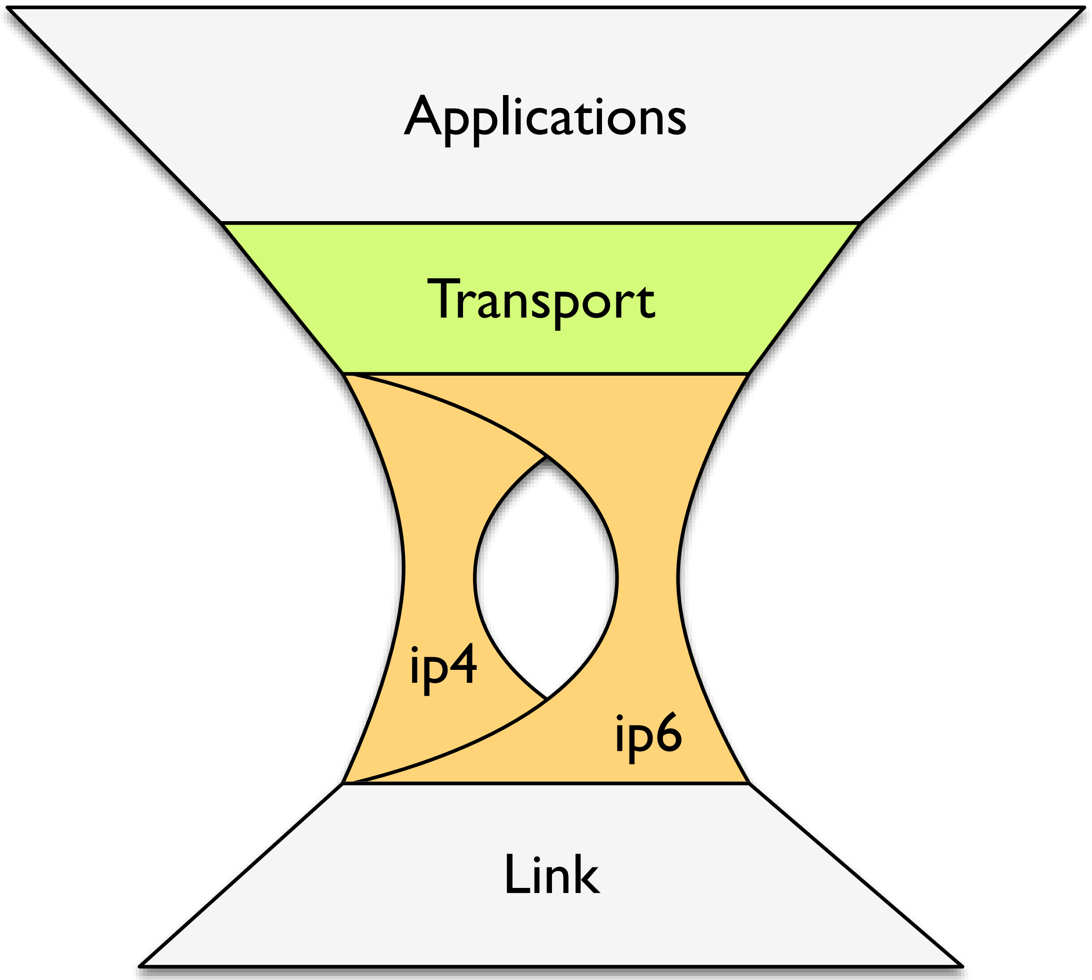
IP Stack Evolution Program

IETF 91 Technical Plenary, Honolulu
Joe Hildebrand and Brian Trammell

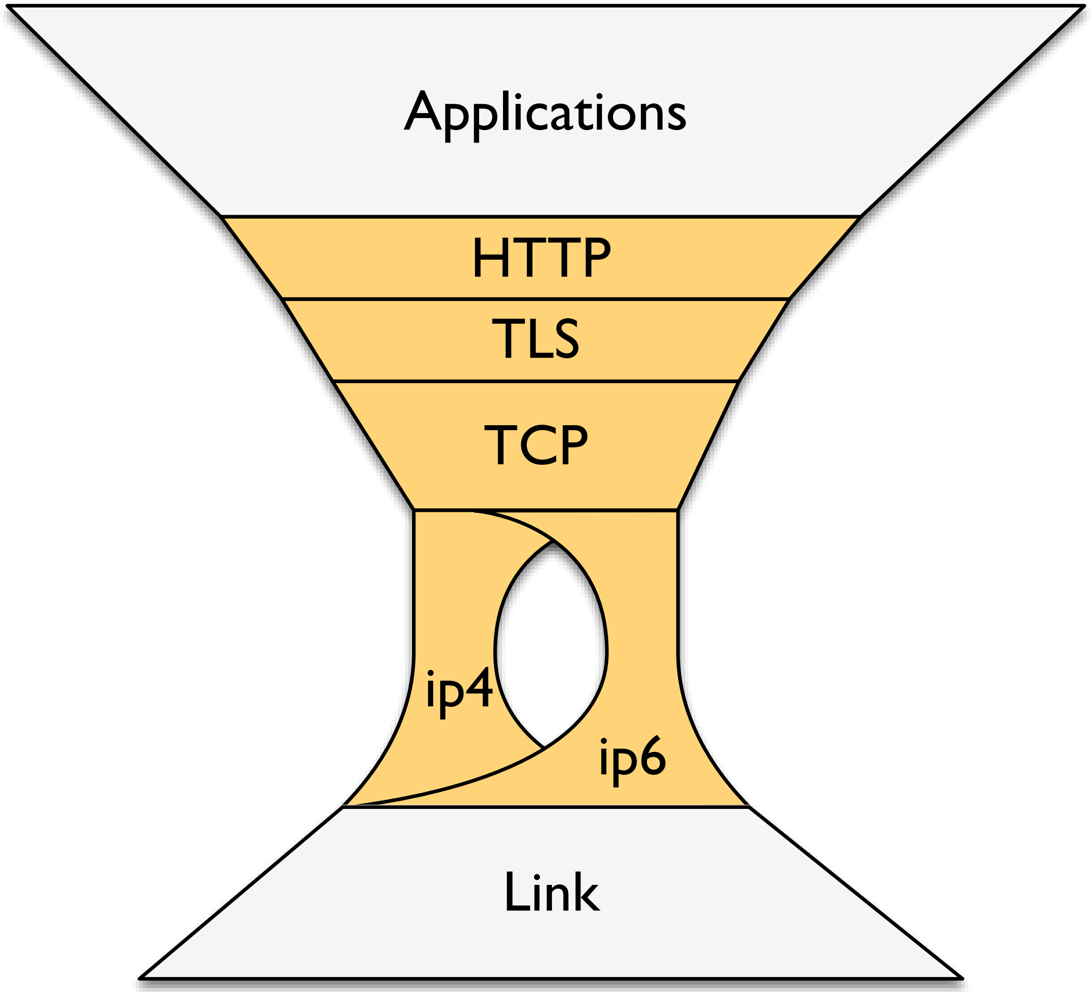
A Taller, Thinner Hourglass

- We've evolved IP to have a dual stack waist…
- …but this picture is decreasingly accurate above the network layer.



A Taller, Thinner Hourglass

- HTTP (+TLS) is a universal session layer
  - Driven by endpoints (browsers as front-end) as well as the network (HTTP-/proxy-only connectivity)
- HTTP implies TCP, which is not always what we want.
  - Transport stagnates, or innovation happens beyond the stack.

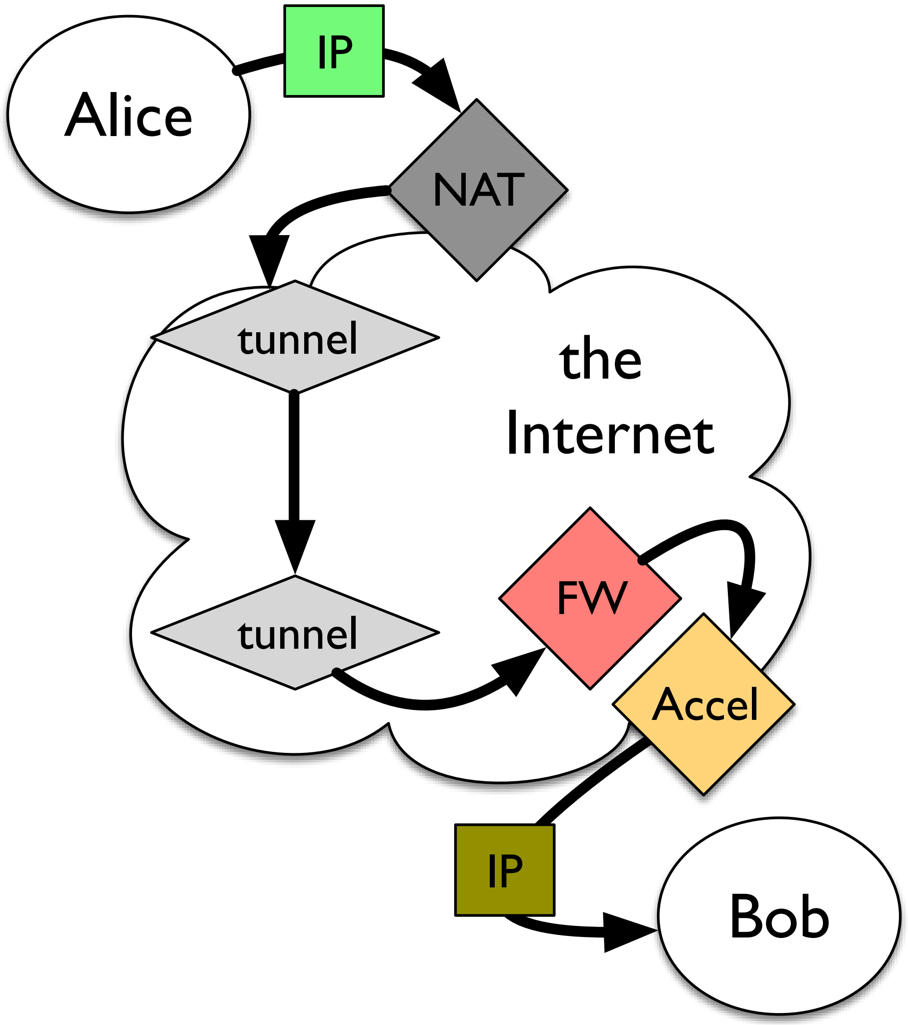


Narrow Interfaces

- `fd = socket()`: the network is a special kind of file. But…
  - `SOCK_STREAM:` single-streaming, full reliability, head of line blocking, no record boundaries…
  - `SOCK_DGRAM:` record-oriented transport, no reliability, MTU issues…
  - All other transports require encapsulation
- Identifier bound to location: roaming is “difficult”.
- Security bolted on, and SSL APIs can be “unique”.
- Kernel/userspace boundary is in the wrong place.

Reasonable Brokenness

- Even if you fix the interface, there's still the little matter of middleboxes.
- Even the worst middlebox isn't *evil*
  - It's there on purpose
  - It's solving a problem
  - Wishing won't make that problem go away
- *Can we make everyone happy **without** breaking end-to-end?*



How Broken is Broken?

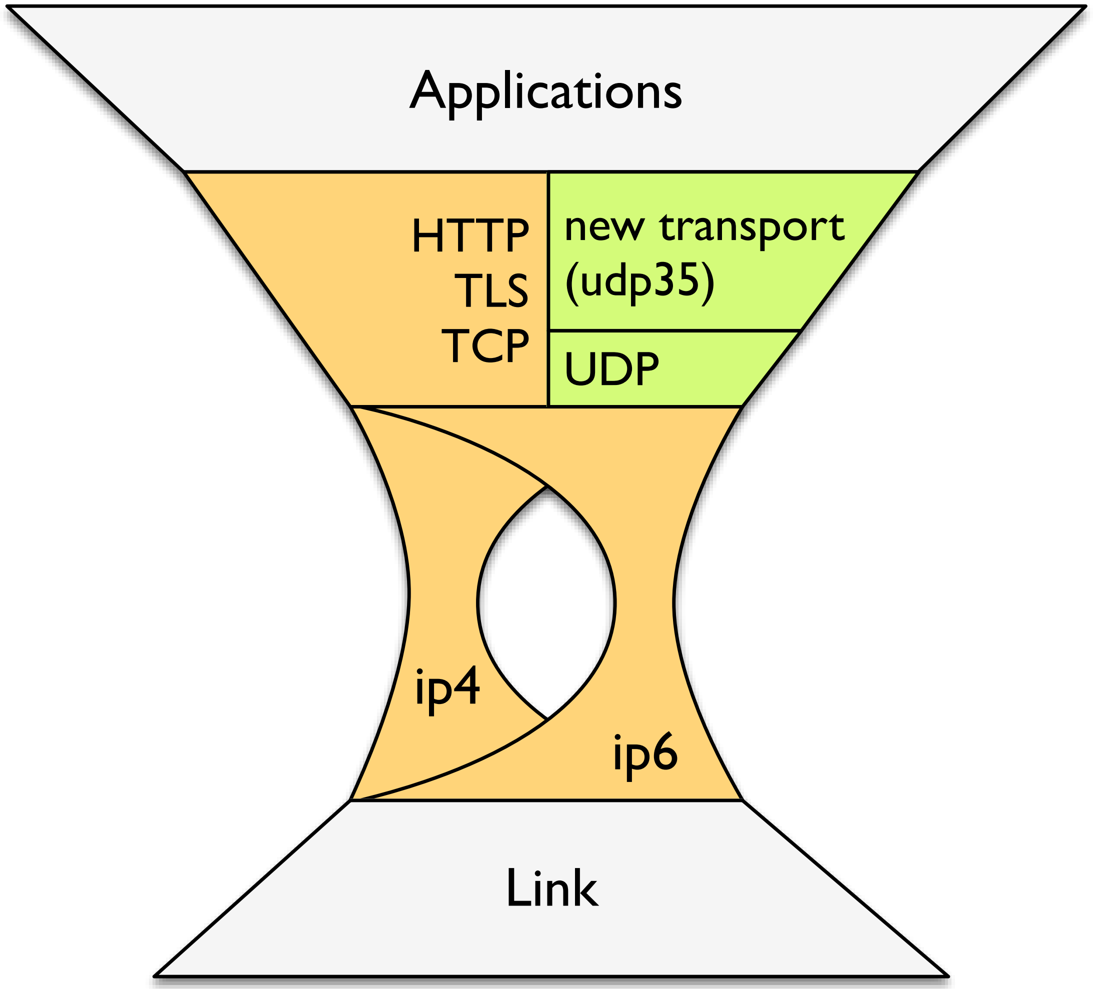
Modification	**Planetlab**	**Ark**
NAT	**74.9%**	**79.0%**
ECN IP	**13.7%**	**13.2%**
ISN	**10.7%**	1.8%
MSS	**10.8%**	5.9%
Exp. Option	8.8%	0.5%
MPCAPABLE	8.4%	0.3%
ECN TCP	0.6%	0.6%
SackOK	0.3%	0.0%
TS	0.3%	0.4%
WS	0.2%	0.2%

- Majority of paths on two well-known testbeds NATted.
- Certain features work unmolested 9 times out of 10
- Variation based on vantage point

Source: R. Craven, R. Beverly, M. Allman. **A Middlebox-Cooperative TCP for a non End-to-End Internet**. ACM SIGCOMM, August 2014.

An Initial Proposal

- UDP gives us a partial defense against middleboxes, provides port multiplexing, and works from userspace.
- Building recommendations / “mix-ins” for transport atop UDP will make this work better.
  - Congestion control in particular is hard to do well.
- Provide hooks for policy decision (e.g. enterprise firewall traversal)
- Allow evolution beyond and coexistence with “Internet over HTTP”.



Stack Evolution Program

- Formed to provide guidance and coordinate effort toward breaking this logjam:
  - Evolving interfaces to transport and network-layer services beyond the traditional `SOCK_STREAM` and `SOCK_DGRAM`.
  - Improving path transparency in the presence of firewalls and middleboxes, including guidelines for the detection of and cooperation with these devices to evolve away from present limitations on which protocols can be used across network boundaries.

Current work in the IETF

- Transport Services (TAPS) WG
  - define transports in terms of the services they provide and provides endpoint-deployable approaches for protocol selection
  - chartered simultaneously and in concert with stackevo
- TCP Increased Security (TCPINC) WG
  - may require additional options space in TCP header, leading to discussions of how to provide this within TCPM.
- Advanced Queue Management (AQM) WG
  - Work on ECN deployability
(it's not much path signaling, but it's a start)
- Looking for more applications area participation!

SEMI

- IAB Workshop on Stack Evolution in a Middlebox Internet (SEMI), Zürich, 26-27 January 2015
- Aim: get people from research and industry working in this space together to refine the scope and solution space considered by the program
- Workshop report will follow in the Dallas IETF timeframe.

Program Documents

- draft-iab-filtering-considerations: inherited from IP Evolution program, comments under review
- draft-blanchet-iab-internetoverport443 to evolve into statement on architectural considerations for HTTPS as a transport
- draft-eggert-tsvwg-rfc5405bis to evolve into statement on architectural considerations for UDP transport encapsulation