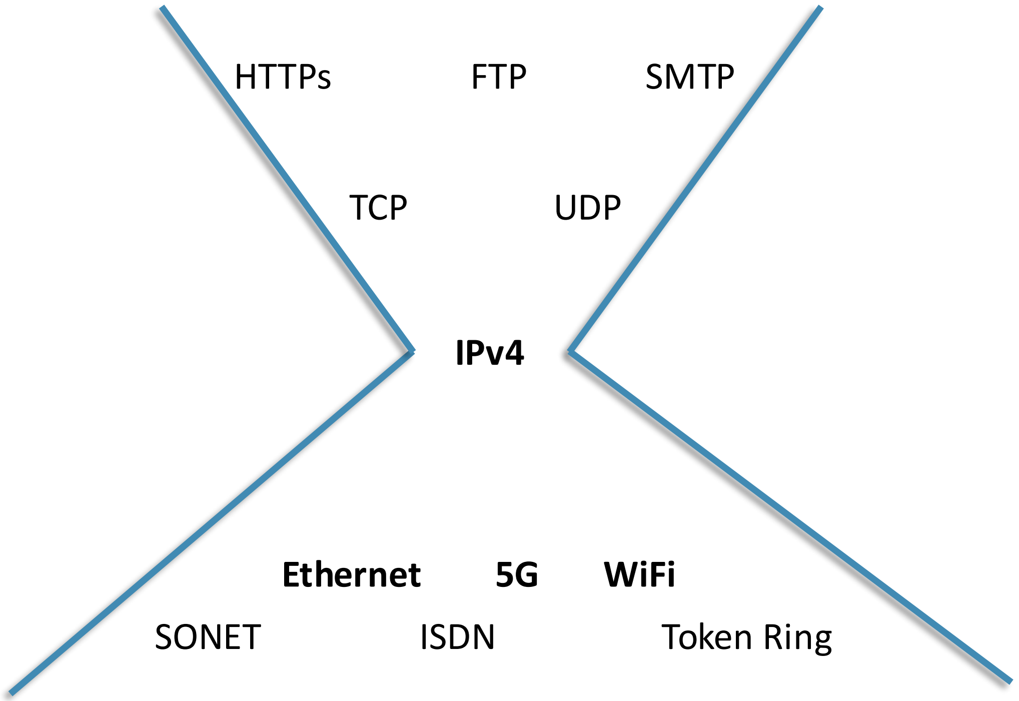
... locality adds a third dimension ...

local/global

how does this look in the protocol world?

TRADEOFFS

TRANSPORT ABSTRACTION



HTTPs FTP SMTP

TCP UDP

IPv4

Ethernet **5G** **WiFi**

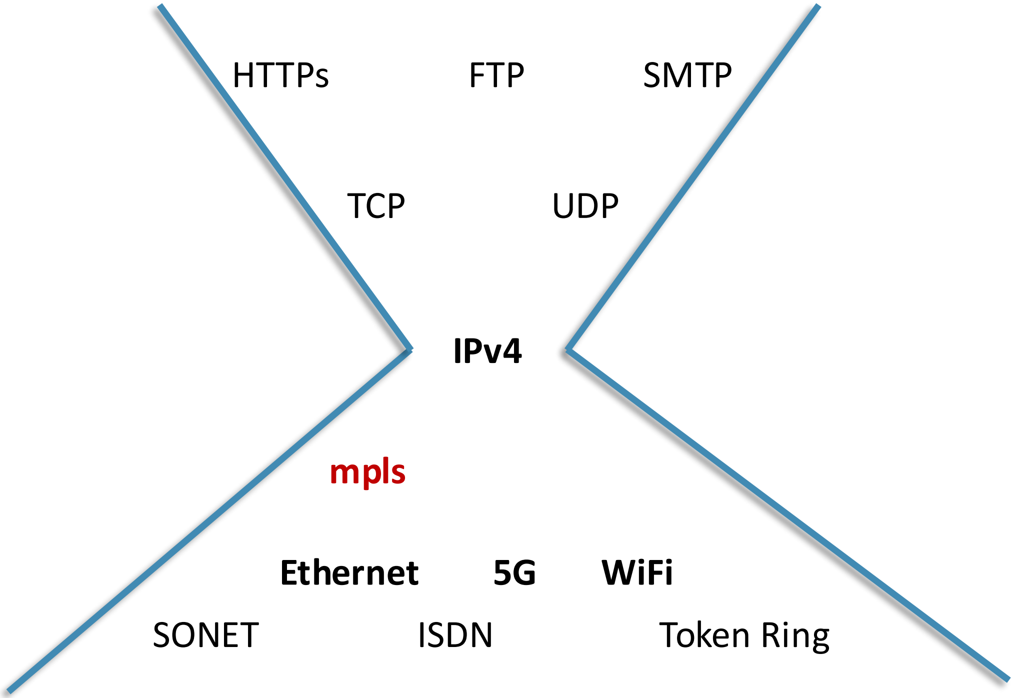
SONET ISDN Token Ring

Layering is the primary abstraction method in protocol design

In the “old world…”

IPv4 was the global simplifying abstraction

Did this work?



IPv4 didn’t work well for …

- … “generic overlay” traffic
- … steering/engineering

Add MPLS

- Between IPv4 and Ethernet
- Additional control plane state

MPLS

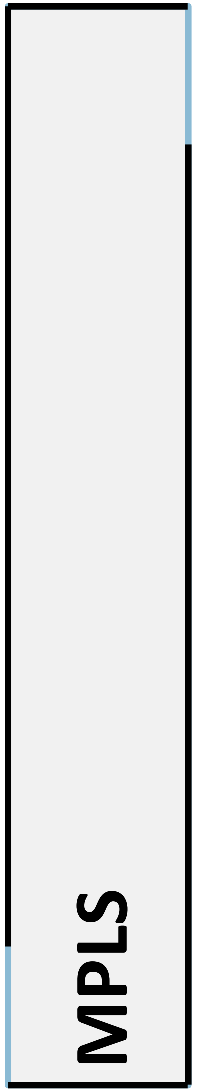
Why didn’t MPLS “eat the Internet?”

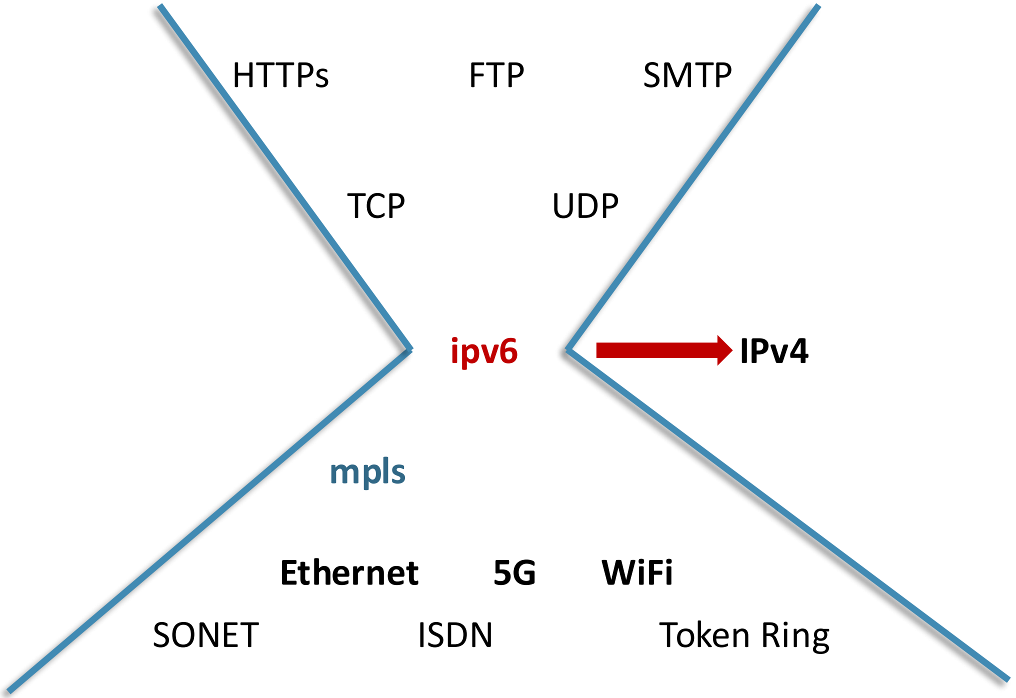
- Shouldn’t every AS be an MPLS/BGP-free core?
- Perceived to be complex, hard to deploy, hard to manage, etc.
- To gain bandwidth optimization, we ...
  - ... added new interaction surfaces
  - ... added new control plane state

Abstracting the new control plane state *necessarily* limited the new optimization too much

So ... MPLS becomes a *localized solution*

- *... the additional complexity isn’t globally practical ...*





HTTPs FTP SMTP

TCP UDP

ipv6 IPv4

mpls

Ethernet **5G** **WiFi**

SONET ISDN Token Ring

Maybe we just need to replace IPv4

- IPv6
- bigger address space
- traffic steering capabilities "built in"

Replacing a universal abstraction is *hard*

- Sheer cost of core component replacement is high
- Traffic steering and other "fancy stuff" is still *too much state*

MPLS

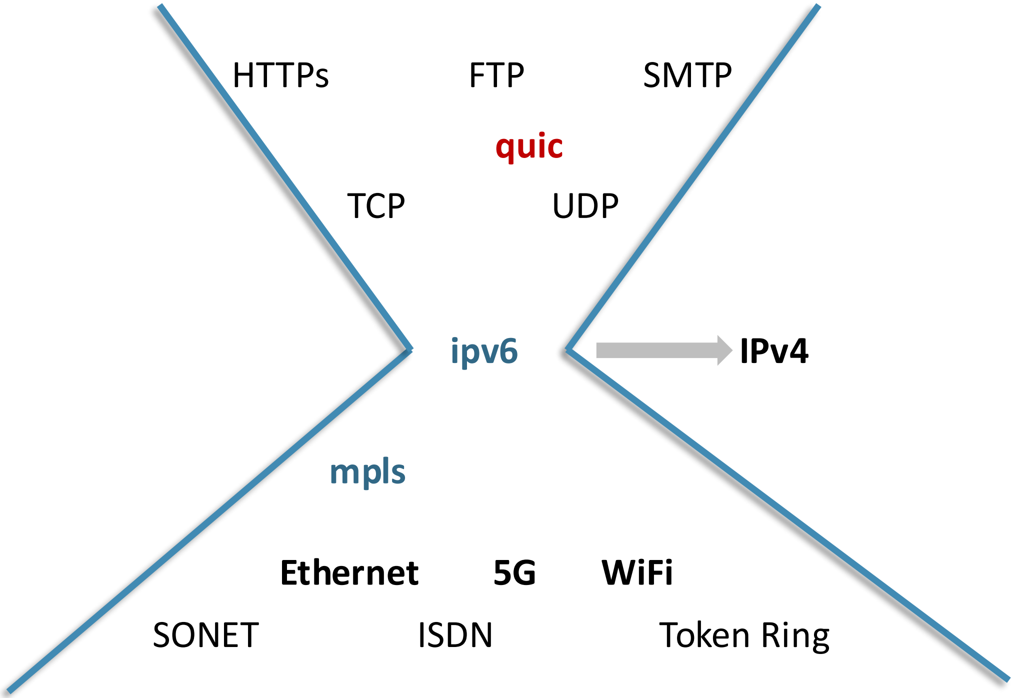
Stated goals are complex

- Increase address space
- Replace DHCP with autoconfiguration
- Get rid of NAT
- Create more rational extensions

Possibly *too complex*

- At least some of these have been “backed off” over time
- Make it simple, make it extensible, make it work ... then extend it

IPv6

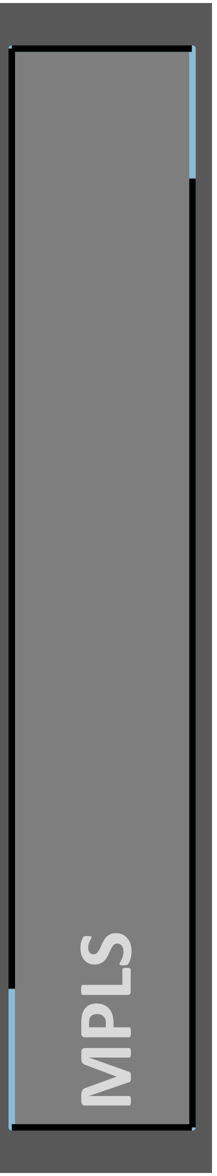


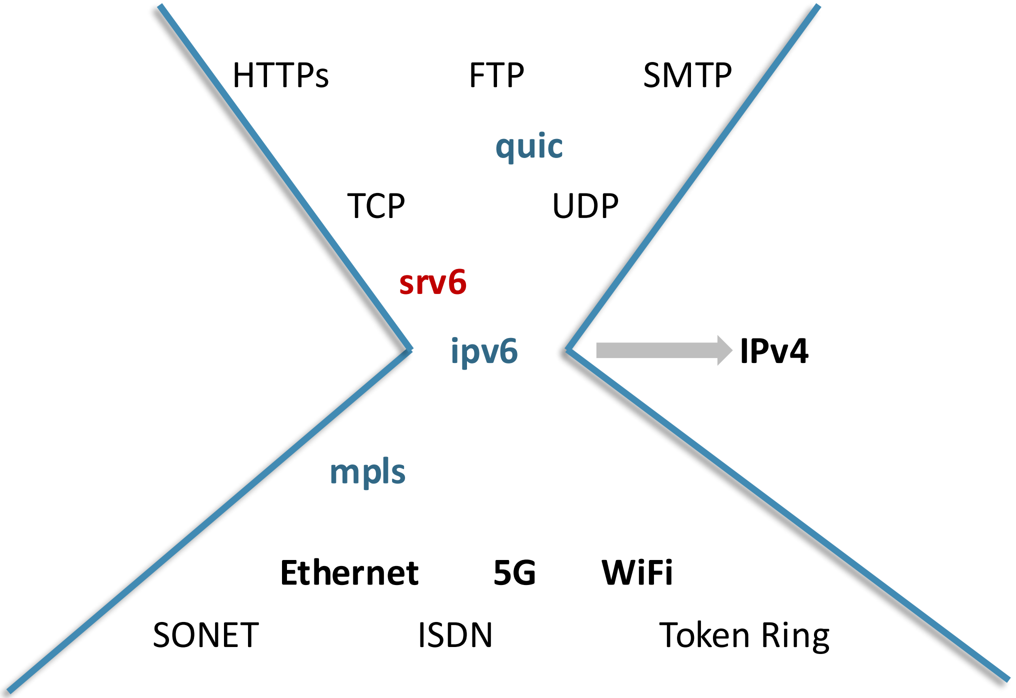
Maybe we can go *above UDP*

- QUIC
- Improve bandwidth utilization and performance

Good for some applications

Not close enough to the universal choke point to be *universally* effective





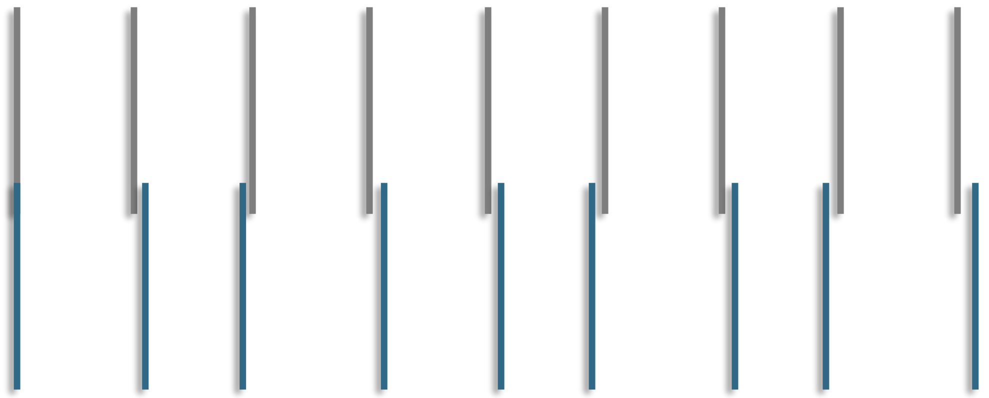
Maybe using the “big IPv6 address space” for traffic engineering will work?

- SRv6
- Additional traffic engineering state can be localized

Will this work?

- *... we will see ...*
- ... this *seems to be* the closest we’ve come to optimizing the local/global tradeoff in a meaningful way

MPLS



Maybe bandwidth isn’t the problem we need to optimize for?

Jitter is the problem

- Increased bandwidth util increases jitter

Solutions?

- fake it
- cache it
- work with queues
- queue elimination
- traffic steering

JITTER

COMMON SOLUTIONS

- Fake it
  - Terminate sessions close to the sender and receiver
- Cache it
  - Cache data close to the end user
- Neither of these
  - Seem to apply to the kinds of high speed problems being addressed here
  - Seem to work well with end-to-end encryption
    - *(though work is ongoing)*

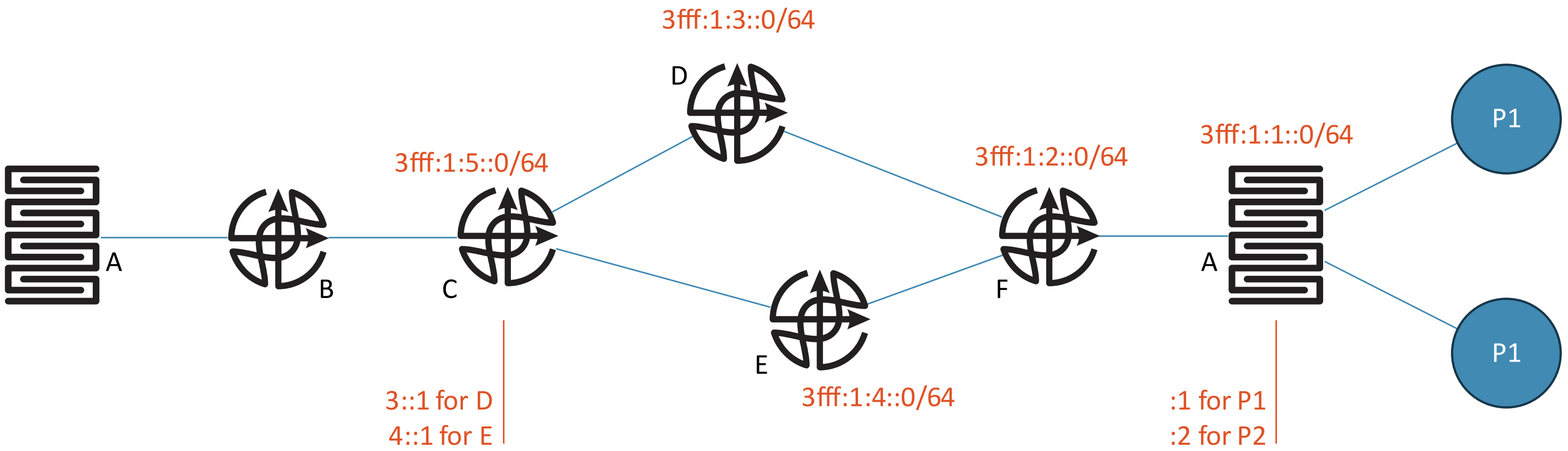
Work with queues

- BBR versus LEDBAT++ (QBit)
- Largely via QUIC
- Avoid buffer bloat

all of these compliment traffic steering

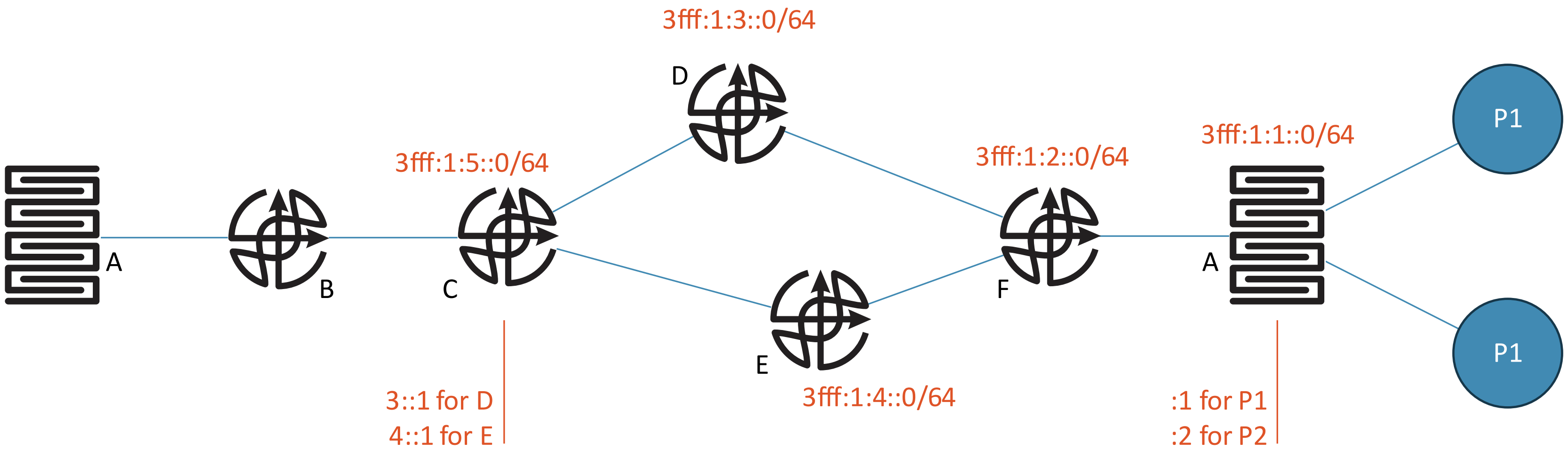
- *let's dive into traffic steering a little more deeply*

SRv6 STEERING



- A can send packets to 3fff:1:1::1 for P1 and 3fff:1:1::2 for P2
    - moves the service identifier into the address space
    - can be used for service chaining, for instance
- A must know about these separate addresses
    - additional state
        - add state to increase optimization
    - does not need to be in the routing protocol
        - distributed database, DNS, many other solutions

SRv6 STEERING

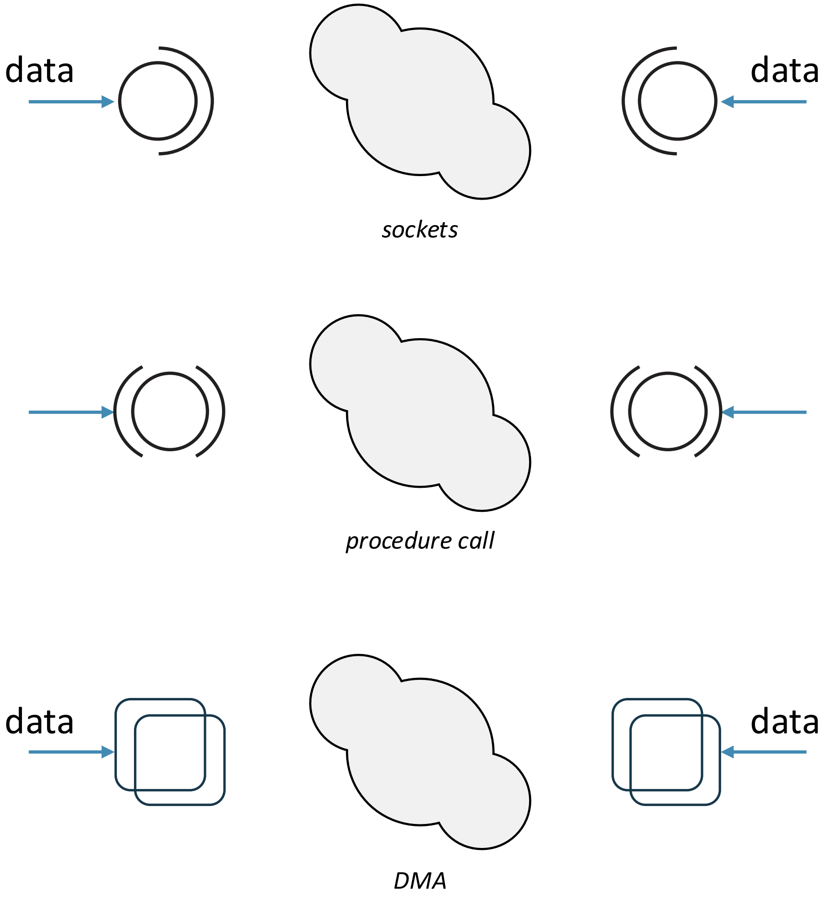


- A can send packets encap’d to 3fff:1:3::1 to push traffic through D
  - D removes outer header and forwards based on inner header
- A can send packets encap’s to 3fff:1:4::1 to push traffic through E
  - E removes outer header and forwards based on inner header
- Policies at D and E are simple
  - Just remove the outer header and forward like any other tunnel
  - Effectively IP-in-IP tunneling

Simplifying Assumptions

- Work with the existing address space
  - Repurpose “slack” addresses within the existing space
  - Set aside for autoconfiguration
  - Repurposed to represent “services”
- Do not change fundamental routing
  - Largely opaque from the network’s perspective
- Directly expose state/optimization tradeoff
  - Increasing steering specificity requires increasing state

SRv6 STEERING

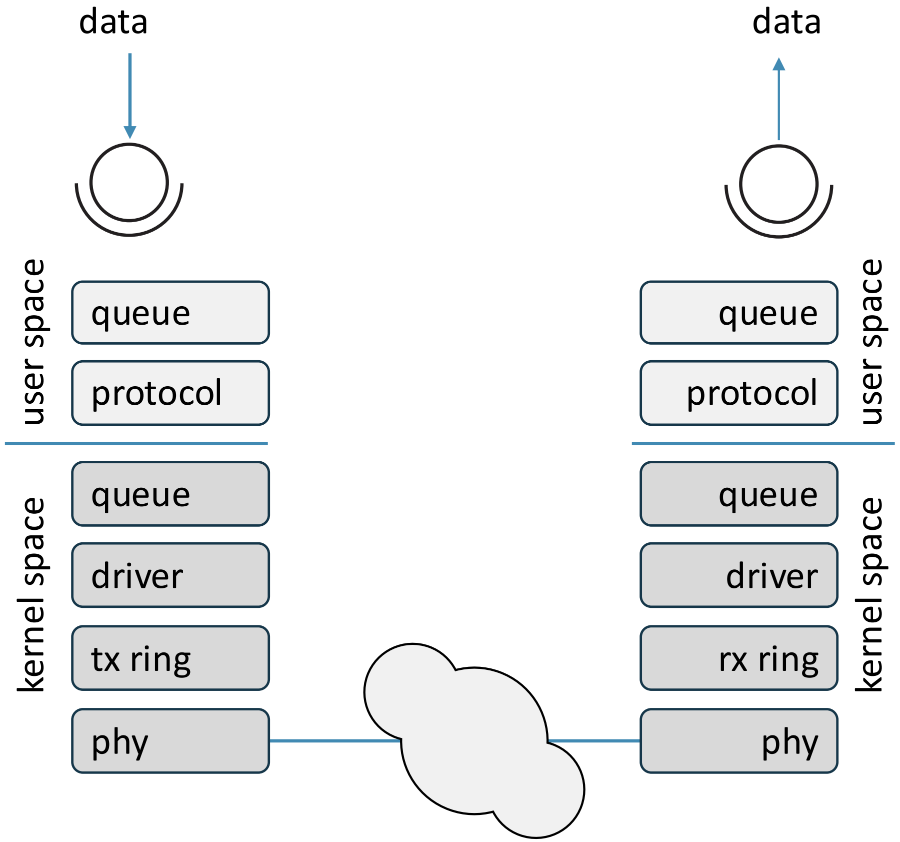


We can also just eliminate the queues and protocol stack ...

Remote DMA (RDMA)

Three ways to model data transmission

RDMA



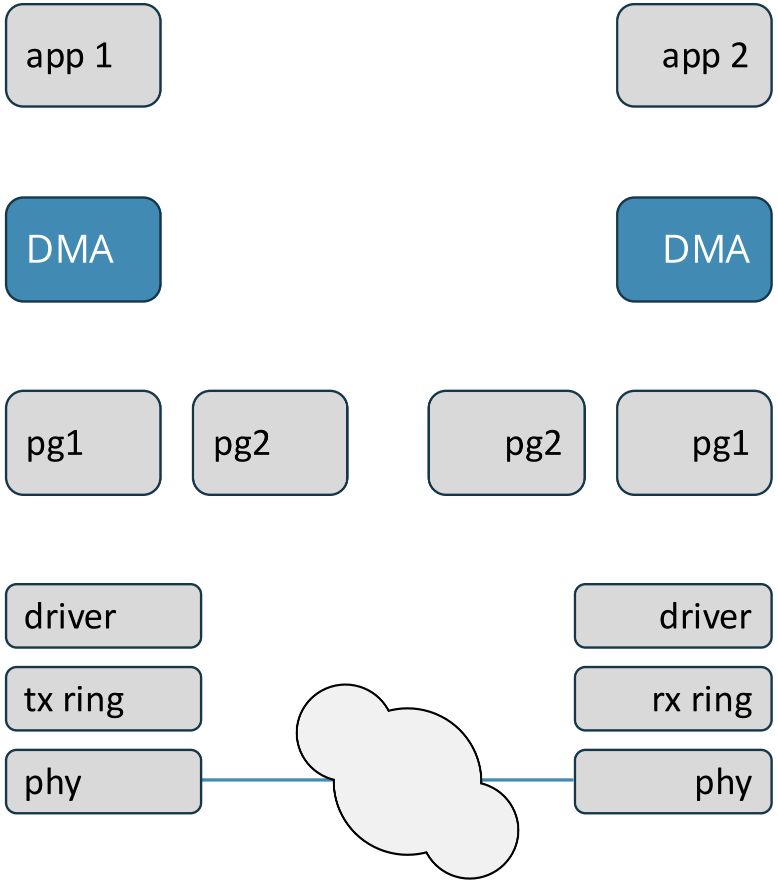
Sockets

- put the network in the queue
- serial stream
- cross the user/kernel space divide

RPC

- put the network in the function call
- call/return
- cross the user/kernel divide

RDMA

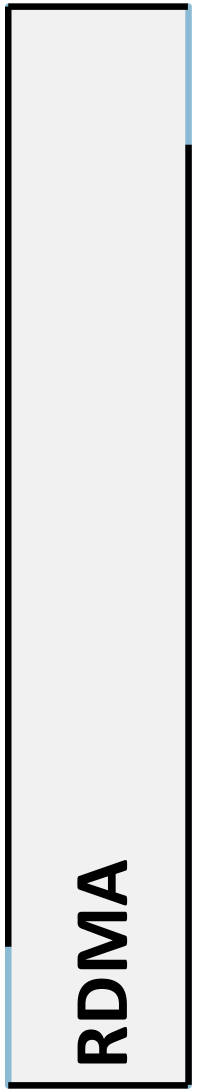


Puts the network in the virtual memory page

Read and write directly to virtual memory locations from user space

RDMA

- Bypasses all the functionality of the network stack
  - Multihop routing
  - Traffic steering
  - Error control
  - Flow control
- *Probably* not good for
  - More than a few hops
  - Anyplace with drops, out of orders, etc.
- Heavy local optimization
  - Not good for global transport use


RDMA

- Complexity limits transport options
  - Global/local
  - State/optimality
- It’s hard to replace the middle of the wasp waste
- SRv6 and RDMA
  - Seem like good candidates for the future of transport in different spaces
  - One size does not fit all

TRANSPORT & COMPLEXITY

Russ White
November 2024

russ@riw.us