Platforms and Tools for Internet Measurement: Current and Future Developments

Brian Trammell

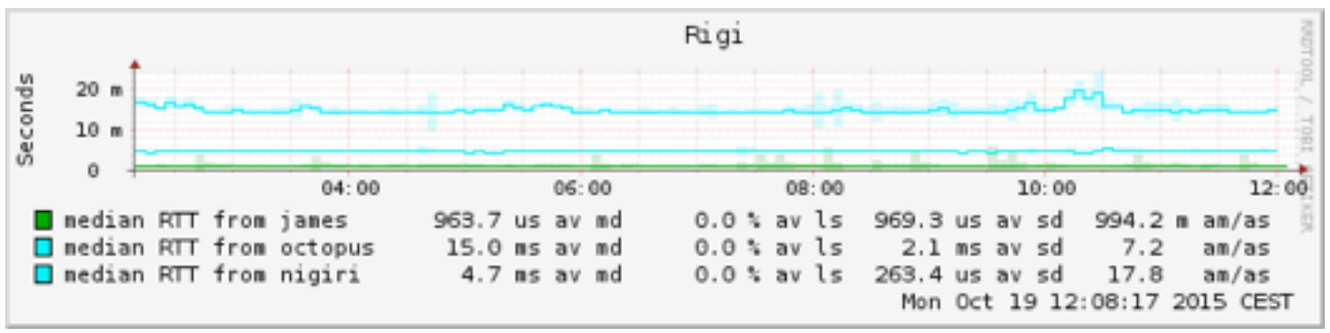
IRTF/ISOC Workshop on Research and Applications of Internet Measurements
Yokohama, Japan, 31 October 2015

In the beginning…

- …there was `ping`, and it was good.
  - (still the only explicit measurement facility in the stack.)
- Periodic measurement via `cron`
- Visualization and storage with `rrdtool`
- ~~Distributed measurement via `telnet`~~
- Distributed measurement via `ssh`
- Glue everything together with `perl`

Echo or Echo Reply Message

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Code      |          Checksum             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Identifier          |        Sequence Number        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Data ...
+-+-+-+-+-
```



- Actually, this is pretty much SmokePing.

Overview

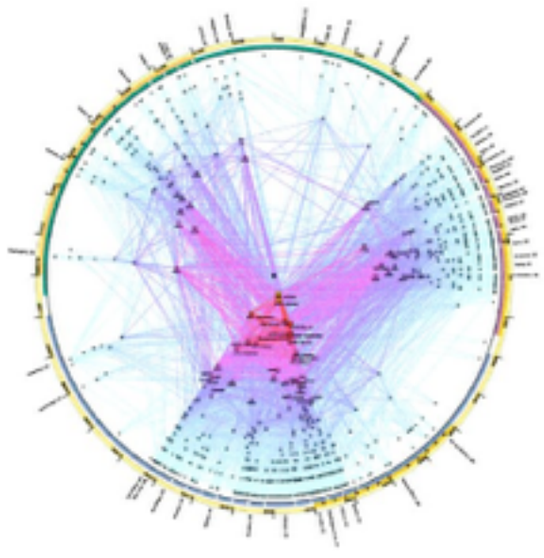
- Dimensions of work in tools and platforms
- State of the world (illustrated with a current project)
- Musings on the bright shiny future

Different views of the Internet

- Topology and (intra-,inter-)domain routing
- Addressing and naming
- End-system and infrastructure security
- Data plane performance and impairment
- Traffic characterization

Different reasons to measure

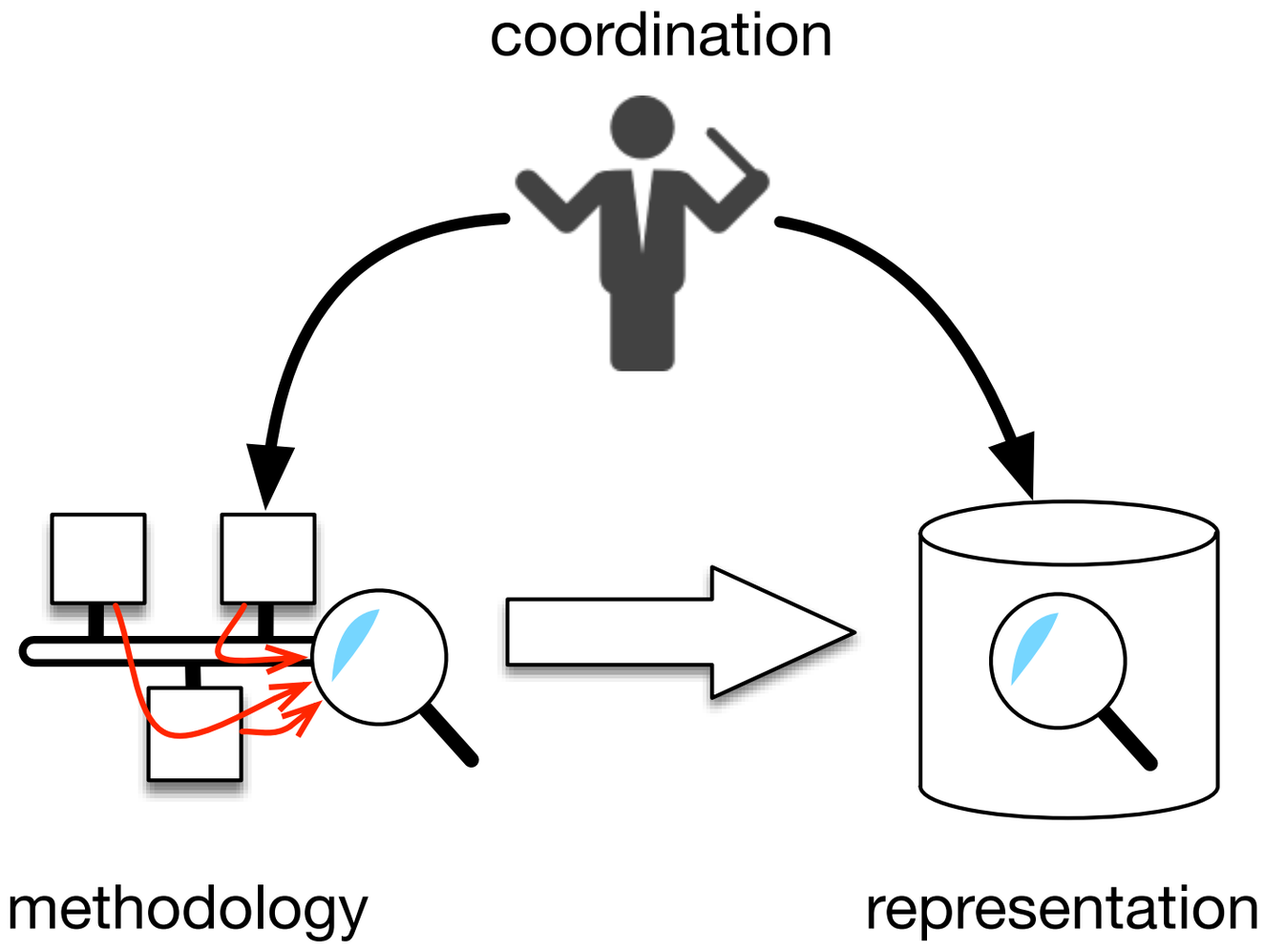
- **Operations**: keep the Internet running, efficiently
  - "What's broken?" (or "who's attacking me?")
  - "Why is it broken and how can we fix it?"
  - "Is everything running as we expect it to?"
  - "How should we invest in our network in the future?"



- **Research**: understand the Internet
  - "What does the network look like?"
  - "What will the network look like tomorrow?"
  - "Hm, that's interesting..."
- **Engineering**: support protocol design decisions with data

- Most platforms designed with only one of these communities in mind.

Different areas for improvement



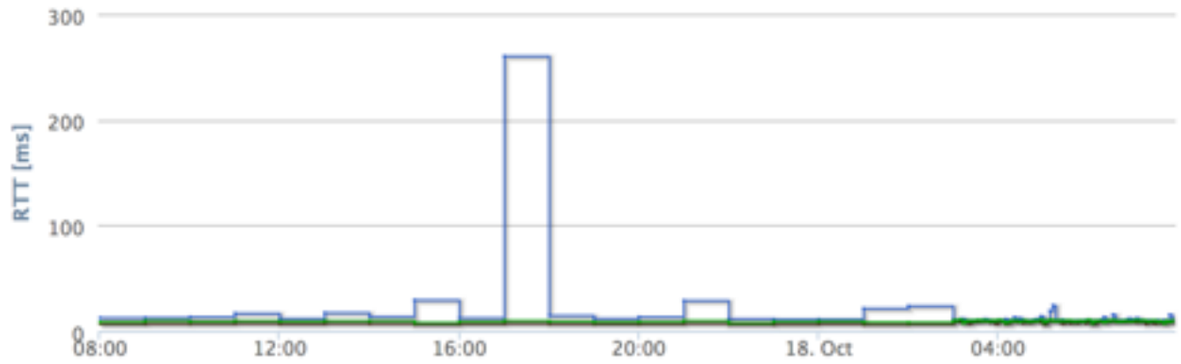
Techniques and Methodology

- `ping` doesn't work everywhere it should
  - ICMP blocking to prevent "reconnaissance"
- It doesn't measure what you think it does
  - ICMP handled by different codepaths/queues
  - ECMP causes different flow labels to take completely different paths
- What it does measure might not be what you want
  - Application latency affected by proxies, transport pacing

- And that's just `ping`.

Analysis and Representation

- Privacy is a problem (even for `ping`)
  - Latency correlates with buffer occupancy correlates with activity.
  - Quiz: find the download
  - “Publish-and-forget” not possible.



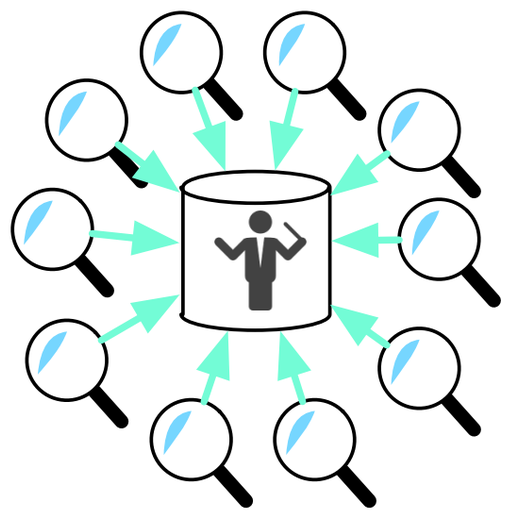
- We lack good standards for data exchange
  - CSV the lingua franca in research
  - Some use of structured data (JSON)
  - Some attempts at normalizing column/element meaning

image: RIPE Atlas

Coordination and control

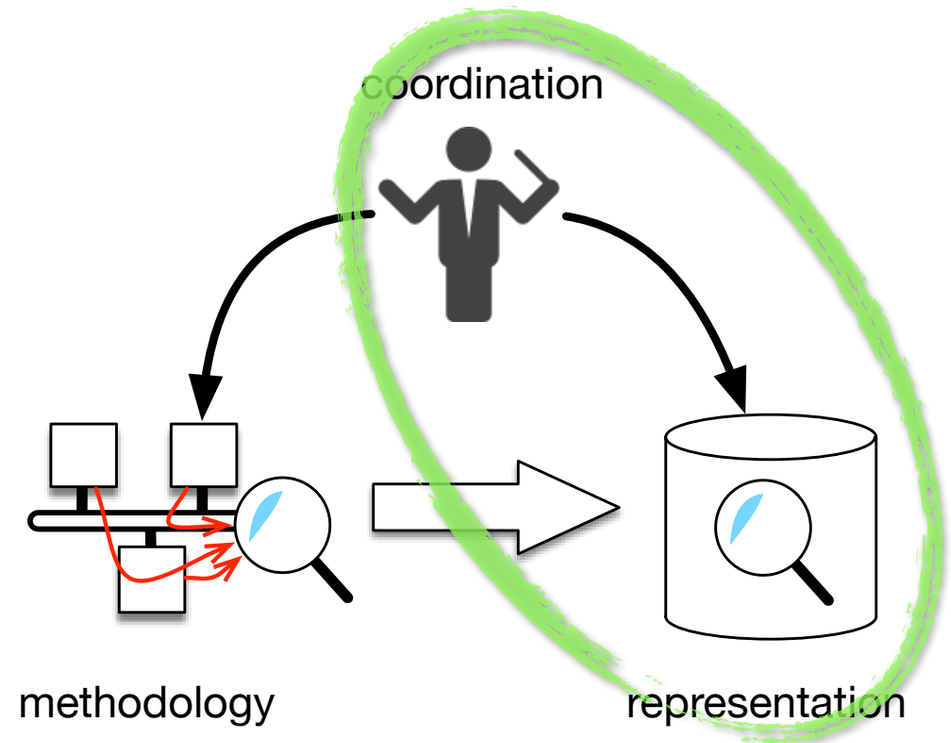
- Single-point measurements are of limited use to understand what's happening on a network.
  - Difficult problems in operations are distributed
  - Internet is heterogeneous
- New tool development should happen with this in mind.

- Currently: centralized architectures for coordination.
  - A surprising amount of effort goes into device management.



Toward platforms for measurement

- Methodology: painstaking attention to detail
- Coordination: allow methodology to scale
- Representation: make measurement universal
- A successful platform is the product of a coherent approach to the latter two areas.

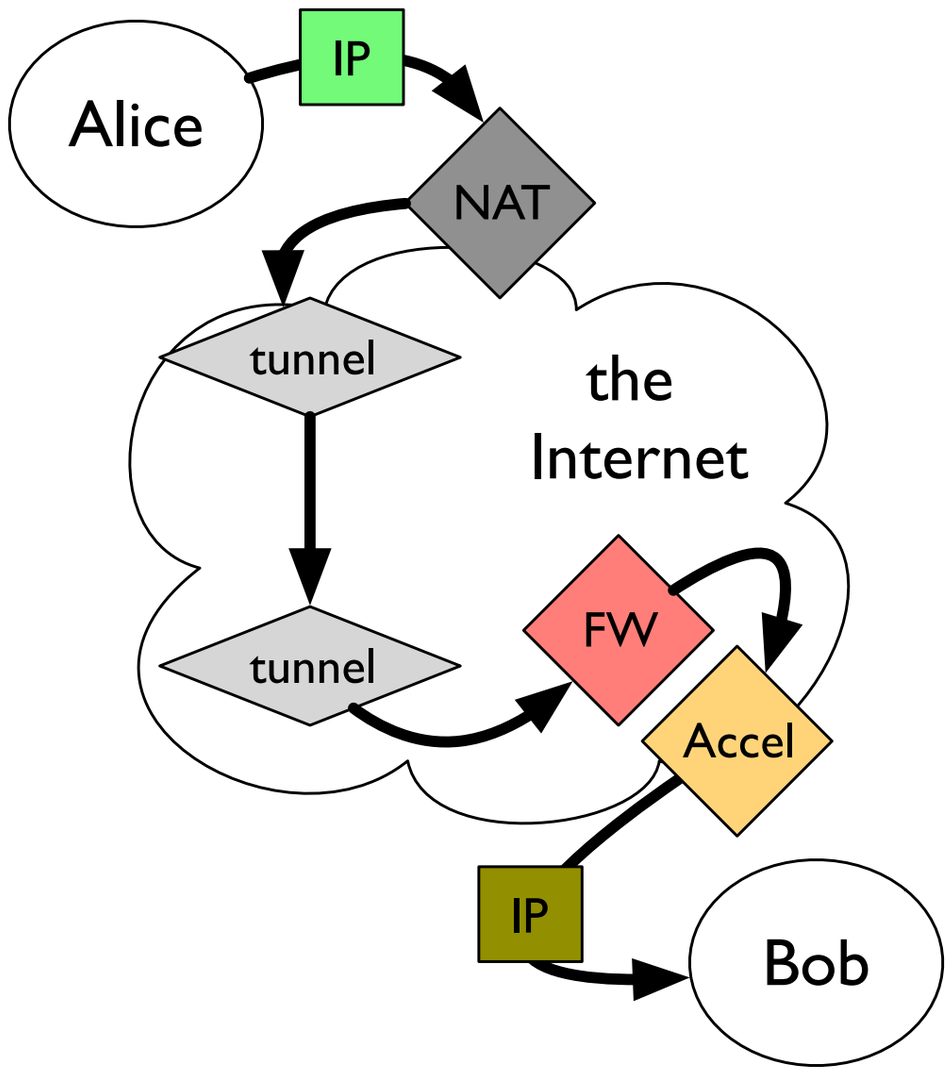


Techniques:

Path Transparency

Path Transparency
(in one slide)

- The Internet is not end-to-end...
  - some of this is policy, but a lot of it is accident
  - deployment of new protocols over IP, transport extensions difficult or impossible
- ...but some paths are worse than others.
  - Goal: data on "how bad" and "where" to guide future protocol design
  - Connectivity impairment
  - Latency and loss differences
- Interested? **HOPSRG** (hops@ietf.org) (Monday, Room 303 (you are here)).



What can go wrong?

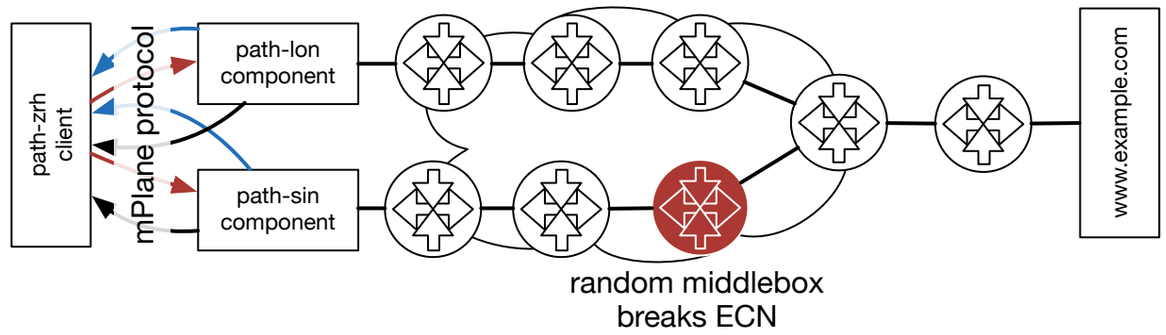
Modification	Planetlab	Ark
NAT	**74.9%**	**79.0%**
ECN IP	**13.7%**	**13.2%**
ISN	**10.7%**	1.8%
MSS	**10.8%**	5.9%
Exp. Option	8.8%	0.5%
MPCAPABLE	8.4%	0.3%
ECN TCP	0.6%	0.6%
SackOK	0.3%	0.0%
TS	0.3%	0.4%
WS	0.2%	0.2%

- NAT everywhere
- Many features mostly work
- Variation based on vantage point
- **Best studies look at O(10k) paths[1].**

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

Measuring Transparency and Impairment

- Lots of tools for doing this:
  - `tracebox`: localize packet modification along a path.
  - `pathspider`: find path-dependent impairments via A/B testing.
  - Anything that can put arbitrary packets on the wire: `nmap`, metasploit, scapy.



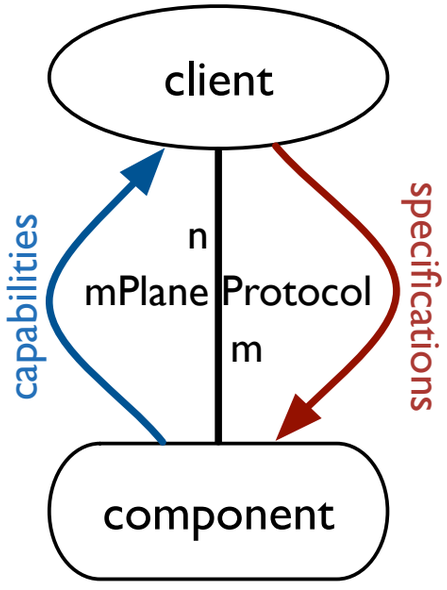
- But impairments aren't just "weird packets get dropped":
  - How much slower are UDP-encapsulated transports than TCP transports? Is the Internet even UDP-transparent?
- The Internet is heterogeneous.
  - look at as many paths as possible.
  - common representation to compare studies with different tools.

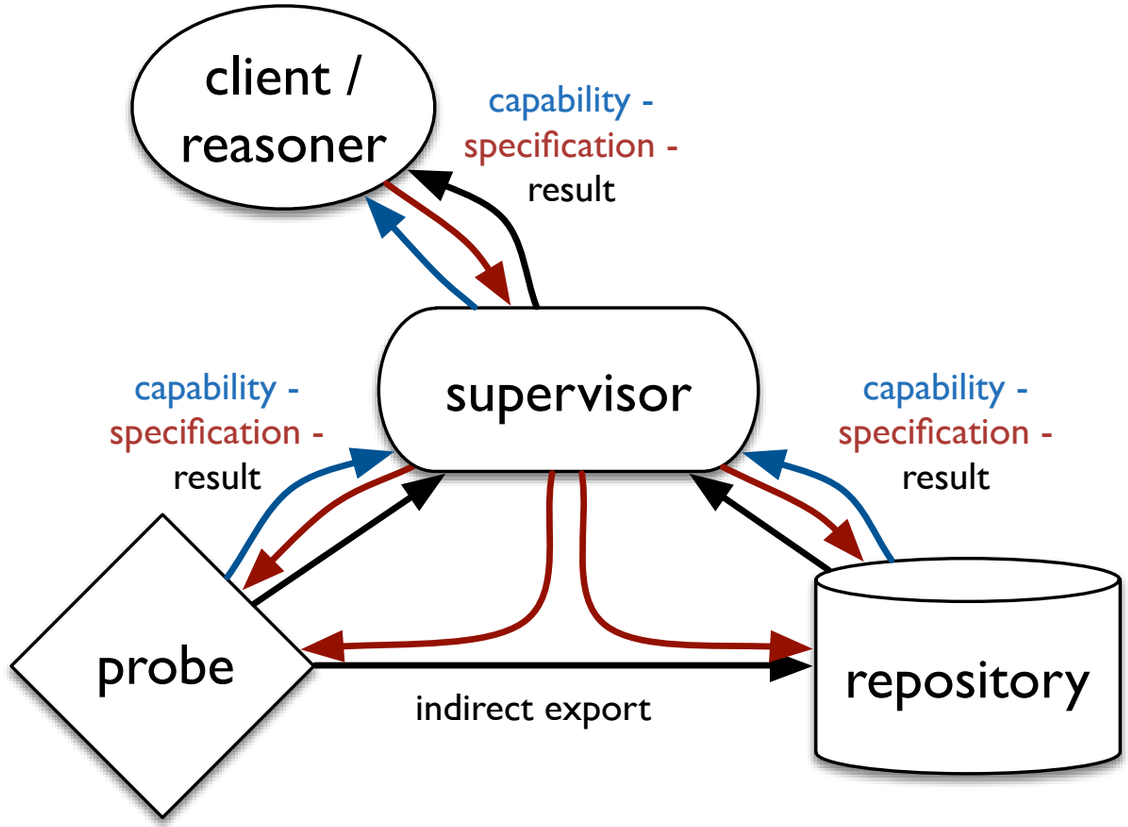
Coordination and Control: Applying mPlane

mPlane

(in one slide)

- Self-descriptive, error-tolerant RPC protocol connecting *clients* with *components* to cooperatively perform network measurements and analysis using **heterogeneous** tools.
- Measurements and analyses described using *capabilities* containing measurement *schemas* defined in terms of a registry of *elements*.
- **Schema defines the measurement** to perform.
- *Supervisors* knit larger infrastructures of components together.





Architectural Principles

- **Schema-centric measurement definition**: a measurement is completely described by the parameters it takes and the columns in the results it produces.
- **Weak imperativeness**: capabilities aren't guarantees, normal exceptions discovered in later analysis, state and responsibility dynamically distributed throughout an infrastructure.
- Component management left out of scope
  - assume components too heterogeneous anyway.

Schema-centric measurement definition

- Traditional RPC:

```
ping -c 3 -w 5 10.2.3.4
ping(count, period, dest) => [int]
```

  - Need to register entry points, argument names.
  - "Can I compare `ping()` to `webping() to nmap_christmas_tree_warning_very_beta()`?"

- Schema-centric:

```
measure(param(singleton_measurement_count,
              period,
              destination_ip4);
        result(delay_oneway_icmp))
```

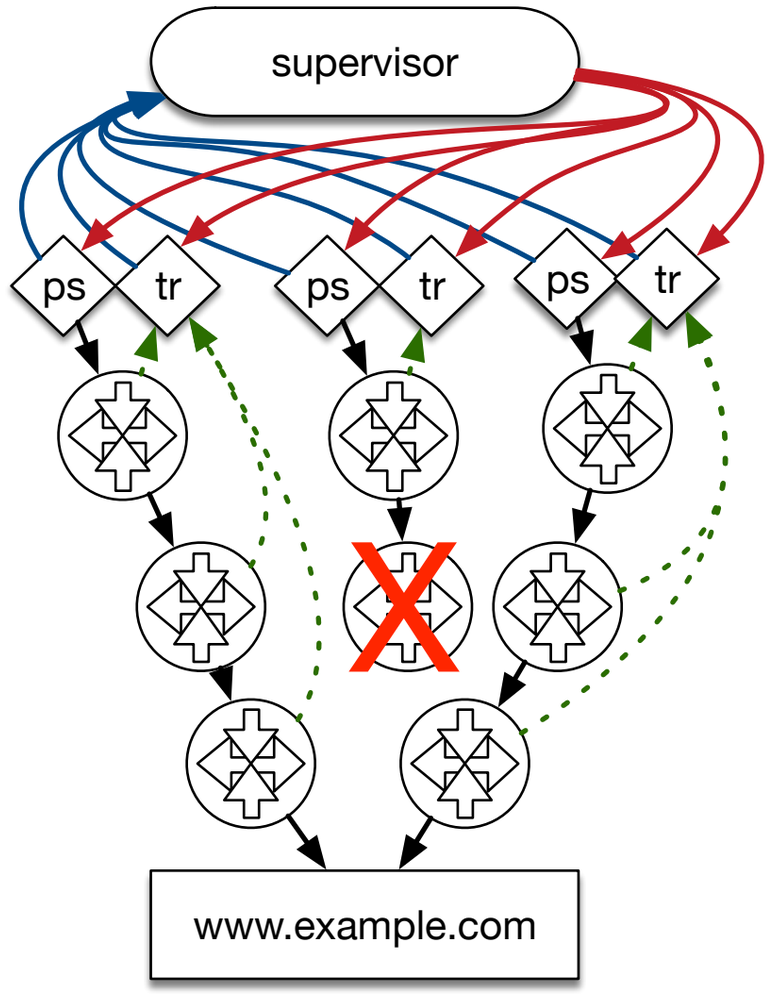
- Requires rigorous control over the set of column names, but allows more or less infinite combination (cf. www.iana.org/assignments/ipfix)

Weak imperativeness

- **Failure is inevitable. Embrace it.**
- Two kinds of failure:
  - Things that are part of what you're measuring (e.g. variable connectivity on mobile probes)
  - Things that need a forklift to fix.
- For the second class, you need completely separate infrastructure monitoring anyway.
- For the first class, export enough metadata to allow analysis *as part of the normal measurement workflow.*

Applied to path transparency

- mPlane-based `pathspider` tool connects to set of targets with feature enabled and disabled.
- `pathspiders` at multiple vantage points find path dependency.
- Triggers `tracebox` to localize impairment.
- mPlane enabled easy integration.



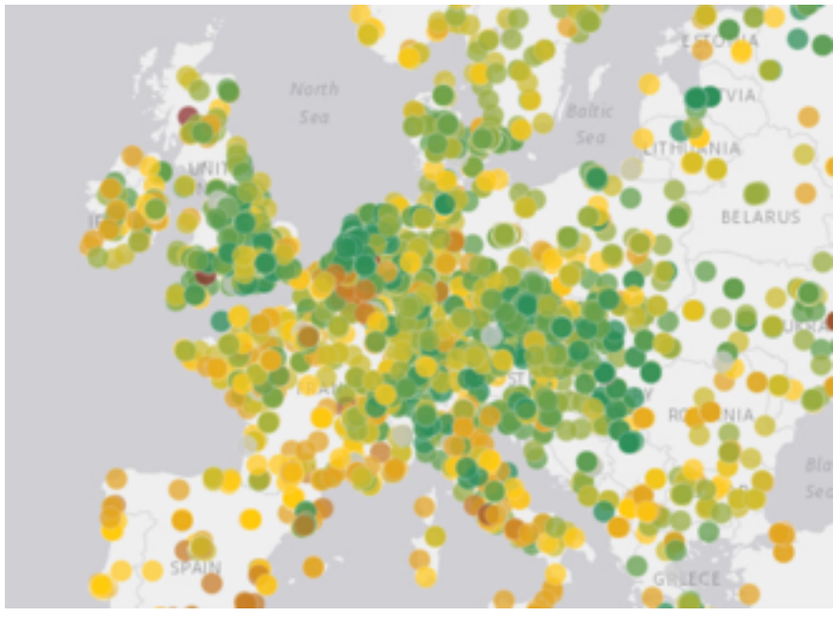
Lessons Learned

- The architecture is experimental in nature:
  - Weak imperativeness is hard to get used to.
  - Schema-centric measurement definition replaces one hard problem with another.
- Managing a PKI is way harder than it needs to be.
- Device management more in scope that we thought.
- but mPlane is a "platform toolkit" instead of a platform at this stage in its development
  - Few vantage points (ECN: *n*=5)

Scaling Up:
RIPE Atlas

What is RIPE Atlas?

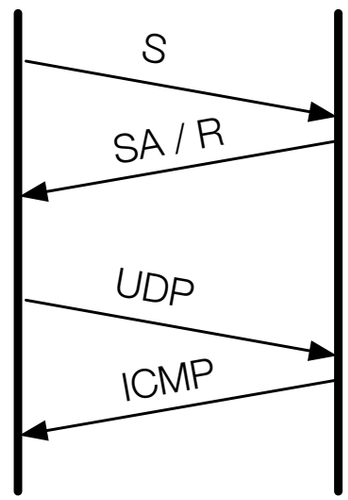
- Active measurement platform using ca. 8,500 distributed probes connected to volunteer networks, under active development.
- Operationally focused: ping, traceroute, HTTP, TLS certificate, DNS, and NTP.
- Centralized control, storage, API, UI provided by RIPE.
- Database of ~3m measurements, many openly accessible.
- Credit system to encourage probe deployment, limit abuse.



images: RIPE Labs

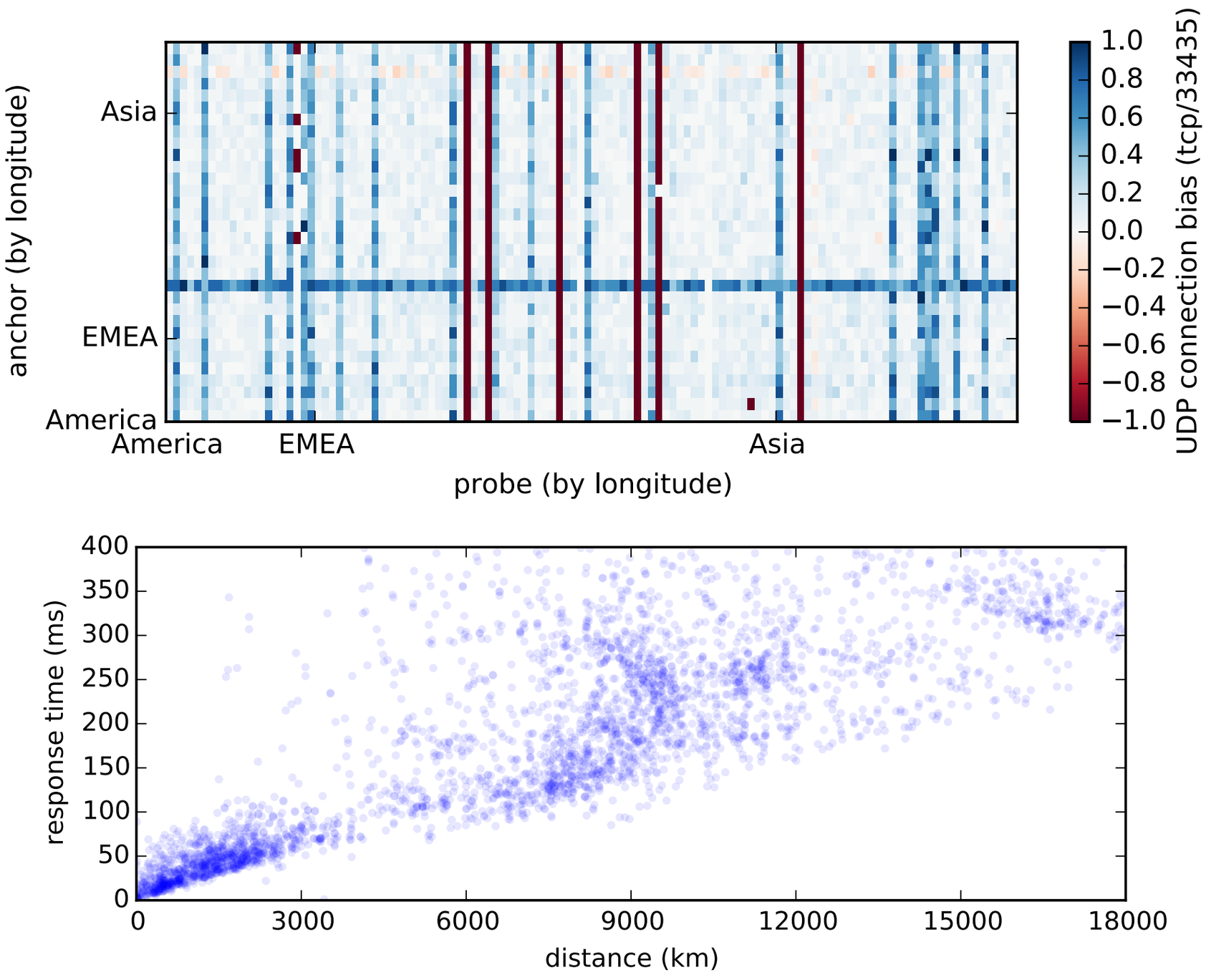
Tweaking Atlas

- Question: “how is UDP treatment different from TCP treatment in the wild?”
- Issue: any sufficiently advanced active measurement platform is indistinguishable from a botnet, so arbitrary connections aren’t supported.
- Hack: Use single-hop traceroute to simulate TCP, UDP connection attempts.
- Attempt UDP and TCP connection simulations “simultaneously”.



- Target Atlas anchors, note return from target (connectivity proxy) and response time (first-packet RTT proxy)

What we found



Open Questions and Next Steps

- Like most Internet measurement, this study consists of the artful stacking of hacks.
  - Does traceroute measure what we really want?
  - Is TCP handshake RTT comparable with UDP to ICMP delay?
    - 8,500 probes is a lot. We chose a diverse sample of 128. Is this diverse enough to generalize to “the Internet”?
- Focus on operations: Atlas evolves, but it is not likely to become the platform we (as researchers) want.
- Ongoing work: integration of Atlas as an mPlane component
  - Amplify information from a few vantage points with context from many.

The ~~Bright, Shiny~~ Future

Recentralization

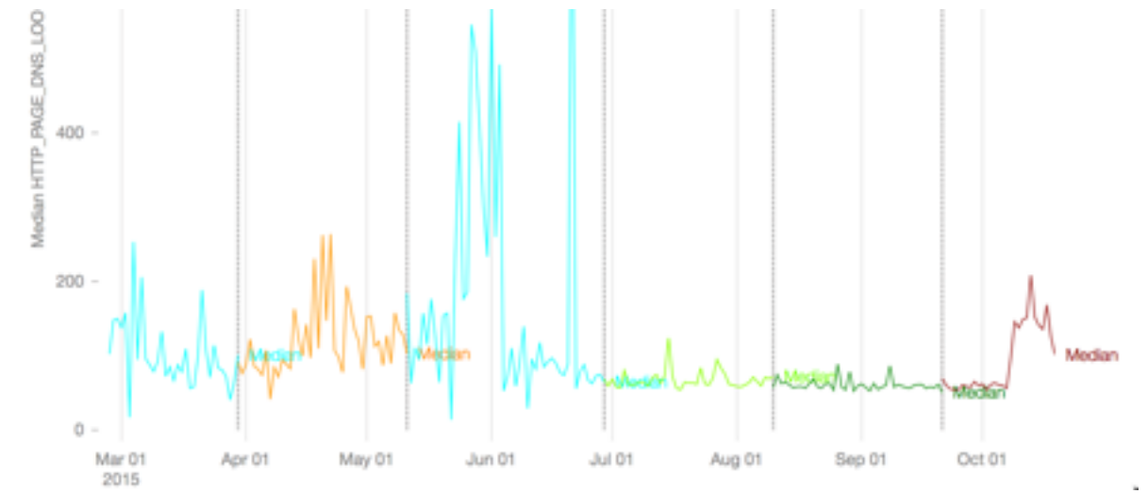
- Very few organizations have the scale to do measurement studies about “the Internet” at large.
- Current work in platforms centralizes development, management, and control.
- Large-scale sharing for network measurement studies is fraught with peril.
  - Any sufficiently advanced passive measurement platform is indistinguishable from the NSA.

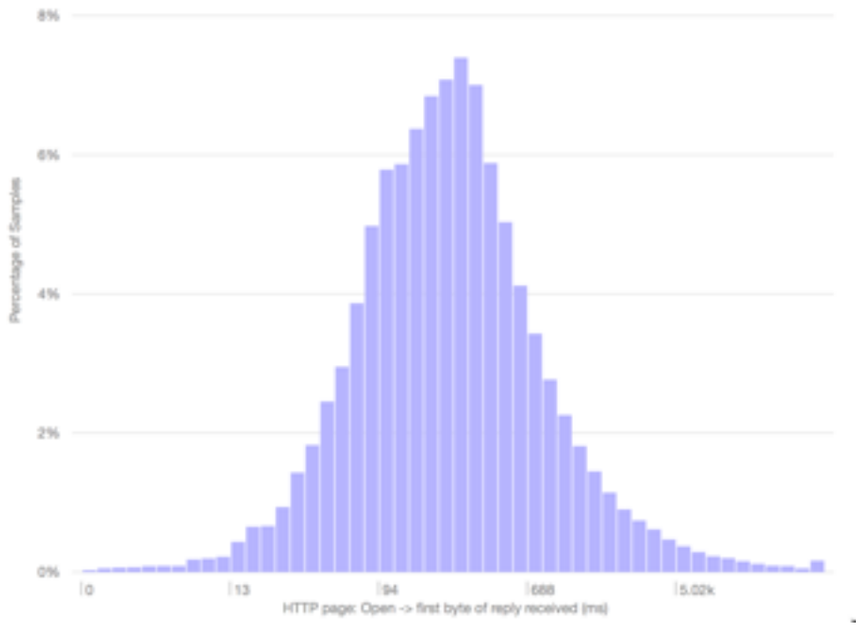
Embracing recentralization

- One way out: repositories and observatories to centralize partially processed data *for a specific purpose*.
- Define an information model all users will share.
- Processes for data contribution and access.
- Community-building around each observatory.
- See our lightning talk later.

Measurement as a first-class operation

- `ping` is the only explicit measurement feature built into the stack. What if we could do better than that?
- Many implementations generate a wealth of information about their operation then stick it in a log nobody ever looks at.
- Phase 1: better instrumentation, out-of-band access.
- (Phase 1.5: open protocols to access better instrumentation)

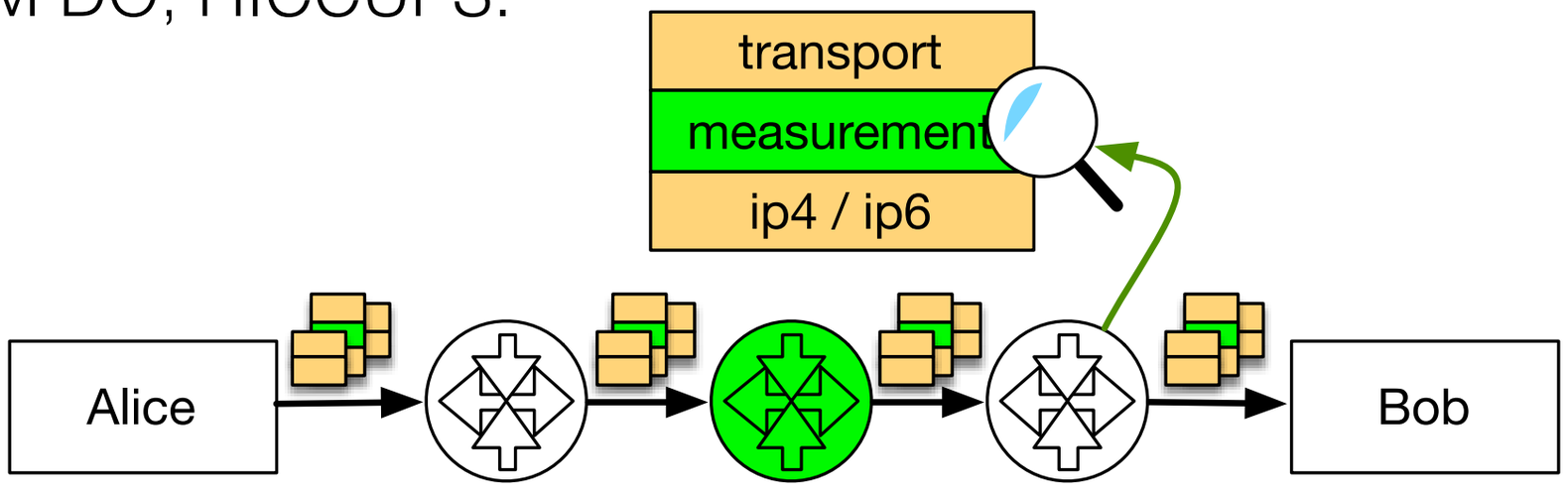




images: Mozilla

In-protocol measurement

- Years of stacking hacks atop one another have made us pretty good at squeezing knowledge out of small datasets.
- Phase 2: apply these insights to explicit exposure of information as part of every protocol exchange in a “measurement header”
  - e.g. IPv6 PDM DO, HICCUPS.

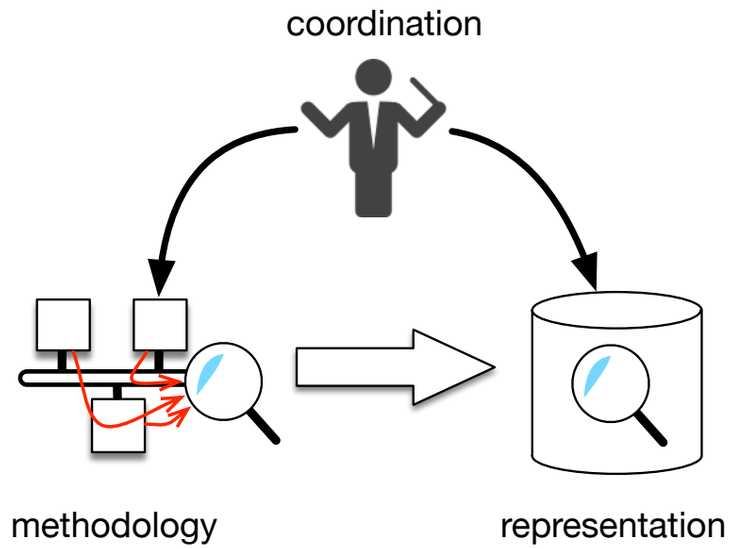


- Selective exposure techniques (e.g. SPUD) provide the infrastructure
- Large numbers of flows means lots of data can be generated with very low sampling rates (i.e., low overhead)

in conclusion...

The Internet is big and measuring it is hard.

- Platform-building: **engineering** that fills the gap between research and practice
  - Coordination: make measurement scale
  - Representation: make measurement portable and universal
- Move toward measurement as a **first-class function** of the stack.



thanks! questions?
<trammell@tik.ee.ethz.ch>