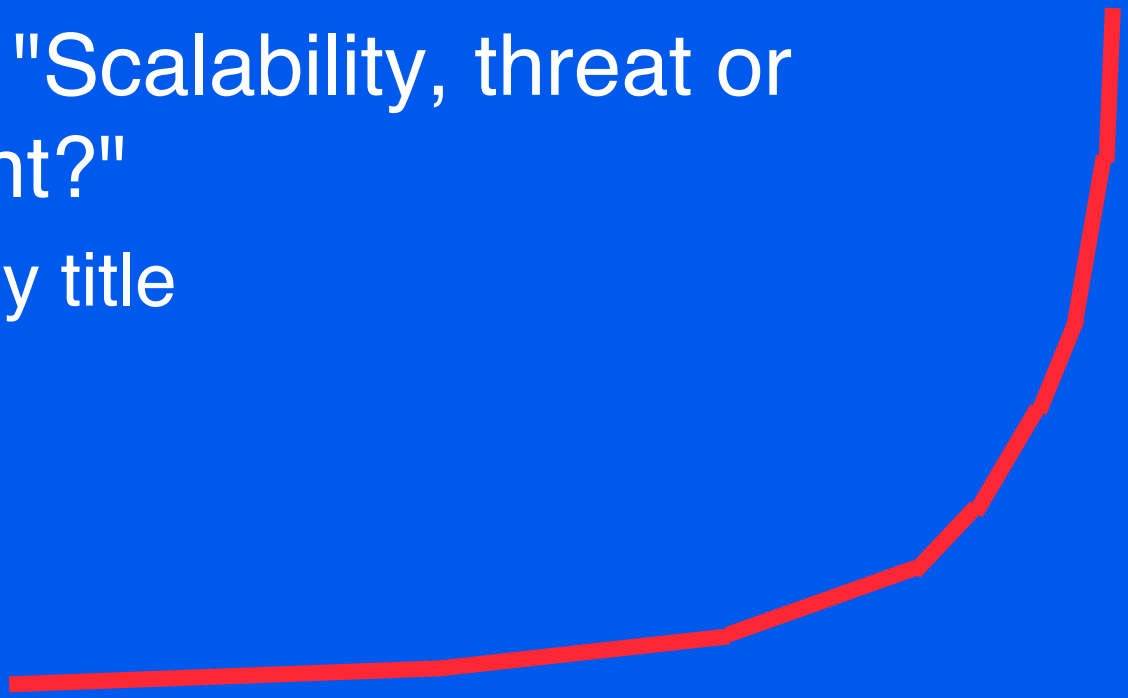
Real QoS versus a Few Traffic Classes

Scott Bradner

Harvard University
IETF Transport Area Director
sob@harvard.edu

A (mis)Leading Title

- title makes assumption that there is only one way to do QoS
  - but note who devised title
- how about "Scalability, threat or requirement?"
  - but that's my title



Deep Desires

- predictability

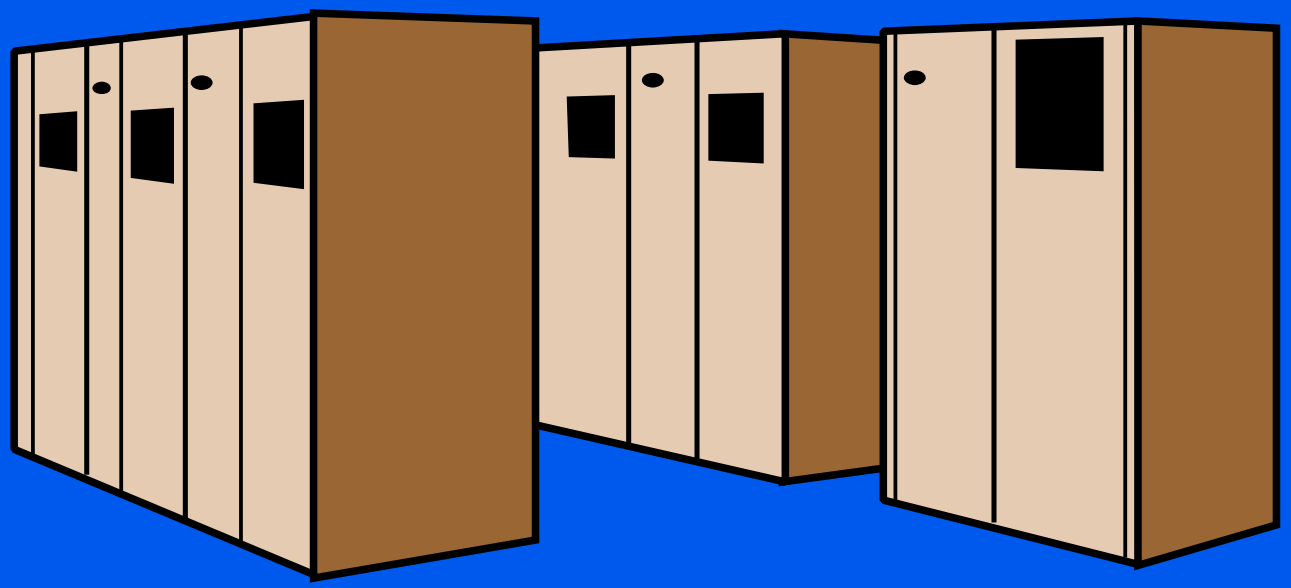
  rats need it, so do people

Manifestations

- SNA-like control of the world

  configure, configure, configure

- denial of "good enough"
- redefining reality



Internet Reality (from ‘92)

from cic nets’ Chicago hub

IP Flow Switching Cache, 16384 active flows, 0 inactive
132159644 added, 124468367 replaced, 4892577 timed out, 2782316 invalidated
statistics cleared 270640 seconds ago

Protocol	Total Flows	Flows /Sec	Packets /Flow	Bytes /Pkt	Packets /Sec	Active(Sec) /Flow	Idle(Sec) /Flow
TCP-Telnet	5222464	19.2	40	89	785.3	32.9	17.3
TCP-FTP	2087345	7.7	6	87	47.9	7.3	22.7
TCP-FTPD	1275958	4.7	95	390	449.5	21.9	23.6
TCP-WWW	83916123	310.0	9	304	2944.5	5.4	20.9
TCP-SMTP	14106833	52.1	8	173	448.9	6.4	21.6
TCP-X	94849	0.3	81	176	28.6	24.1	17.8
TCP-other	16095661	59.4	38	274	2290.8	20.9	21.5
UDP-TFTP	339	0.0	1	207	0.0	2.3	21.0
UDP-other	5059444	18.6	11	217	208.4	9.4	26.0
ICMP	4201689	15.5	2	83	46.0	5.2	26.8
IGMP	39809	0.1	30	398	4.4	48.2	29.4
IPINIP	9431	0.0	1808	254	63.0	147.1	18.6
GRE	32811	0.1	594	204	72.0	62.1	18.8
IP-other	909	0.0	3	223	0.0	1.2	31.8
Total:	132143665	488.2	15	260	7389.7	0.0	0.0

More Reality

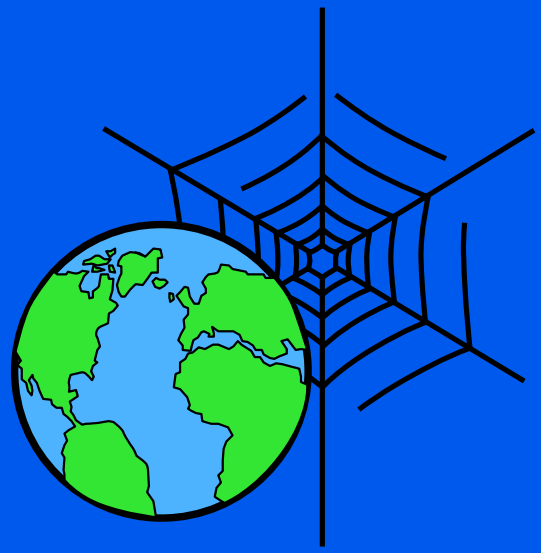
- on FIX-west & MAE-west connection
  - research net & commercial net interconnect
  - generally a low traffic point
    - restricted AUP on fed nets

```
10/28/98 - 10:26 am - 5 min sample
   average pps                - 12,546
   max # active flows         - 73,317
   average # active flows     - 70,713
   average # new flows/sec    - 764
```

Design for What World?

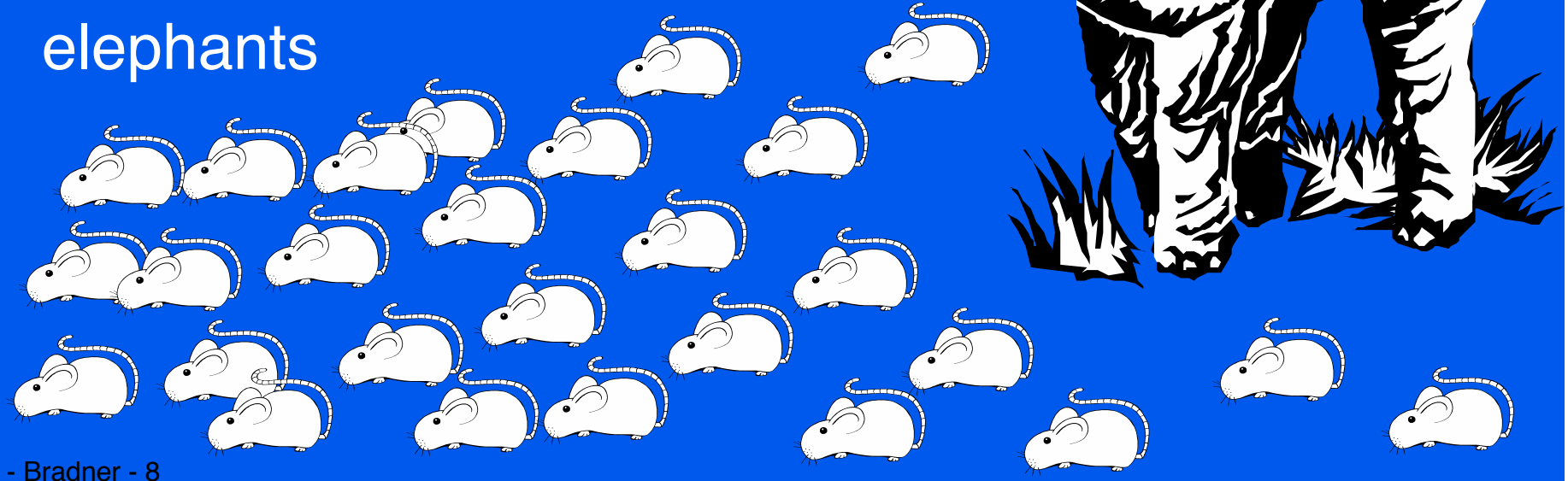
- is a flow-based QoS system the answer?
- the Internet is not about long-lived flows
- phone calls vs web traffic
- VPNs vs email
- but still do want good quality web & email
- remember reality

  not a bad design goal



Implications of Reality

- per-flow reservations can work some of the time
  - VPNs
- but not for everything
  - email, web, EDI etc are mice
  - per-flow reservations are good for elephants



Other Fun Aspects of Flows

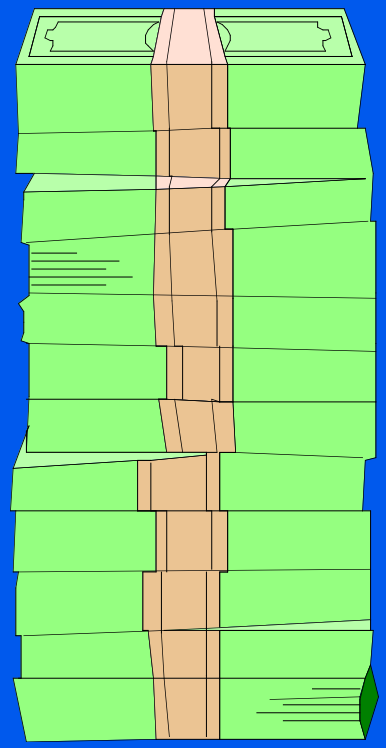
- AAA

  authentication, authorization & accounting

- not *too* hard with phone calls

  multi second setup time

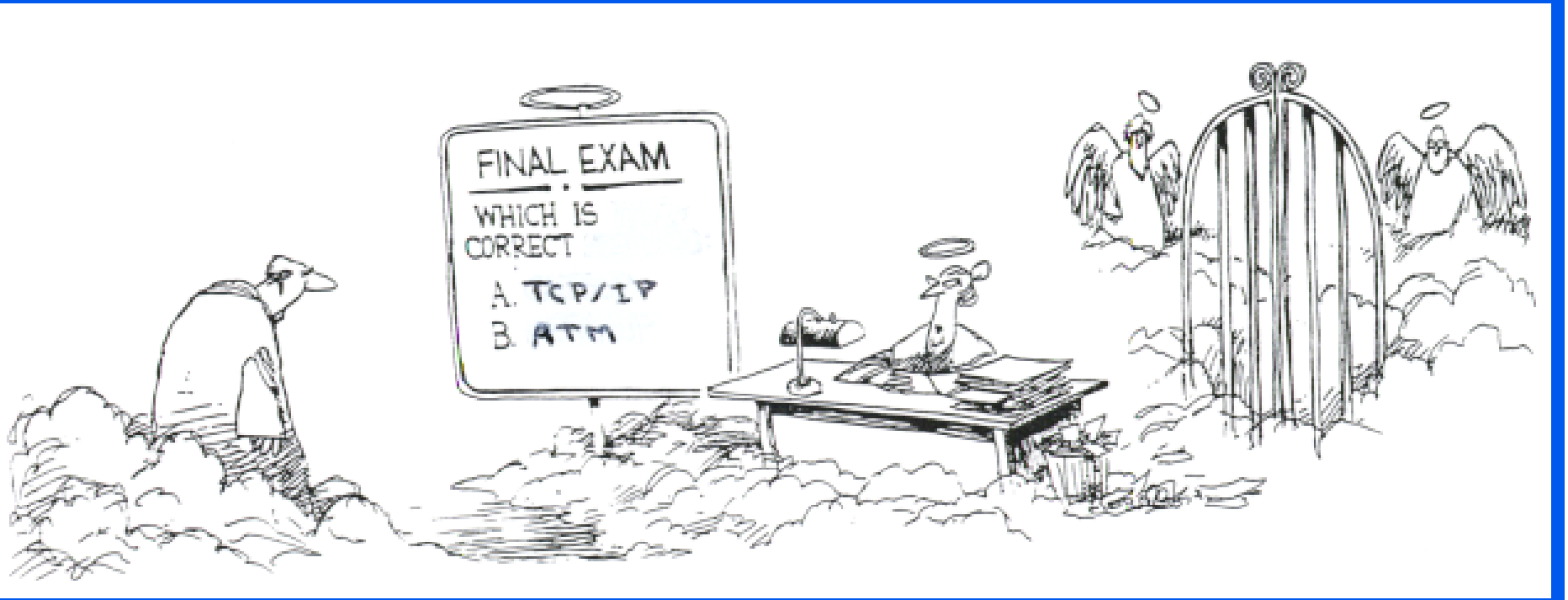
  significant % of phone co costs



Note on ATM as *the* Answer

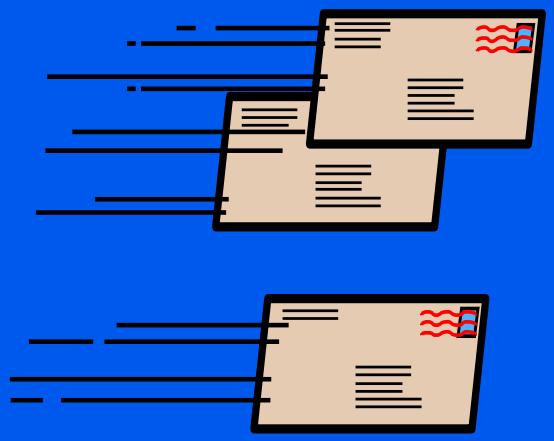
- ATM QoS designed to be end-to-end
- how many expect ubiquitous ATM soon?
- edges are ABA (anything but ATM)

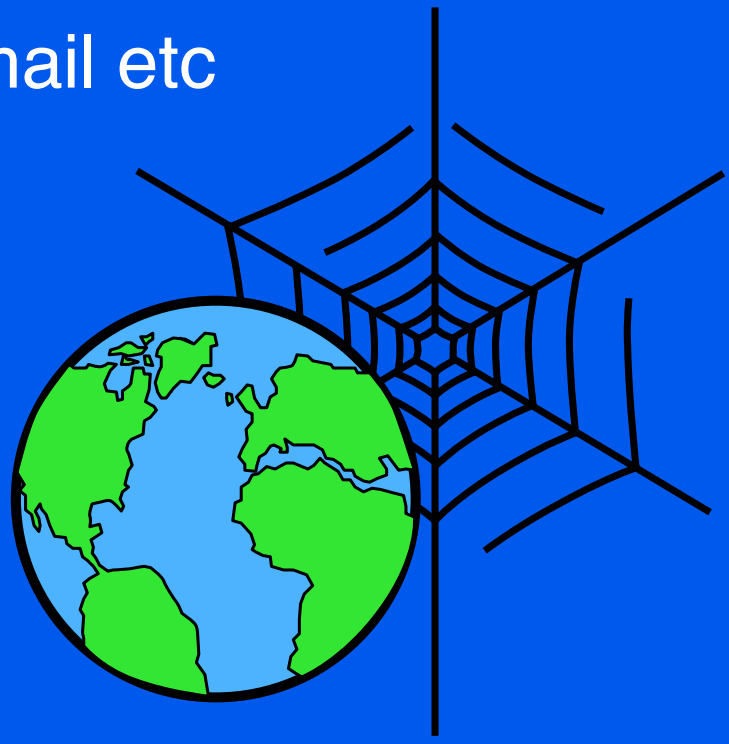
  actually Ethernet (using 80/20 rule)



If Ideal Can't Work . . .

- use flow-based where reasonable
  - long-lived flows
  - aggregate flows - router to router
- give up on rest?
  - note "rest" includes all web, email etc
  - that would be a shame





Class-Based

- separate traffic into classes
- police ingress traffic
  - per class contracts
- at congestion points
  - process packets based on class

Selecting Classes

- if evaluate at each hop

  add complexity to core

- mark at edges

  move complexity to edges where knowledge is

Diffserv

- IETF differentiated services working group
- redefine use of part of "TOS Byte" in IP header
  - high-order 6 bits now “DS Field”
- currently focused on packet processing in congested hops
  - “congested” means not enough time to send data that needs to be sent
  - other functions later
    - e.g. edge shapers & conditioners

Per Hop Behavior

- difserv defining per hop behaviors (PHBs) not services
- not enough bits to define services
  - create many services from simple PHBs
    - by changing edge functions
- map between bit pattern (code point) & PHB
  - to permit flexibility by ISP

Code Points (so far)

- 000000 = best effort
- xxx000 = compatible with precedence bits
- 101100 = Expedited Forwarding (EF)
- 01xxx0 = Assured Forwarding group

Precedence Compatibility

- set of 7 relative priority code points
- compatible with historic use of IP precedence bits
- assumes input policing
- 111000 & 110000 currently used for routing

Expedited Forwarding

- can be used to build a low loss, low latency, low jitter, assured bandwidth, end-to-end service
- "virtual leased line"
- requires strong policing at edges

  like a leased line has - drop excess traffic
- fun time allocating between customers

  “bandwidth brokers” proposed

Assured Forwarding

- set of code points
- define 4 classes
  - could be queues
- define 3 drop preferences per class
  - within contract
  - within burst contract
  - exceed contract

Should I Trust You?

- who marks packets?
- AAA problem
- policing problem
- can the host be trusted?

- (same question about end to end VCs)

Preserving the Stupidity of the Net

- Internet != phone network
- very old argument - Baran, 1964

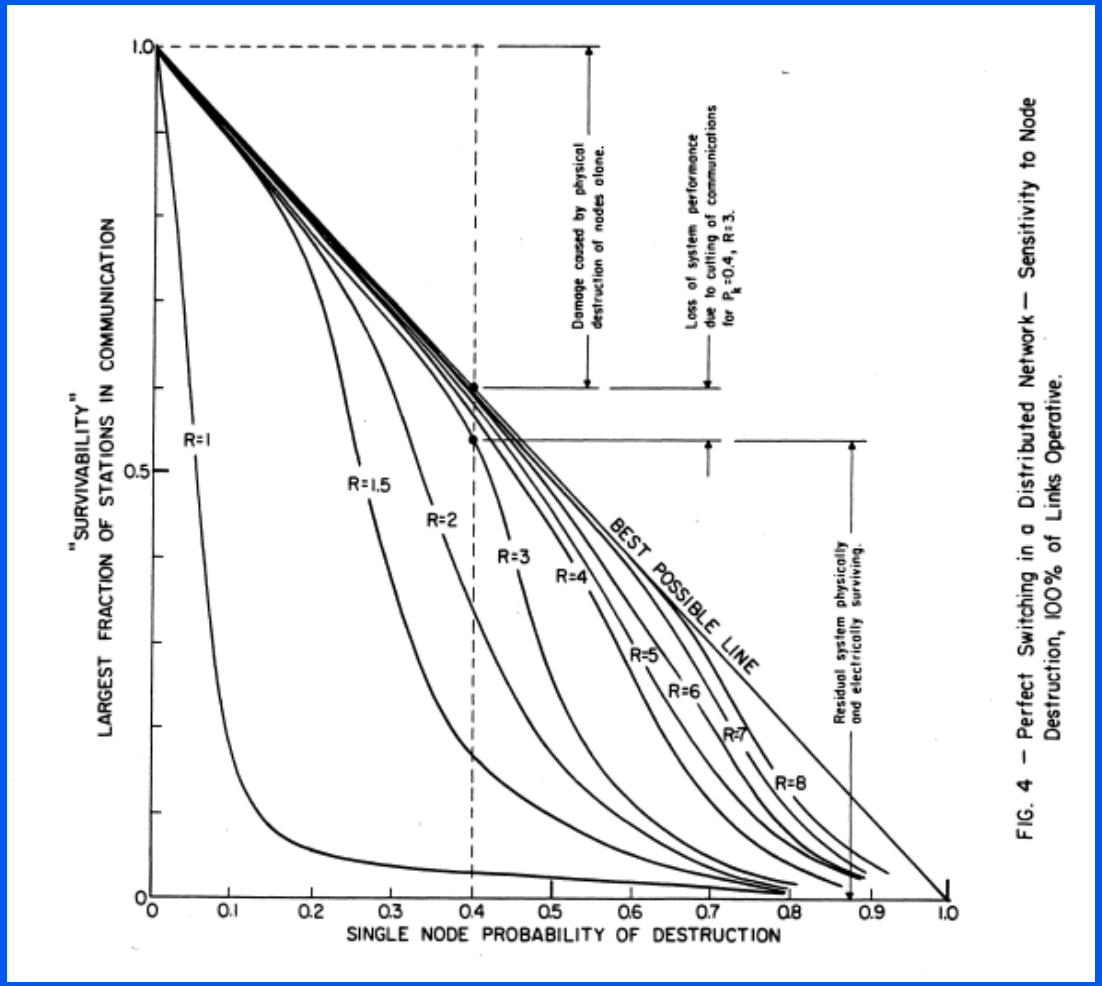
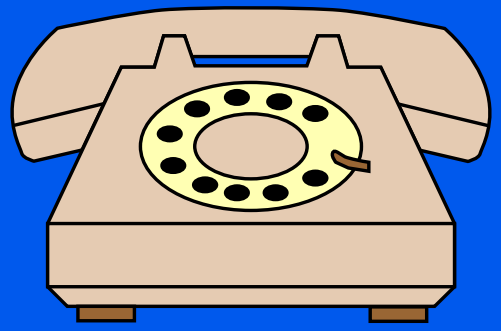


FIG. 4 — Perfect Switching in a Distributed Network — Sensitivity to Node Destruction, 100% of Links Operative.

The Intelligent Network

- phone company-speak for phone network
- implication is that the supplier knows what you want (need)
- you must want new things slowly

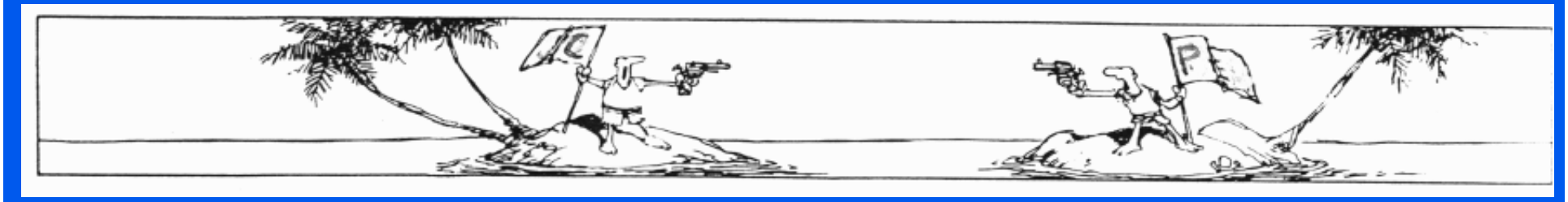


Circuit vs. Packet

- "real QoS" == VCs == circuit switching
- Internet == packet switching
- IBM once said "can not build corporate net with TCP/IP"

  have now seen the light
- but some others still have not understood Baran

  phone co planning process is "careful"



Answering John's Question

- "yes"

  circuit-based QoS for long lived things

  class-based for other "important" traffic

  best-effort for remainder