15 years of Policy Routing

Susan Hares
NextHop Technologies

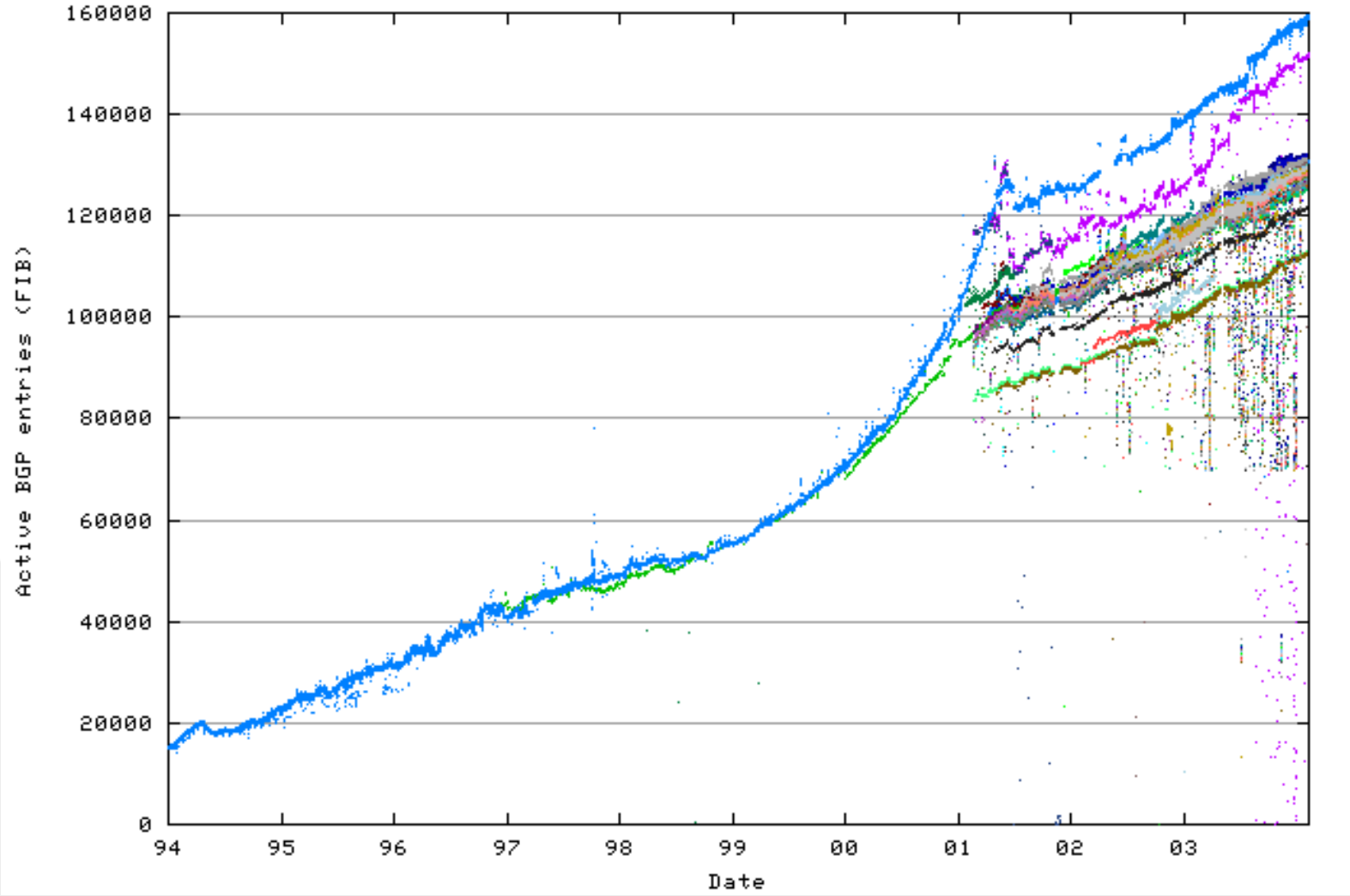
NANOG 30

February 9th, 2004

What's New? (1989 vs. 2004)

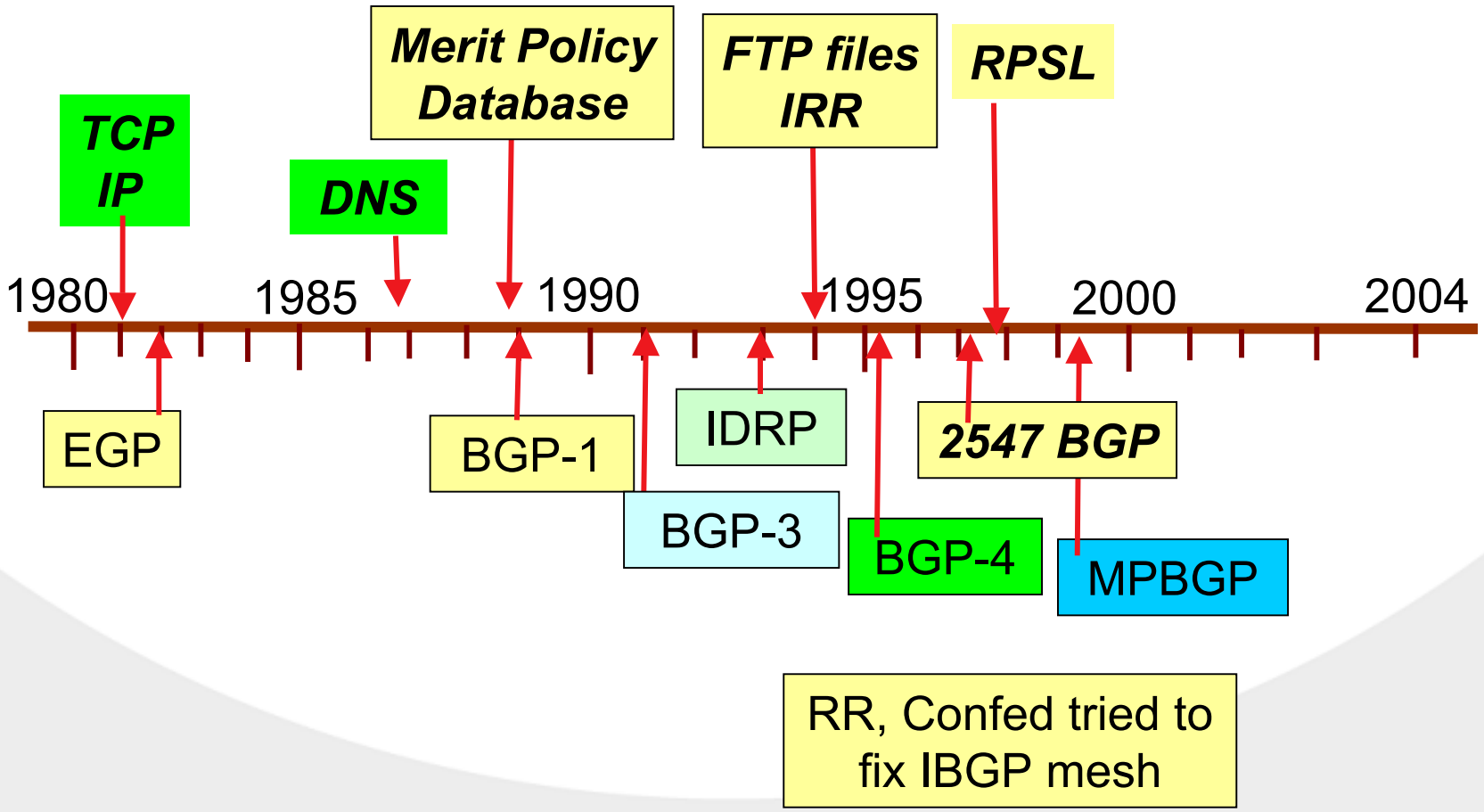
- Commercial Internet is Critical Infrastructure
  - Fierce Competition for revenue, little sharing
  - Security attacks occur regularly
  - Carrier's need
- Multihoming, NATs, VPNs
  - Shortage of IP address space (official rationale)
  - Enterprise-friendly demarcation (unofficial driver)
- Policy is complex
  - Multiple independent policies frustrate convergence
  - VPNs create better and more complex router configurations
  - It is critical that SLAs turn into the appropriate router configuraitons

BGP Routing seems simple

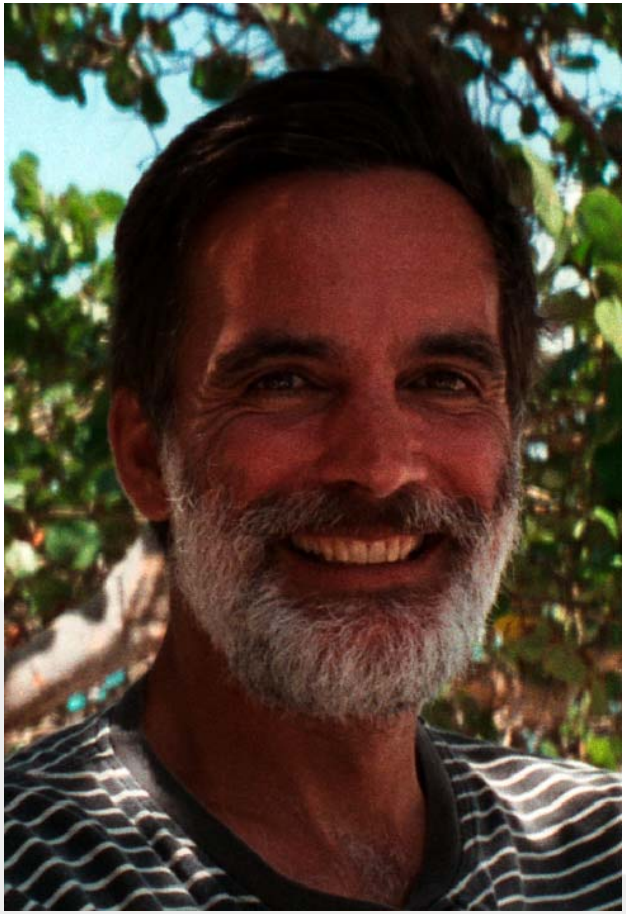


Route Views 2/8/04

Yet.. Policy Routing is still changing



Routing is not simple



Lyman Chapin

"On the surface, routing is simple - just a matter of figuring out how to get from here to there. Twenty years (at least) of research and experimentation with different routing protocols and strategies, and the evolution of our routing toolkit from HELLO to BGP, suggest that below the surface routing is anything but ***simple.***"

Fortunately, we've been able to make progress in understanding, developing, and deploying new routing techniques continuously over that long period of time without suffering too much delay or other damage from the "protocol wars" of the "TCP/IP vs. OSI" era.

Dave Katz's version of History

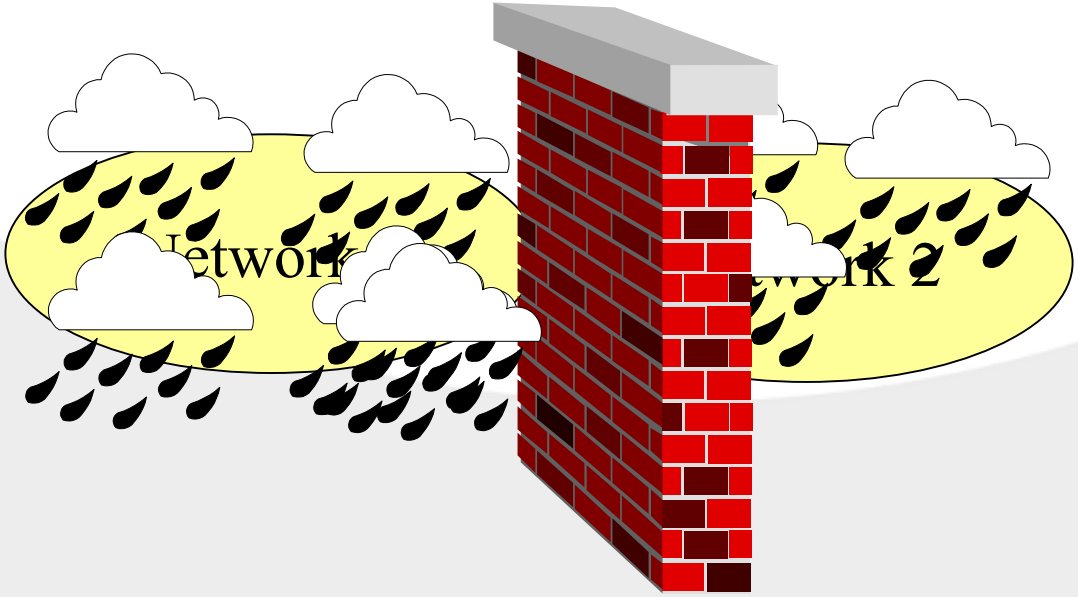
- 1990 - Correctness, Stability, Scalability, Speed: choose 1
- 1994 – Correctness, Stability, Scalability, Speed: choose 2
- 1995 – Correctness, Stability, Scalability, Speed: choose 2.5
- 2002 – Correctness, Stability, Scalability, Speed: choose 3.5

- 2003: Correctness, Stability, Scalability, Speed, HA: chose 4.5

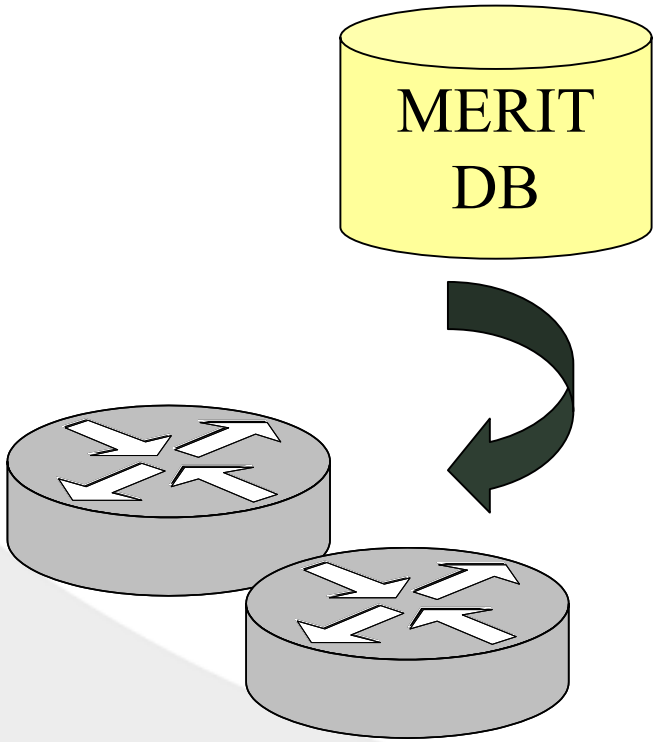
- Software implementations & new algorithm improvements are just starting to pay off (Sue Hares' addition to Dave Katz)
  - Careful engineering should be able to provide speed, scalability, correctness, high availability and stability
  - The only effect of a heavily loaded system should be a gradual slowing in convergence (not to crash and burn
  - Management and Policy Based routing/switching can scale

Why Policy Routing

Technology	Problem we tried to solve	Technologies input	Lessons
Policy Routing	"No Route Storms", limit by policy	1) BGP, EGP 2) IRR, RPSL	1) Policy Routing can limit storms 2) BGP is TLV carrier 3) Convergence matters 4) IBGP full mesh hurts

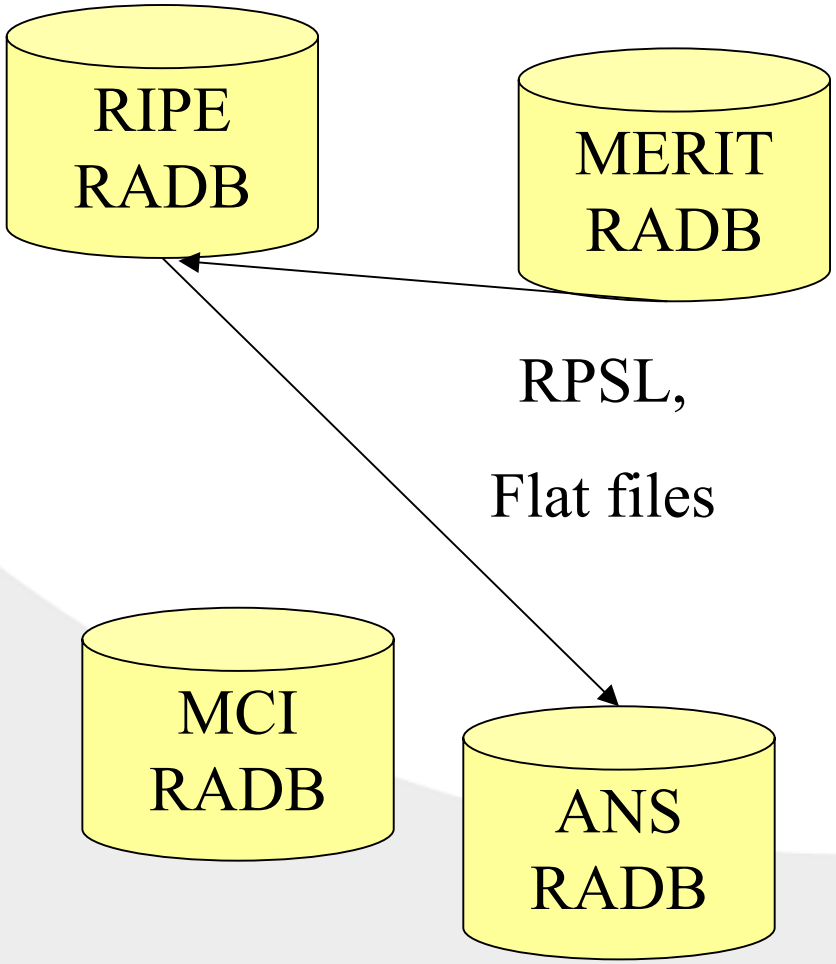


What was Policy in NSFNET



- Find Public BGP policy
- Outgrowth of my desire my poor typing ability
- Created configuration files
  - Router physical
  - BGP policy

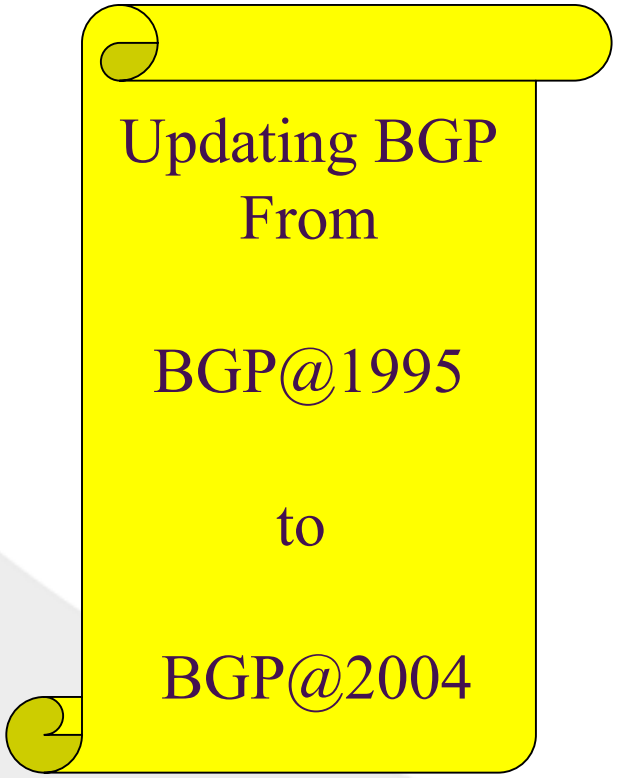
Why an IRR



- Find Public BGP policy
- Outgrowth of Configuration files from NSFNet to Commerical Internet

- Transfer protocol

IETF BGP in 2004



IDR Working Group

- Base Specification is being updated to match current implementations
  - FSM additions
  - Tie-Breaking rules
- Associated Drafts are being upgraded
  - MIB
  - Standardization Reports:
    - Analysis of Protocol,
    - Experience with Protocol,
    - Report on Implementations

We are just about to unleash the new BGP information.

IETF BGP – 2547bis

RFC 2547bis: Is it Carrier's Salvation or Routing Protocol Abuse?

- 2547 Related work
  - Layer L3 VPN will focus on additions to make RFC 2547bis networks work
- Routing Area Discussions
  - Routing Area Meeting at July IETF discussed whether Multi-protocol additions to BGP (and ISIS or OSPF) are protocol abuse
  - Area Director states types of info that can be distributed is:
    - "Information to calculate routing tables
    - Route tagging, Administrative, policy-related information,
    - Routing Security information
    - Information closely related to routing, especially when synchronization with routing information is needed"

Summary of BGP problems

1. BGP convergence problems
2. BGP-4 has no ECMP paths or Traffic Engineering
   - Cisco or Juniper utilize a proprietary QOS
3. BGP policy
   - Inflexible boundaries to BGP (iBGP or EB)
   - Lack of policy verification prior to load
   - No ability to synchronize BGP policy
4. BGP security
   - does not scale to millions of routes
   - BGP security does not protect against replay attacks
5. Concerns about VPNs and NAT overloading BGP
   - Two opinions
     - 2547 == Protocol abuse
     - “Shakespeare” in BGP if customer pays

BGP Convergence Problems

- IBGP
  - fail-over takes seconds instead of ms
  - Route Reflectors or AS Confederations do not converge do to MED problem
  - Route Reflectors of AS Confederations cannot utilize multi-level hierarchy
- Use of TCP requires full mesh topology
- Parallel BGP computations limited by MED comparison
  - must do comparisons on all routes restricting good parallel CPUs must
  - Extra BGP path information to solve convergence overloads the BGP data stream

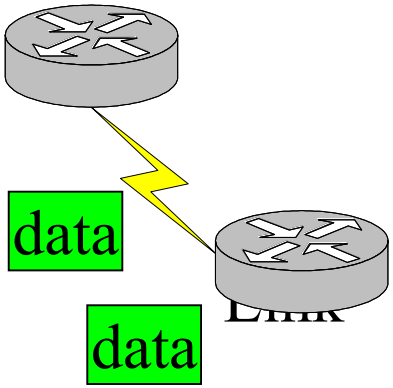
BGP Features coming of Age

- Current IDR Drafts deployed Final Blessing
  - MP-BGP for IPv6, MP-BGP for Multicast, Graceful Restart
  - Extended Communities
  - ORF (Prefix, Communities)
- Drafts in early deployment
  - Cease Codes, ORF ASPath
  - Router-ID extension (Redback)
- L2 and L3 VPN knobs
- New work

New BGP ideas

- AFI/SAFI Isolation
  - AFI/SAFI drafts (Cisco), SAFI Attribute (Cisco), Tunnel Attribute (Cisco), Bundling Multiple TCP Sessions (Cisco)
  - NextHop Revision (Cisco)
  - Inform (Cisco), Soft-Notify (Cisco)
- Better CEASE & Maximum Prefix limits
  - Additional Prefix sent to fix MED oscillation (Cisco))
  - Extended Cease (Redback)
  - Maximum Prefix Draft (Nortel, AT&T, NextHop)
- MED Fix and ECMP
  - Additional Prefix sent to fix MED oscillation (Cisco)
  - ECMP proposal (Samsung)
- Security configs
  - Bogon Requests (Cisco, Security Team)

IETF: Searching for Scalable Security



- IETF
  - RP Security working Group looking at requirements (RP-ESC)
  - Link security based on virtual links security
    - TCP MD5 for BGP-4/BGP-5 links
    - GRE/IP-SEC for BGP-5 links
- Routing Data security
  - S-BGP: Certify BGP information
    - (up to 700% overhead) due to multiple copies of information
  - S-O-BGP: Secure the Origin AS-Route mapping
    - Just secures the origin and some parts
  - Choice for operators: 700% overhead or just the origin
- Security certification hierarchy
  - Either PKI based or Registry based (S-oBG or INV)

Randy Bush:

"We'll wait for the NG hardware"

Scaling – Implementations matter

"If it scales, all else follows" – Mike O'Dell

"Better a good implementations of a bad protocol, than a bad implementation of a good protocol" – Tony Li

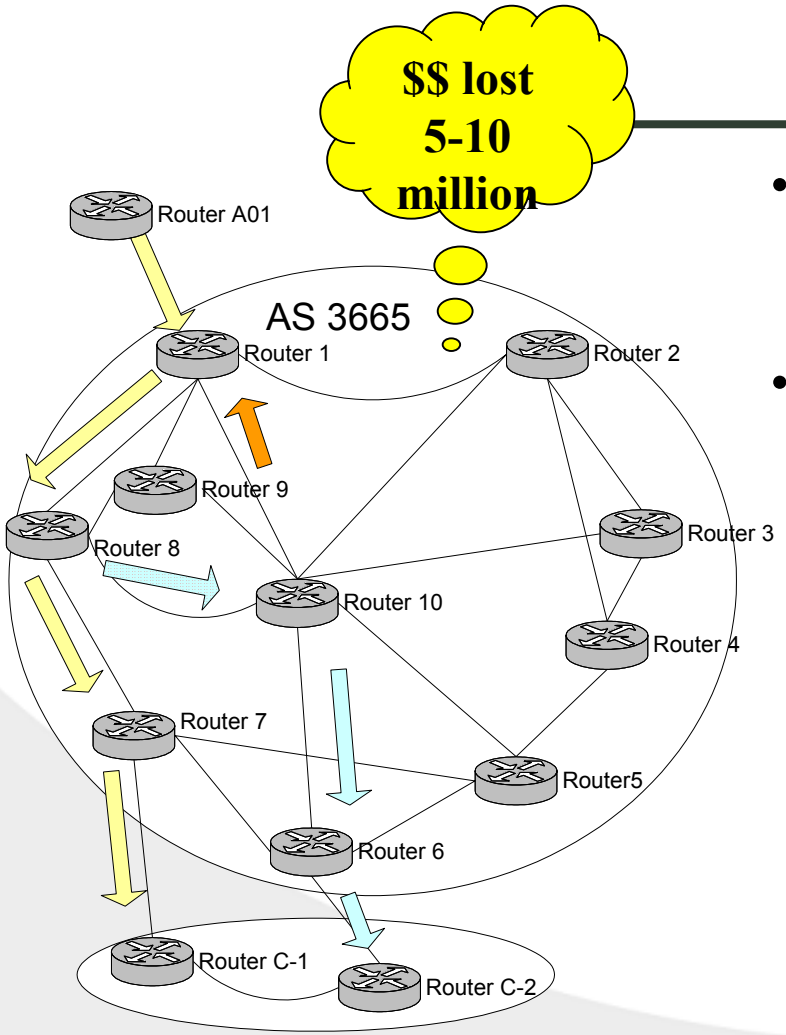
Why Improve Security?

- Business: Security issues are key to enterprise and carriers
  - "Making sure network is hacker proof" is key concern of IT Managers"
    - IT managers rated it a 8.3 on scale of 10 **(NWFusion, 7/21/03)**
  - "Security Drives End-User Spending Up 47%
    - on Data Centers and Hosting Services—from $10.6B to $15.6B between 2003 and 2007" **(Infonetics 7/14/03)**

- Technical: Configuration and Protocol Security
  - ❑ Without tighter control on both configuration and security, it is difficult find the attacks anomalies from the errors.

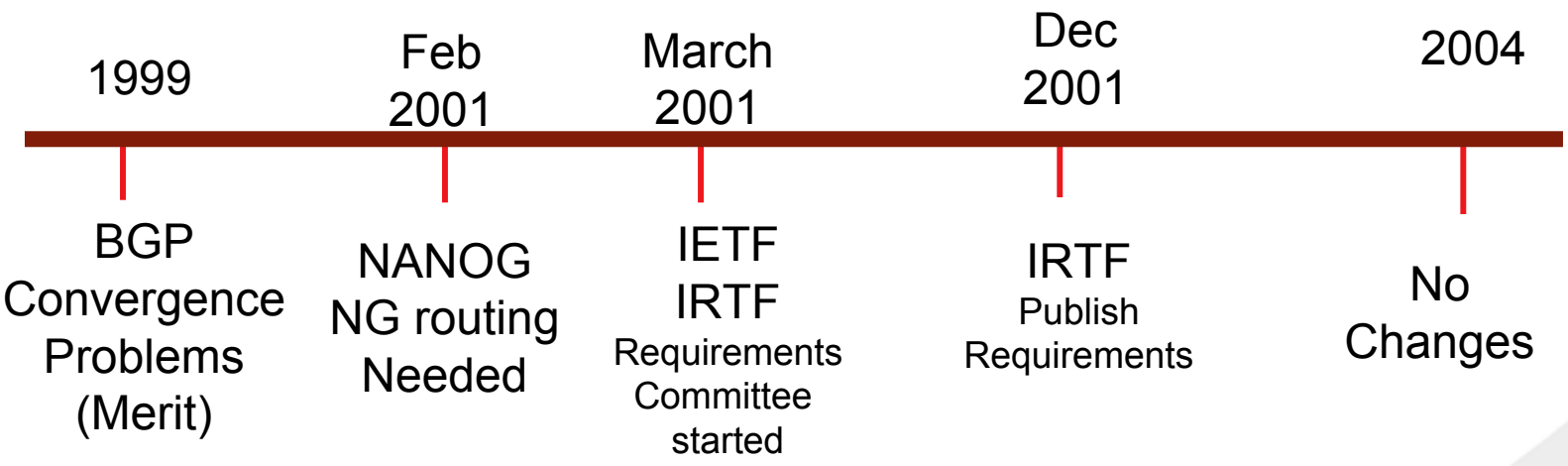
- US Government: Has deems BGP infrastructure critical
  - [National Strategy to Secure Cyberspace, 5 year plan) 2/17/03 infrastructure]

Why Fix Policy?



- Complex Policy
  - Complex Policy s not just a BGP problem
  - BGP is Policy at PhD. Level
- 25% to 50% of network outages related to configuration [**Infonetics 3/2003]**
  - **Outages due to Configurations:** Up to 8-10 million dollars for large enterprises
  - **Total outages:** 0.1% to 1% of revenue lost due to outages (up to 74.6 million up to $96,632/hour of downtime (productivity losses for poor service not included)
  - **Cause of Outages:** 1st Servers, 2nd network devices
    - 50% of network devices outages caused by configurations
- 33% (general) to 50% (Carrier) of outages due to configuration errors [**Yankee Group 2002**]

NG Routing in IETF/IRTF

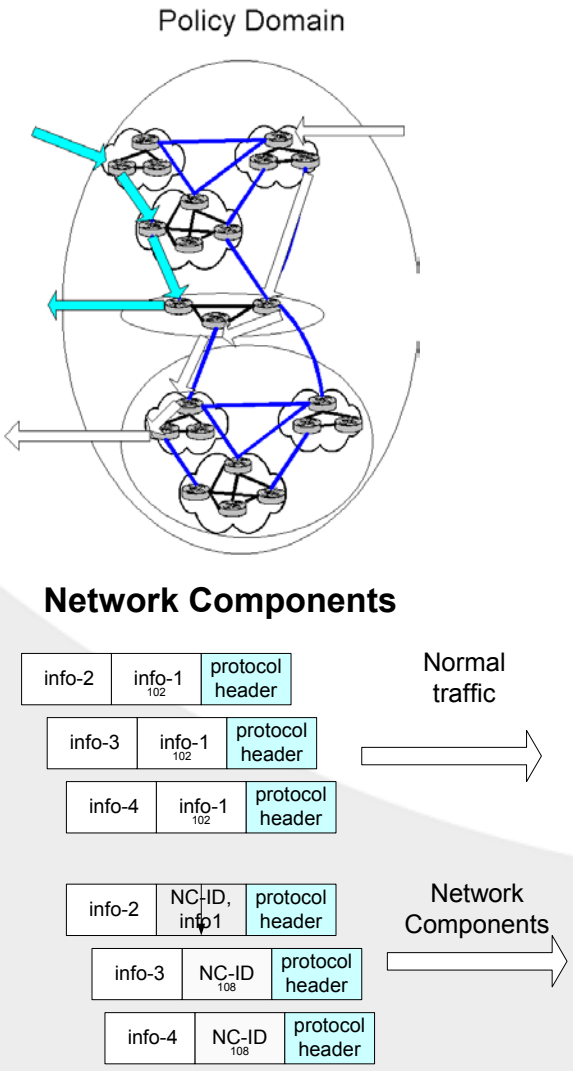


Why BGP Next Generation (BGPng)

- Scalable routing infrastructure
- Reduce management costs of BGP
- Scalable, manageable security
- Remove bottle neck from computation process

All with incremental deployment (No flag day)

BGP NG – New Algorithms



- Policy domains – areas of consistent policy
  - Policy verification mechanisms to ensure synchronized and consistent policy
- Link state path vector algorithms
  - Network Component mechanism to reduce traffic flooded by link state path vector mechanisms and secure information
- Network Components
  - Secure BGP at Data Level while providing network layer compression