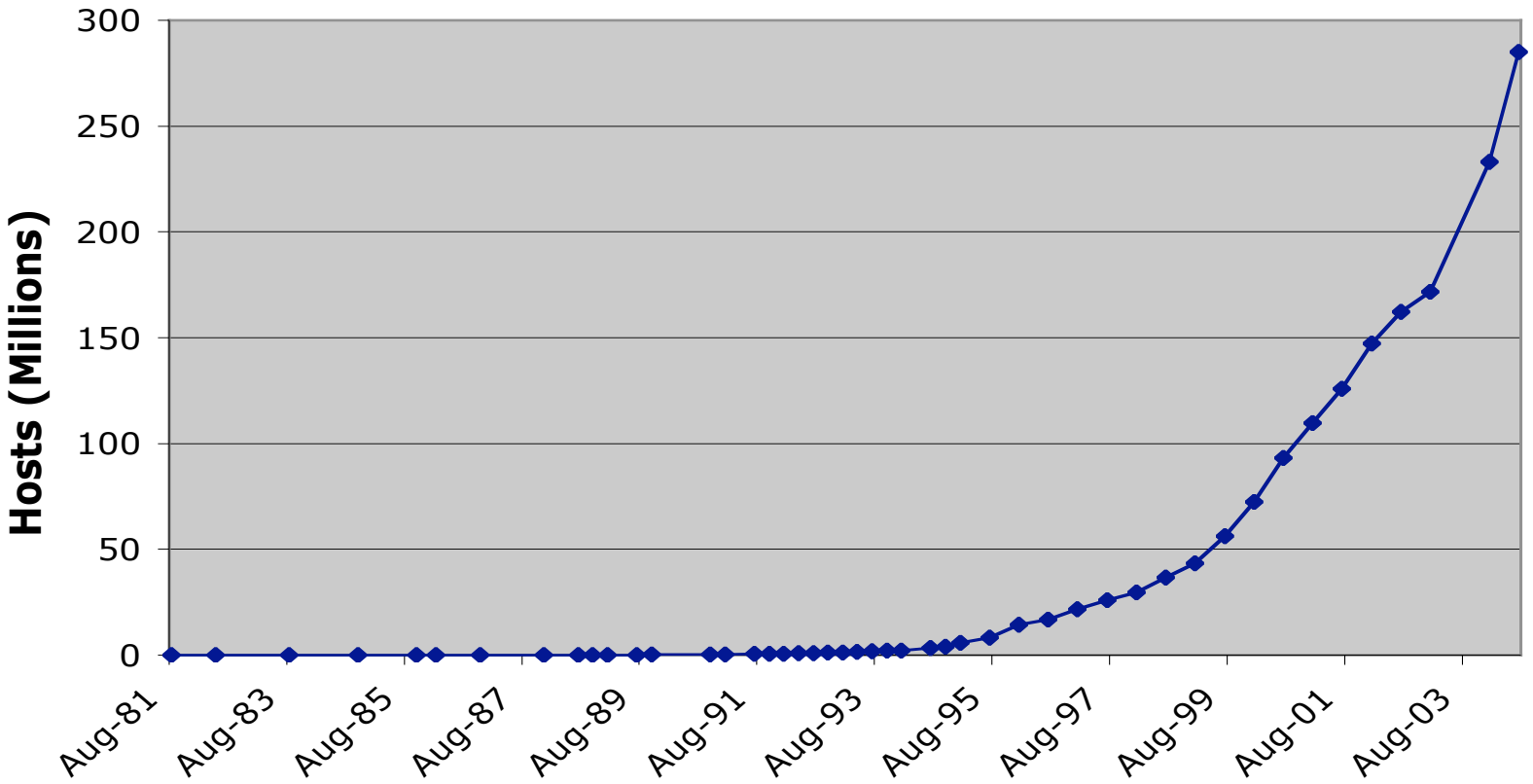
Evolving the Internet Architecture: Challenges and Opportunities

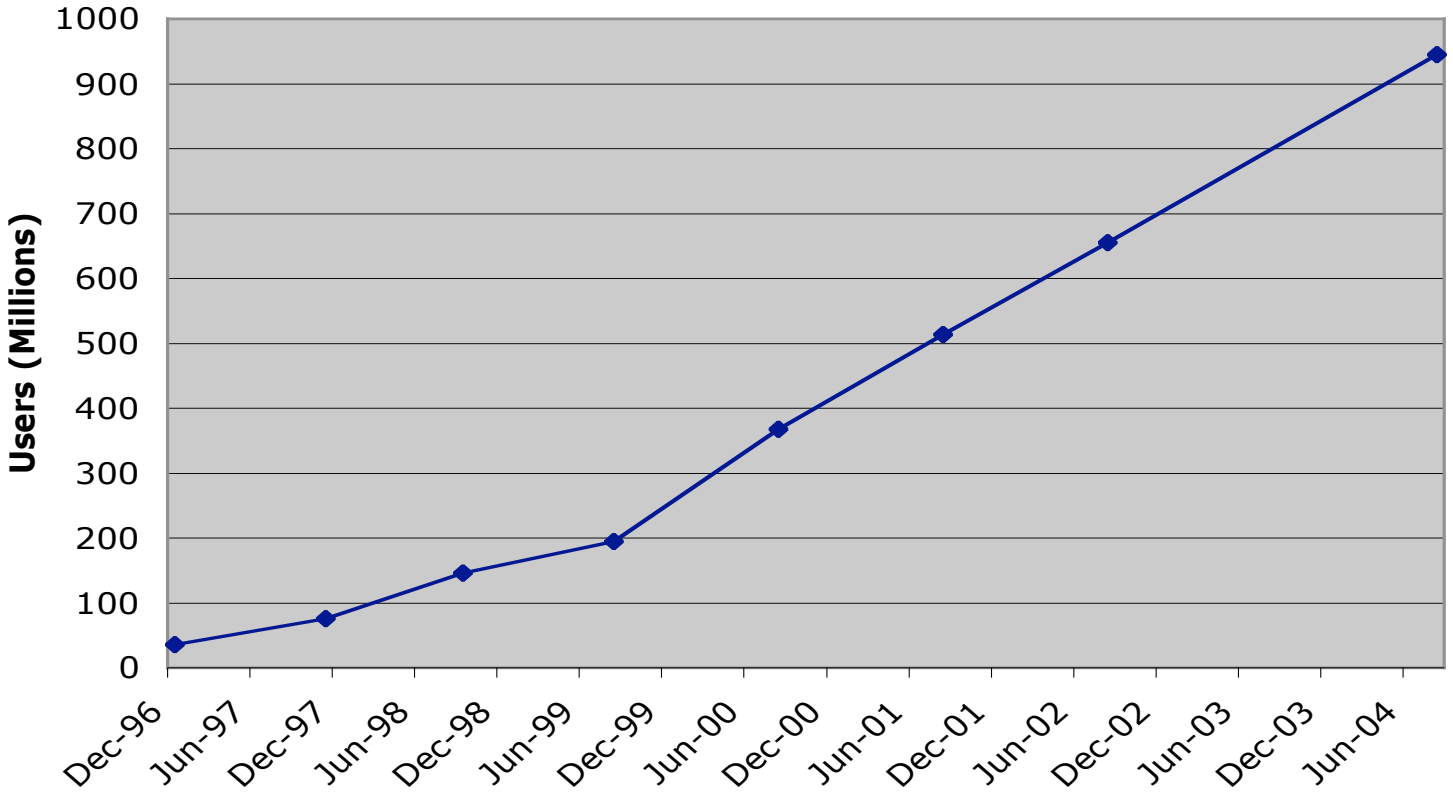
Mark Handley
Professor of Networked Systems
UCL Department of Computer Science

Computers on the Net



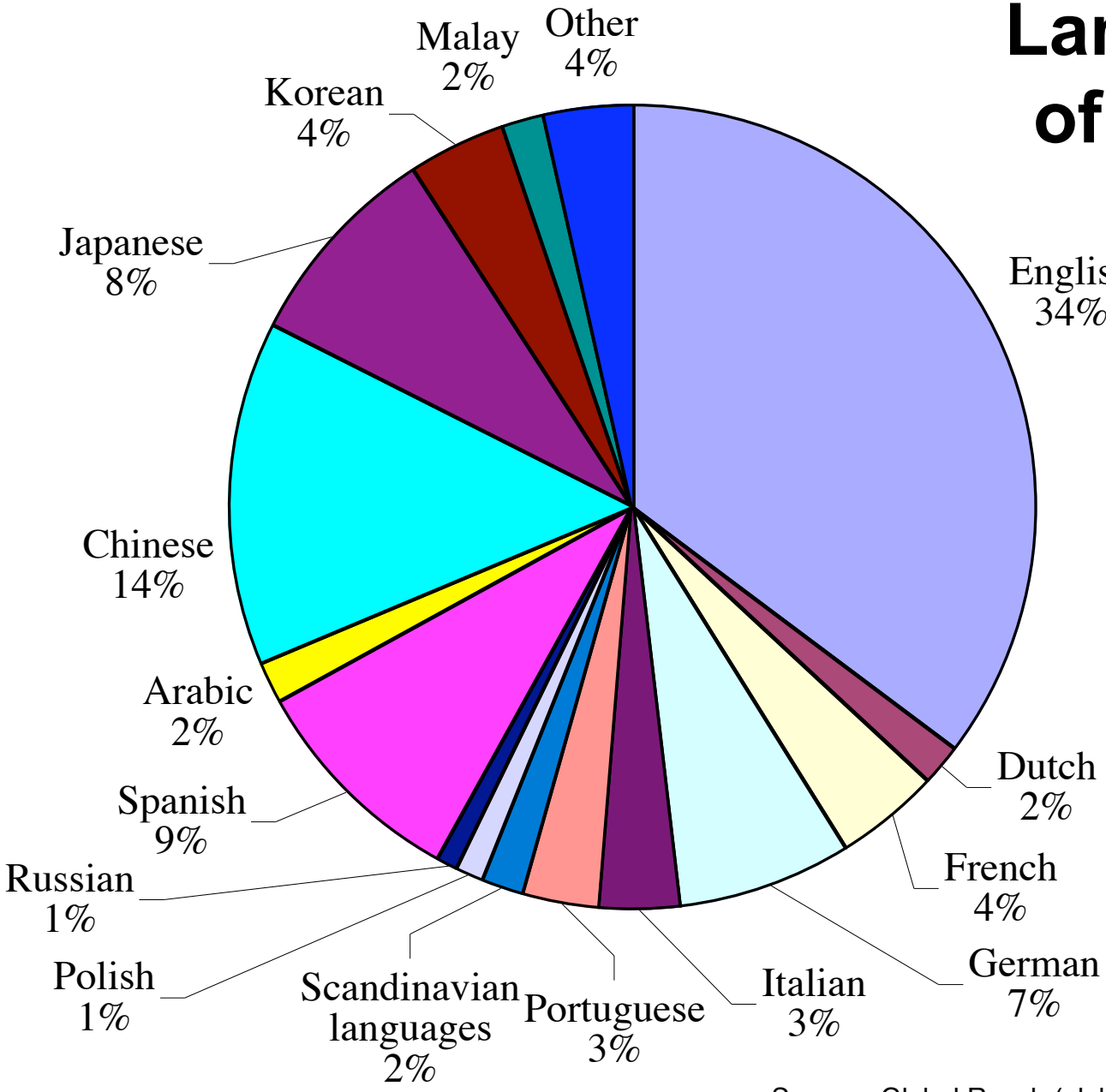
Source:Internet Software Consortium (http://www.isc.org/)

People on the Net



Sources: Reuters, ITC, NUA, ITU

Languages of Internet Users (Sept 2004)



Source: Global Reach (global-reach.biz/globstats)

The net is a success!

- The problem:
  - In almost every way, the Internet only just works!

The net only just works?

It's always been this way:

1975-1981:

TCP/IP split as a reaction to the limitations of NCP.

1982:

DNS as a reaction to the net getting too large for hosts.txt files.

1980s:

EGP, RIP, OSPF as reactions to scaling problems with earlier routing protocols.

1988:

TCP congestion control in response to congestion collapse.

1989:

BGP as a reaction to the need for policy routing in NSFnet.

Changing the net.

- 1st Jan 1983.
  - Flag day.
  - ARPAnet switched from NCP to TCP/IP.
  - About 400 machines need to switch.
- As the net got bigger, it got a lot harder to change.

Sweden Changeover to Right Hand Traffic 1967

Before web...

- Prior to the 1990s the Internet was primarily academic and scientific.
  - Common goals.
  - Low cost of failure.
- Then came the web, and commercialization of the Internet.
  - Exponential growth.
  - Financial costs of failure.
  - ISPs struggling to keep ahead of demand.
  - Huge innovation in applications.

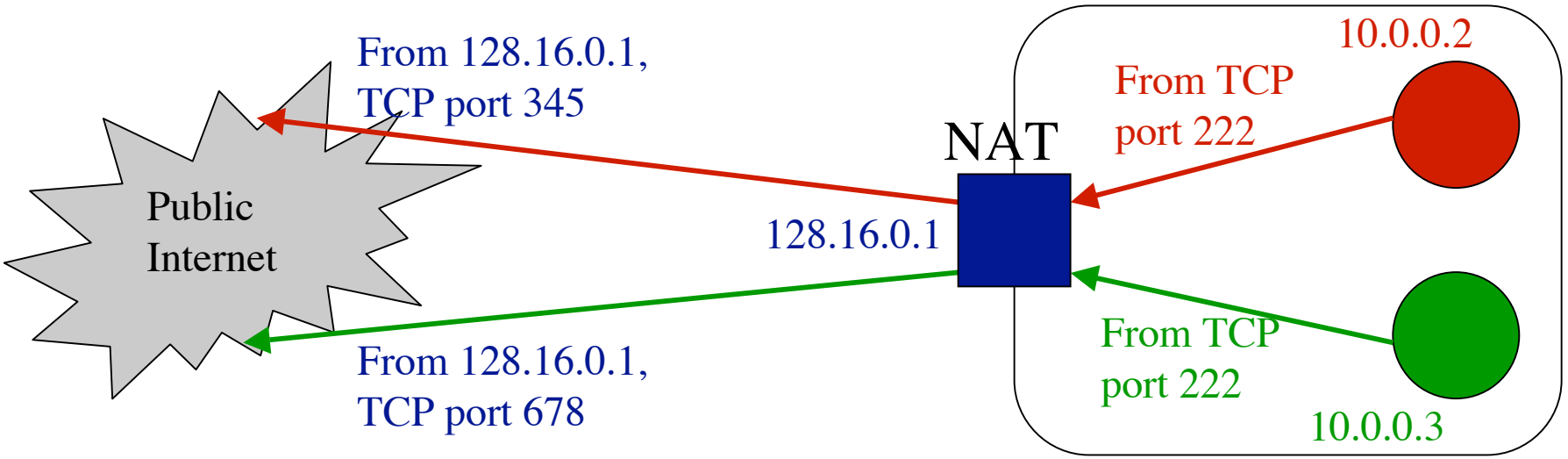
Development Cycle

An Example:
Running out of addresses...

- The current version of the Internet Protocol (IPv4) uses 32 bit addresses.
  - Not allocated very efficiently.
  - ~~MIT has more addresses than China.~~
    MIT + Interop trade show + Halliburton = China
- IPv6 is supposed to replace IPv4.
  - 128 bit addresses.
  - We don't need to be smart in address allocation.
  - How do we persuade people to switch?

Network Address Translators

- ~~Scarcity of addresses~~ tiered pricing has made addresses expensive.
- NATs map one external address to multiple private internal addresses, by rewriting TCP or UDP port numbers in flight.



Network Address Translation

- Introduces asymmetry:
  - Can't receive an incoming connection.
- Makes it very hard to refer to other connections:
  - Eg. SIP signalling causes the phone to ring.
  - On answer, set up the voice channel.

- Application-level gateways get embedded in NATs.
  - Can't change the ends until you change the middle.
  - It should be *easy* to deploy new *applications*!

The sky is falling!!!

- No.
- But we're accumulating problems faster than they're being fixed.
- There has been no significant architectural change to the network core in a decade.

Imminent Architectural Problems

- Spam.
- Security.
- Denial-of-service.
- Application deployment issues.

Medium Term Architectural Problems

- Congestion control.
- Routing.
- Mobility, Multi-homing
- Architectural ossification.

Long Term Problems

- Address space exhaustion.

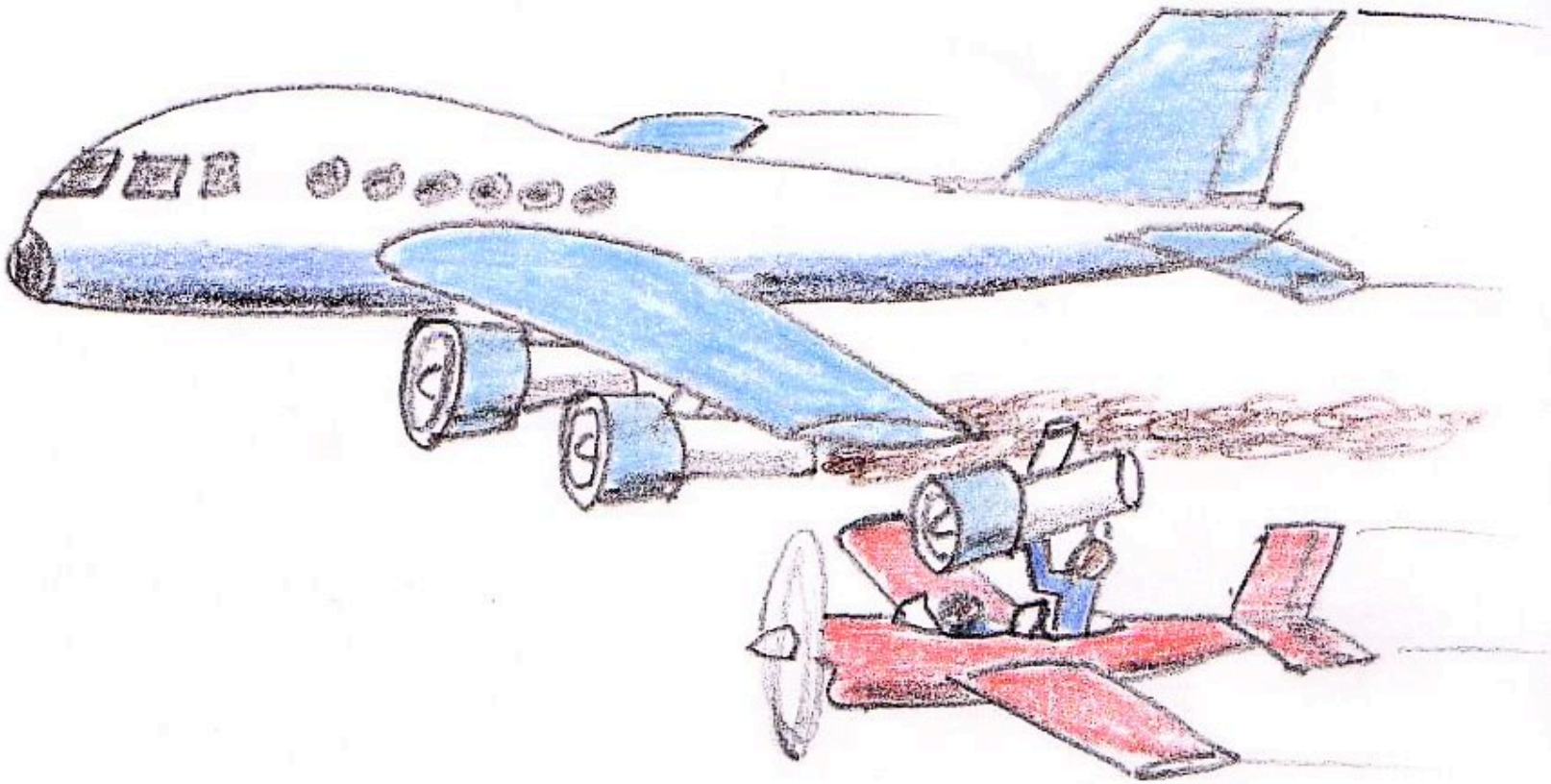
Key Challenge

Is it possible to change the Internet architecture in a planned way, so as to achieve long-term goals?

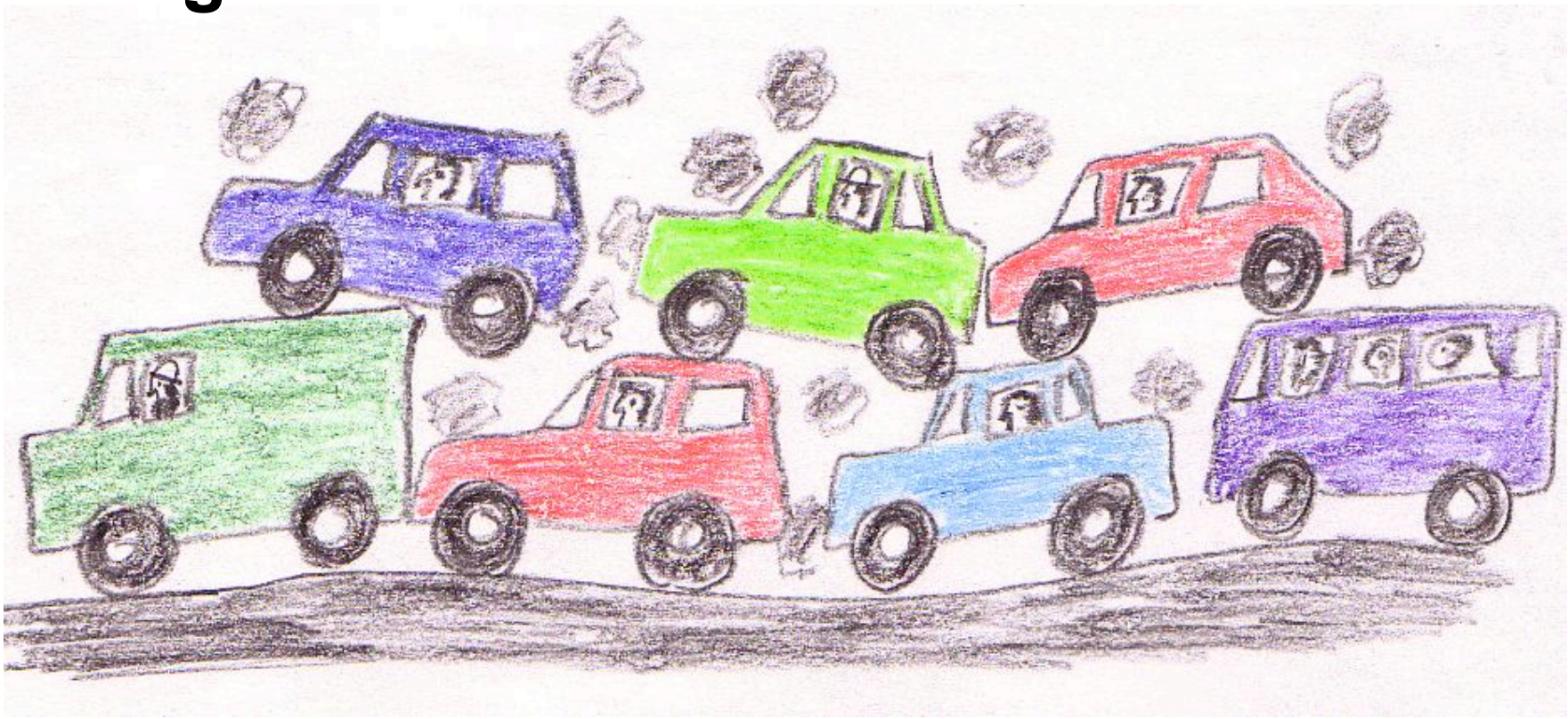
(or is it only possible to patch the pieces repeatedly until it gets too expensive and unreliable, and eventually something better comes along and replaces it?)

Evolving the Internet Architecture:

Changing the Engines in Mid-Flight

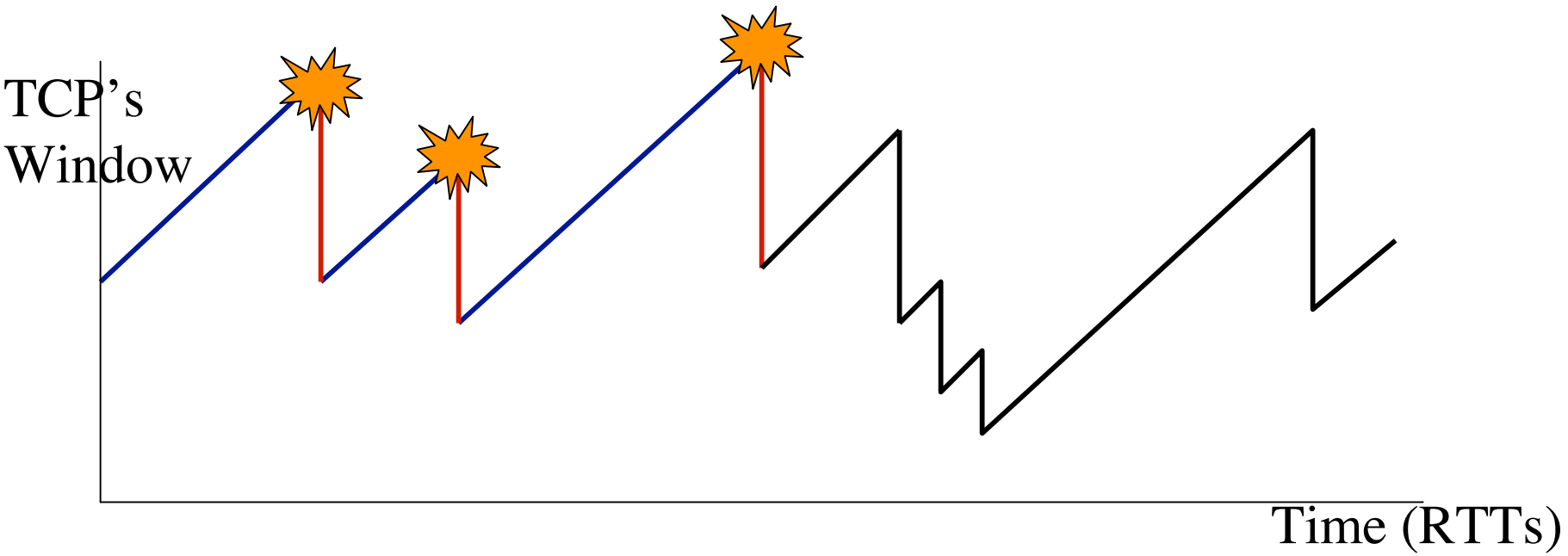


Congestion Control



Congestion Control

- The Internet only functions because TCP's congestion control does an effective job of matching traffic demand to available capacity.



Limitations of TCP Congestion Control

- Failure to distinguish congestion loss from corruption loss.
  - Wireless

- Limited dynamic range.

$$\text{transmit rate} \sim= \frac{\text{packet size}}{\text{RTT}\sqrt{\text{loss rate}}}$$

TCP: Limited Dynamic Range

One loss every half hour, 200ms RTT, 1500bytes/pkt.

⇒ 9000 RTTs increase between losses.

⇒ peak window size = 18000 pkts.

⇒ mean window size = 12000 pkts.

⇒ 18MByte/RTT

⇒ 720Mbit/s.

⇒ Needs a bit-error rate of better than 1 in 10^{12}.

⇒ Takes a *very* long time to converge, or recover from a burst of loss.

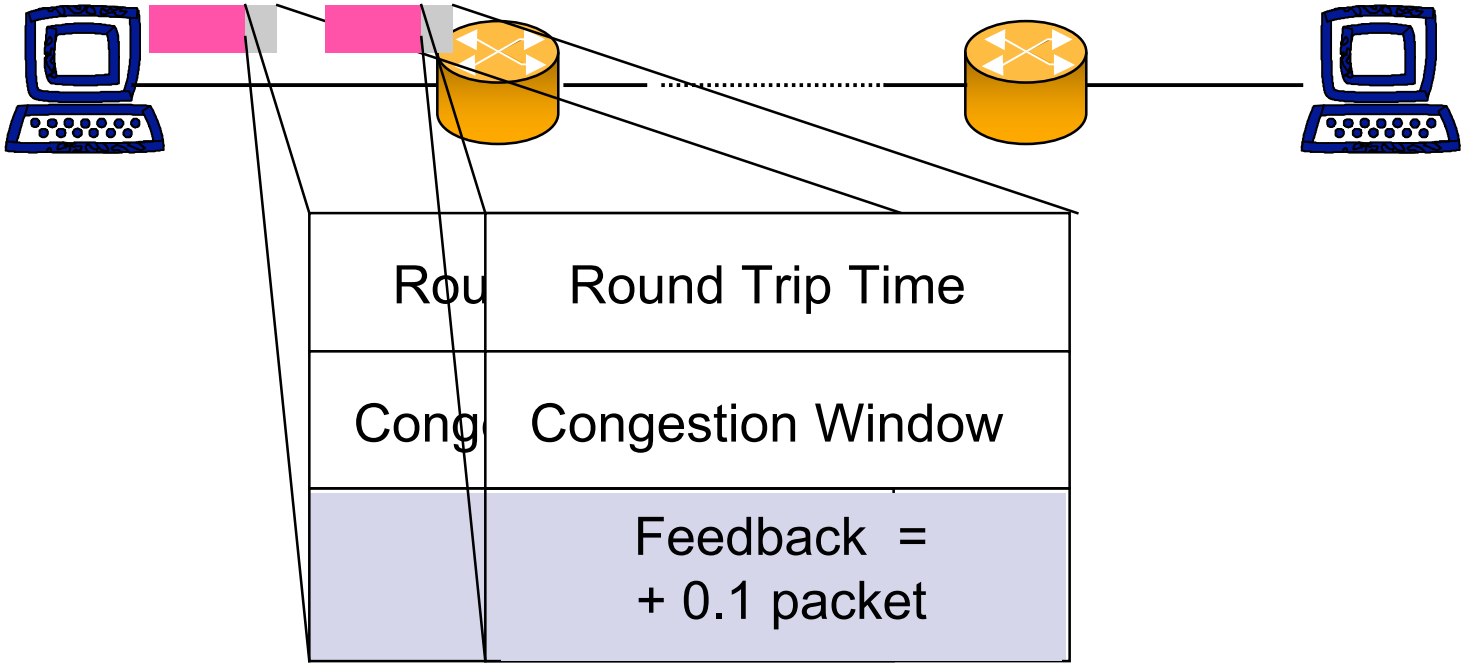
Opportunity

- We *will* need to change the congestion control dynamics of the Internet.
- This presents an opportunity to do it right and solve many additional problems at the same time.
  - Wireless?
  - Smooth throughput for multimedia?
  - Low delay service?
  - DoS resistant?

- But the temptation is always to solve only the immediate problem.
  - Key is having a good solution available at the right time.

XCP: eXplicit Control Protocol

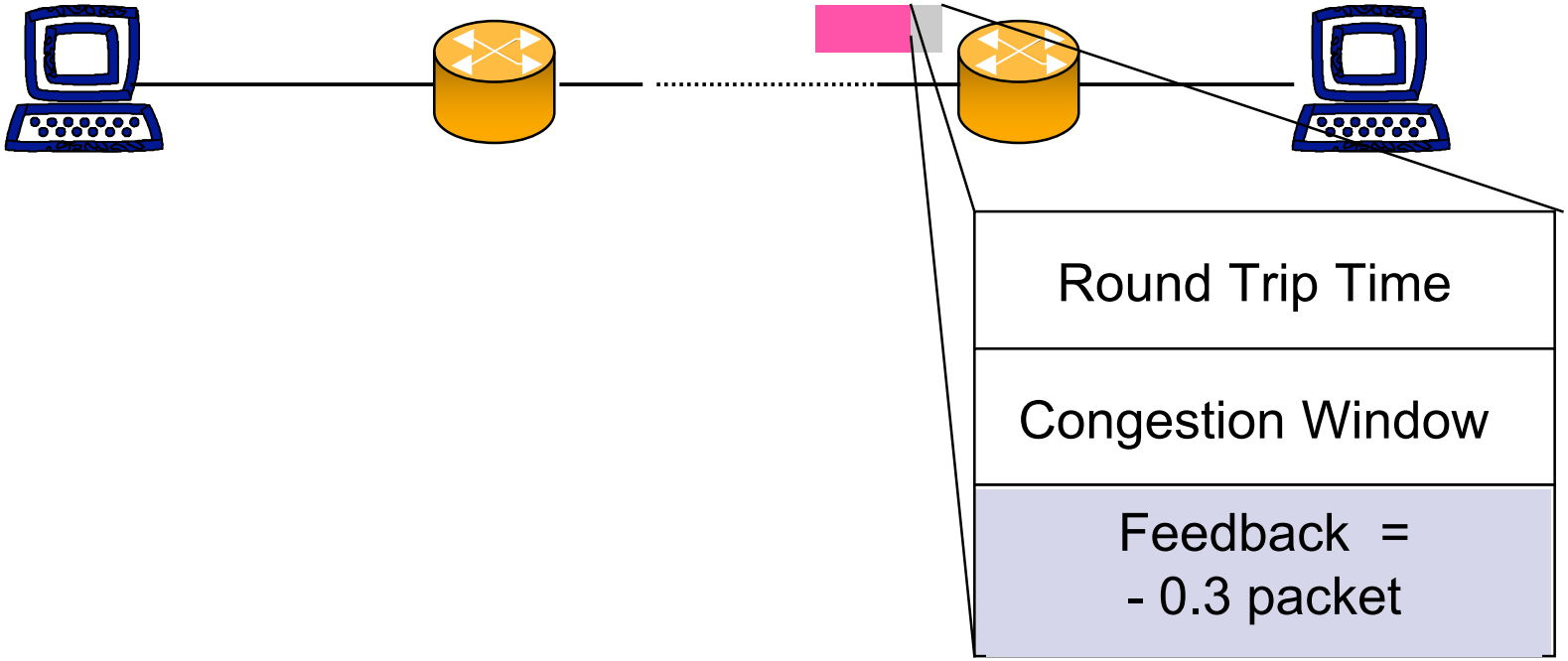
Katabi, Handley, Rohrs, Sigcomm 2002



Congestion Header

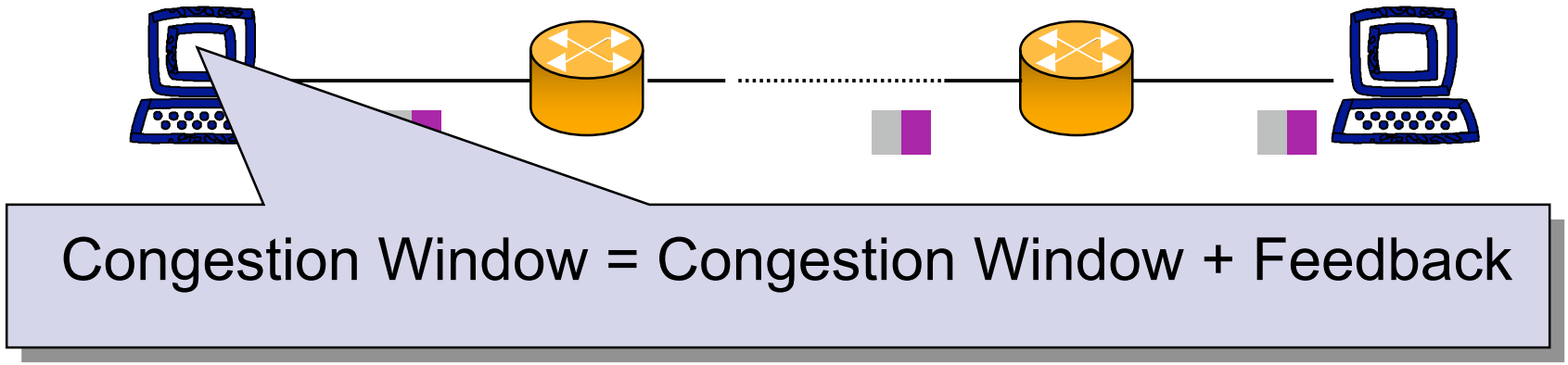
XCP: eXplicit Control Protocol

Katabi, Handley, Rohrs, Sigcomm 2002



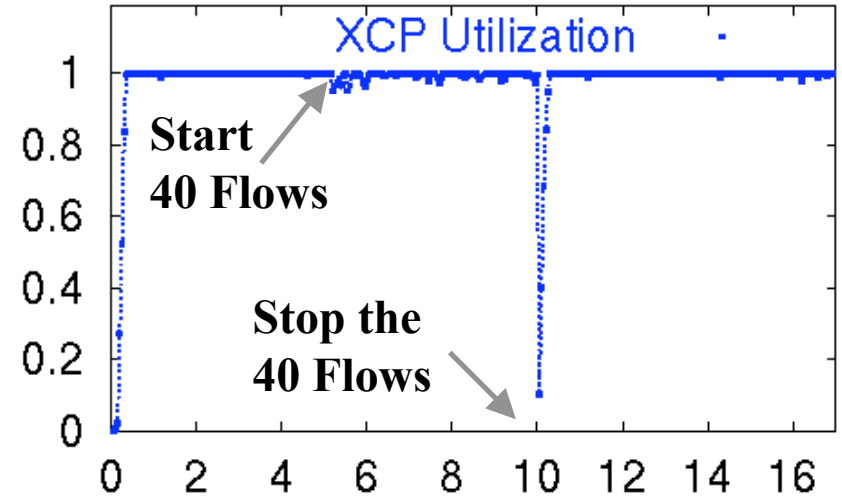
XCP: eXplicit Control Protocol

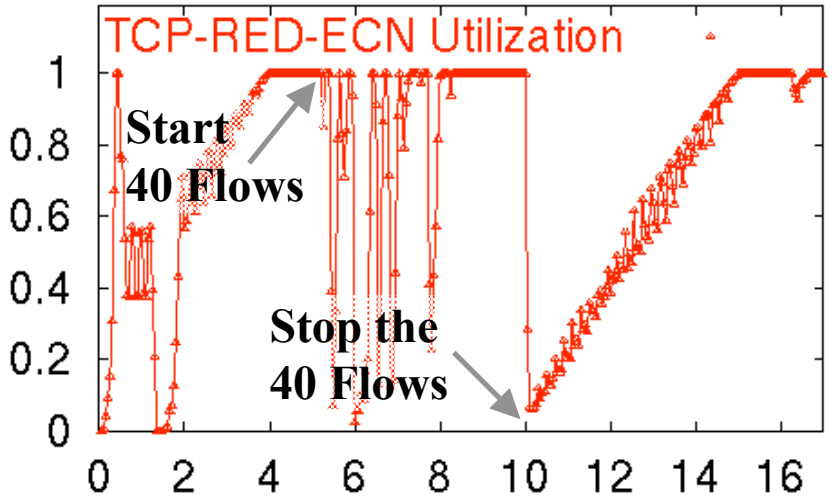
Katabi, Handley, Rohrs, Sigcomm 2002

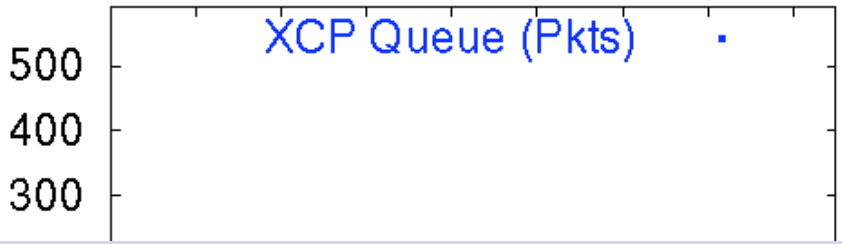


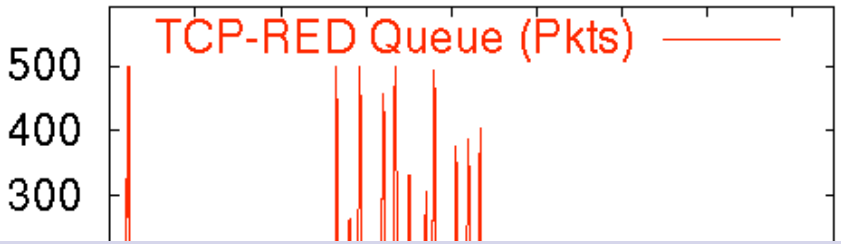
Routers compute feedback without any per-flow state

XCP vs TCP









XCP responds quickly to change, gives smooth throughput, low delay, and low loss.

Time (seconds)

Time (seconds)

So why isn't everyone doing it?

- XCP was intended as a *blue-sky* idea to see what was possible.
  - Needs all the routers on the path to play.
  - Lots of bits in packet headers.
  - A couple of multiplies and a few adds per packet.

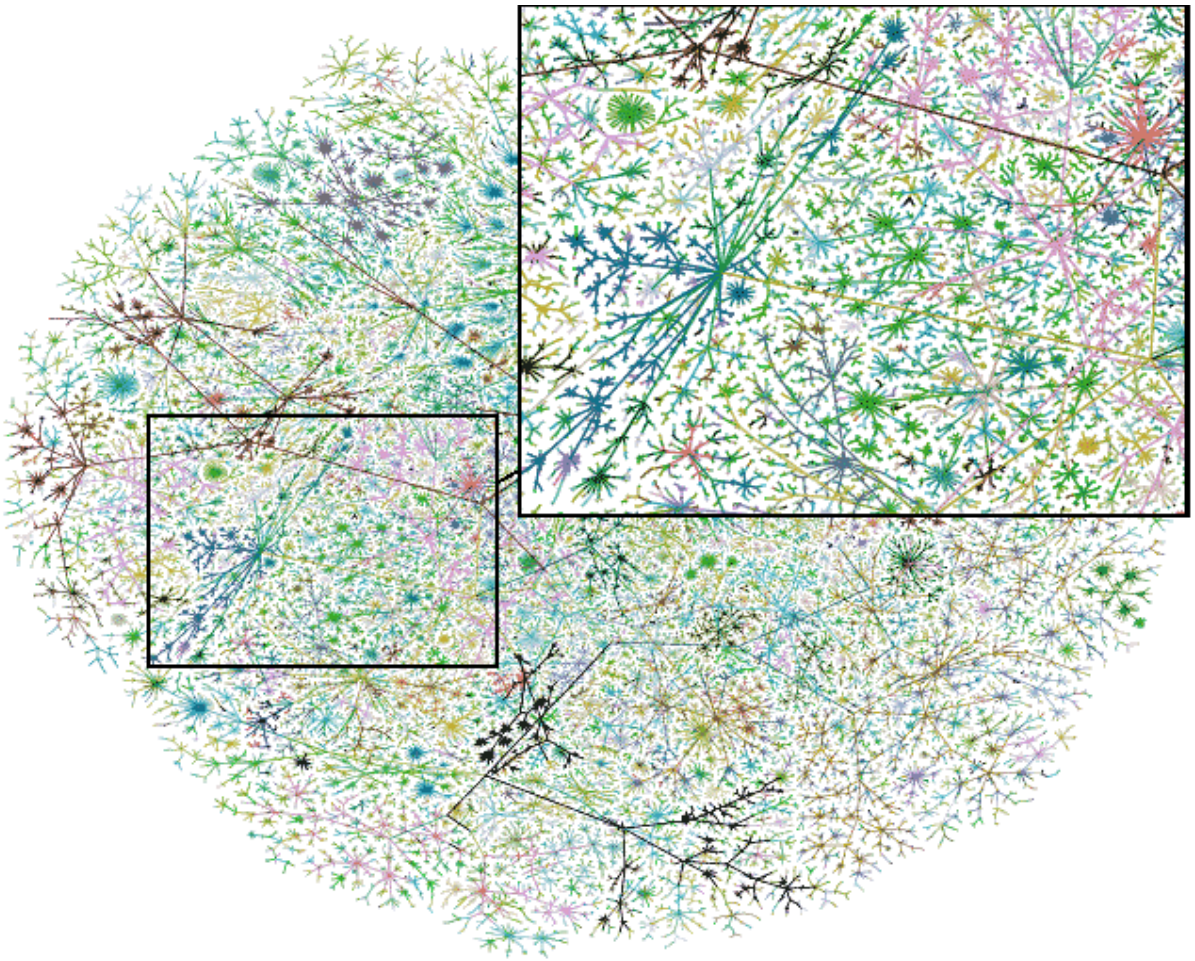
- Now we need phase 2: *Can we make it economically viable*?
  - Reduce costs without destroying benefits.
  - Enable incremental benefit with incremental deployment.

Plenty of Solutions

- High-speed TCP (S. Floyd)
- Scalable TCP (T. Kelly)
- FAST (S. Low)
- H-TCP (D. Leith)
- Bic-TCP (I.Rhee)

- Need a forum for evaluation and consensus that includes researchers and vendors.
  - IETF is not good at this.

Routing
(Internet map, 1999)



Source: Bill Cheswick, Lumeta

Routing

BGP4 is the only inter-domain routing protocol currently in use world-wide.

- Lack of security.
- Ease of misconfiguration.
- Policy through local filtering.
- Poorly understood interaction between local policies.
- Poor convergence.
- Lack of appropriate information hiding.
- Non-determinism.
- Poor overload behaviour.

Replacing BGP?

- BGP works!
- BGP is the most critical piece of Internet infrastructure.

- *No-one* really knows what policies are in use.
  - And of those, which subset are *intended* to be in use.
- No economic incentive to be first to abandon BGP.

Criteria for Successful Replacement

- Interoperate with BGP without any serious degradation in capability during transition.
- Provide incremental improvement when customers and their providers both switch
  - outside-in deployment.
- Concepts must be familiar to ISPs.

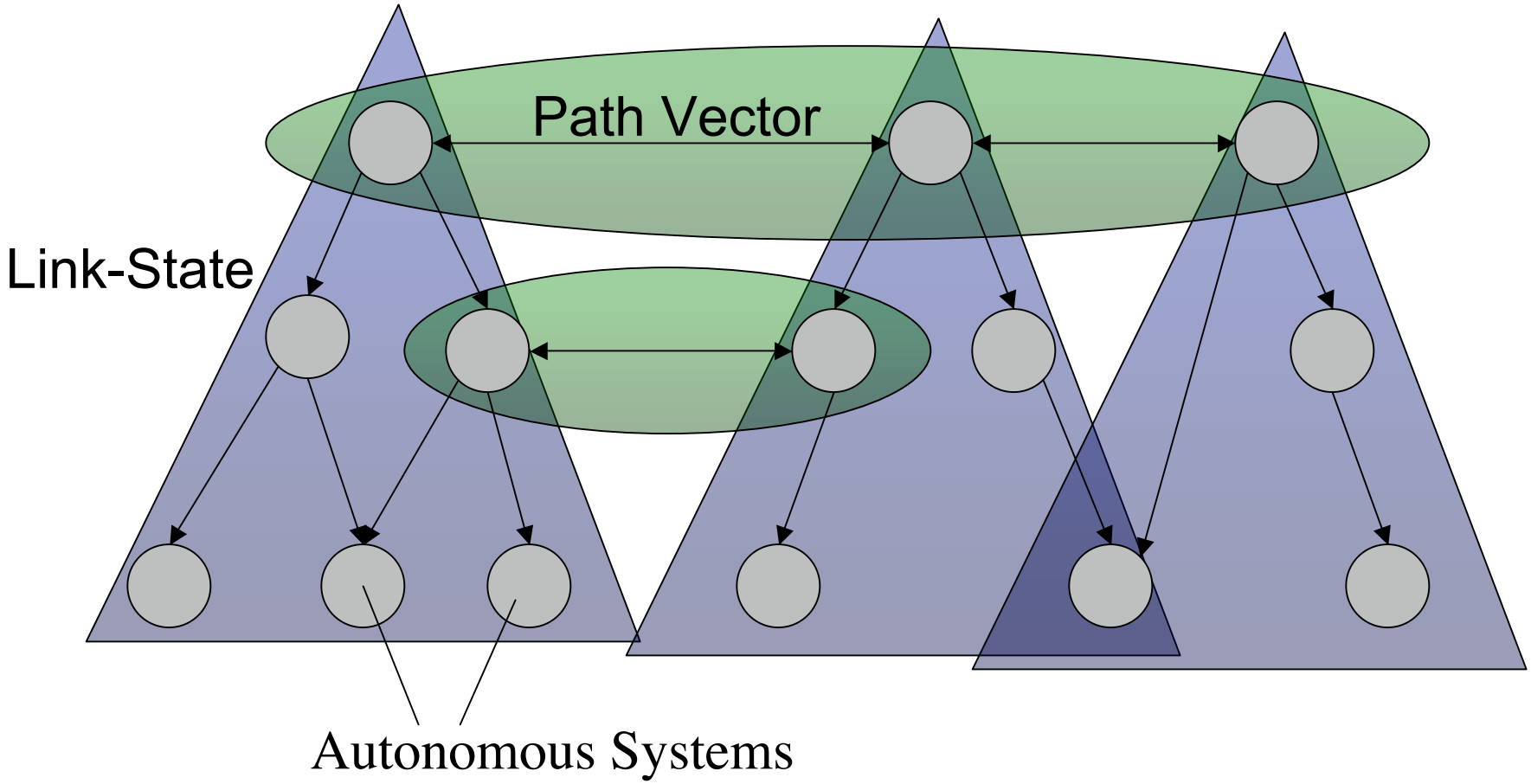
Opportunity for Replacement?

- BGP must be seen to be failing.
  - Security problems being actively exploited?
  - Convergence problems too slow for high-value traffic (VoIP, IP-TV)?
  - Growth of multi-homing causes routing table growth/churn that is unsupportable?

BGP Replacements

- Hybrid Link-State/Path-Vector (Hotnets 2004)
- Specification-Based Routing (Griffin)

HLP



Specification-Based Routing

- Current routing protocols are “implementation-based” - the semantics are embedded in the implementation by the router vendor.
- Goal of SBR is to define a routing grammar, and ways to express and transport both routing information and routing semantics.
- Four layers, each constrains the subsequent ones:
  1. Core RPML.
  2. Core “protocol” semantics (expressed in RPML)
  3. Local policy semantics (expressed in RPML)
  4. On the wire routing information and route-specific semantics.

Denial of Service

The Register » Security » Network Security »

US credit card firm fights DDoS attack

By John Leyden

Published Thursday 23rd September 2004 11:13 GMT

US credit card processing firm Authorize.Net is fighting a sustained distributed denial of service (DDoS) attack that has left it struggling to stay online.

In a statement to users posted yesterday, Authorize.Net said it "continues to experience intermittent distributed denial of service (DDoS) attacks. Our system engineers have successfully minimised the impact of each attack and have quickly restored services to affected merchants. Industry experts are onsite and working with Authorize.Net to expedite a resolution. Please be aware that the stability and reliability of the Authorize.Net platform remains our top priority; and we are doing everything we can to restore and maintain secure transaction processing despite these unforeseen attacks."

Denial of Service

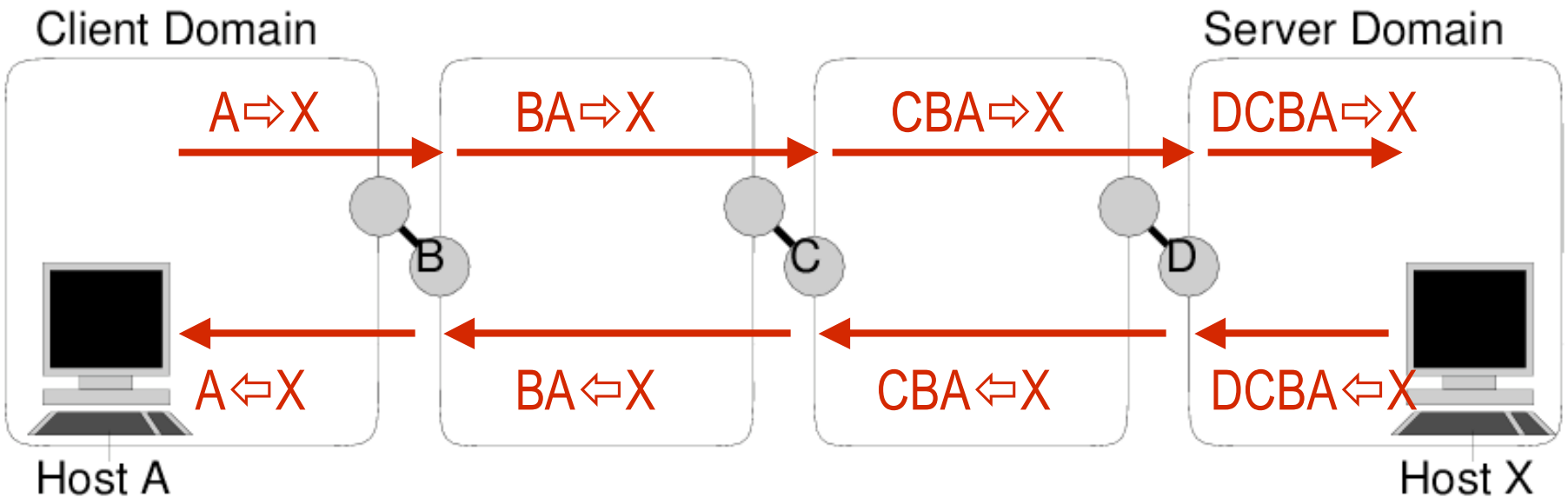
- The Internet does a great job of transmitting packets to a destination.
  - Even if the destination doesn't want those packets.
  - Overload servers or network links to prevent the victim doing useful work.

- Distributed Denial of Service becoming commonplace.
  - Automated scanning results in armies of compromised zombie hosts being available for coordinated attacks.

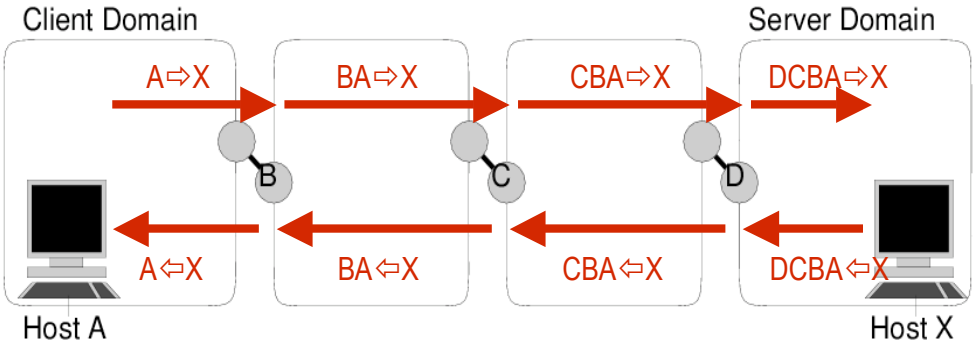
Denial of Service

- Traditional security mechanisms are useless for defending against DoS.
  - Attacker can force you to do expensive crypto operations.
- Many DoS point solutions have side-effects that can be exploited by an attacker:
  - Sendmail SPAM blocking.
  - BGP Flap Damping.
- The death of ping.

Path-based Addressing (FDNA 2004)



Path-based Addressing



Benefits

- Prevents address spoofing.
  - Thus reflection attacks on remote hosts not possible.
- Prevents DDoS of clients: their addresses are not guessable.
- Prevents fast worms. Must go client→server→client.
- Paths are symmetric at the inter-domain level
  - Unidirectional traffic is then clearly visible as malicious, even in the core of the Internet.
- Remote subversion of client routing not possible.
- Provides a safe architecture for deploying automatic pushback mechanisms to shut down malicious hosts.

Architectural Ossification

- The net is already hard to change in the core.
- IP Options virtually useless for extension.
  - Slow-path processed in fast hardware routers.
- NATs make it hard to deploy many new applications.
- Firewalls make it make to deploy anything new.
  - But the alternative seems to be worse.
- ISPs looking for ways to make money on “services”.
  - Many would love to lock you into their own private walled garden, where they can get you to use their services and protocols, for which they can charge.

Summary

- In almost every way, the net only just works.
- This is a *critical* time.
  - The net is moving out of it's infancy.
  - The problems are significant.
  - Increased expectations of performance and robustness.
  - Problems create opportunity for architectural change that would not otherwise be economically viable.
- *We* get to influence how this plays out.
  - Do we patch, or do we fix the underlying issues?