Building and maintaining a measurement infrastructure,

What the RIPE NCC did right and did wrong.

Henk Uijterwaal
RIPE NCC
5 April 2007

This talk

- Introduction
  - Disclaimer
  - What is the RIPE NCC
  - How we entered the measurement arena
- Measurement services we offer
- Do try this at home
- The world changes
- Conclusions

Disclaimer

- 10 years in the Internet measurement business
- Personal opinion of what we did right and could have done better
- The views in this talk do not necessarily represent the views of the RIPE NCC

RIPE and RIPE NCC

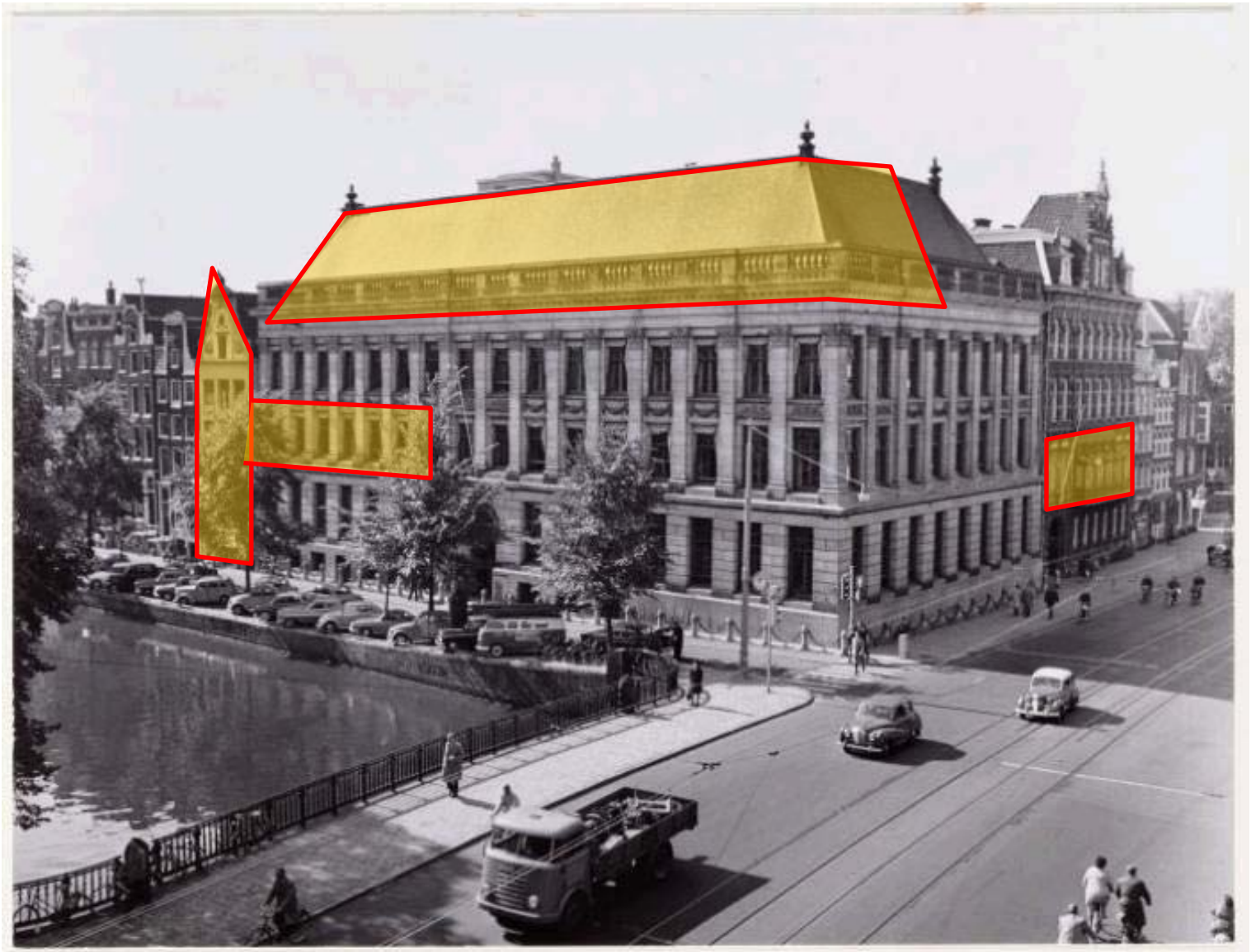
- 1989/RIPE: Réseaux IP Européen
  - Informal organization of people interested in wide area IP based networks
  - Platform for the administrative and technical coordination necessary to operate the Internet within the RIPE region
  - No formal membership
  - Volunteers doing work in working groups and through mailing lists

- Some activities became more and more work

RIPE and RIPE NCC

- 1991/RIPE NCC: RIPE Network Coordination Centre
  - RIPE ≠ RIPE NCC
  - Organization to perform activities that its members need to organize as a group, even though they are competitors in other areas
  - Membership association: ± 4500 members
    - ISP's, Telco's, research networks, corporations
  - Neutral, independent and not-for-profit
  - 100 staff from about 20 countries
  - Located in Amsterdam, NL

Our office



RIPE NCC's services

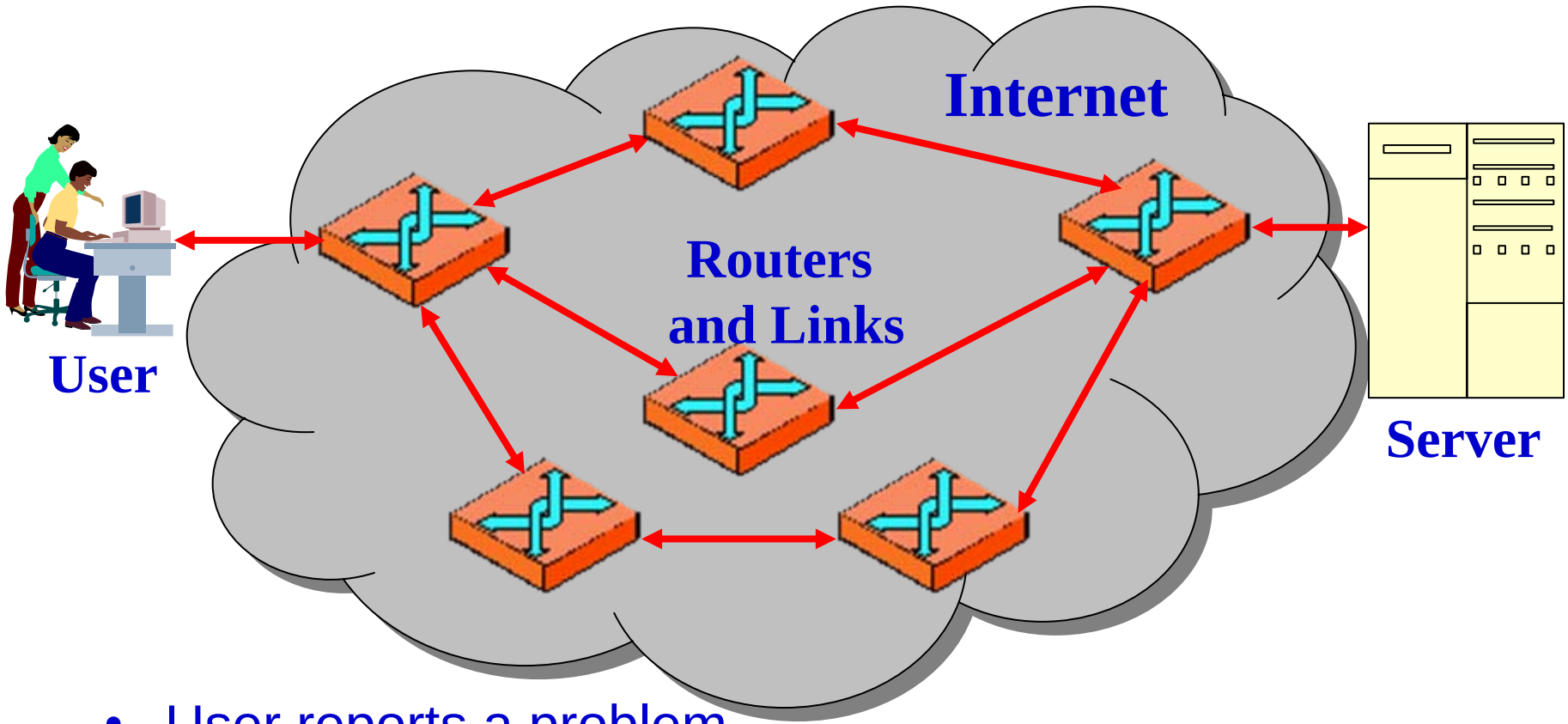
- Regional Internet Registry (RIR)
  - 1 of 5 worldwide
  - IP and AS registration (1/3…1/2 of staff)
- "whois" data-base with operational information
- k-root servers
- Training courses
- Administrative support for RIPE
- Liaison with EU, gvmts, ICANN, IETF, ITU, …
- *New projects and Information Services*

New Projects / Information Services

- NP until 2006, IS from 2007
- Goal: Collect data on the Internet and make it available to interested parties
  - ISP's to researchers, and everything in between
- Current activities:
  - RIS: Routing Information Service
  - TTM: Test Traffic Measurements
  - DNSMON: Monitor root servers
  - Hostcount and statistics on the Internet

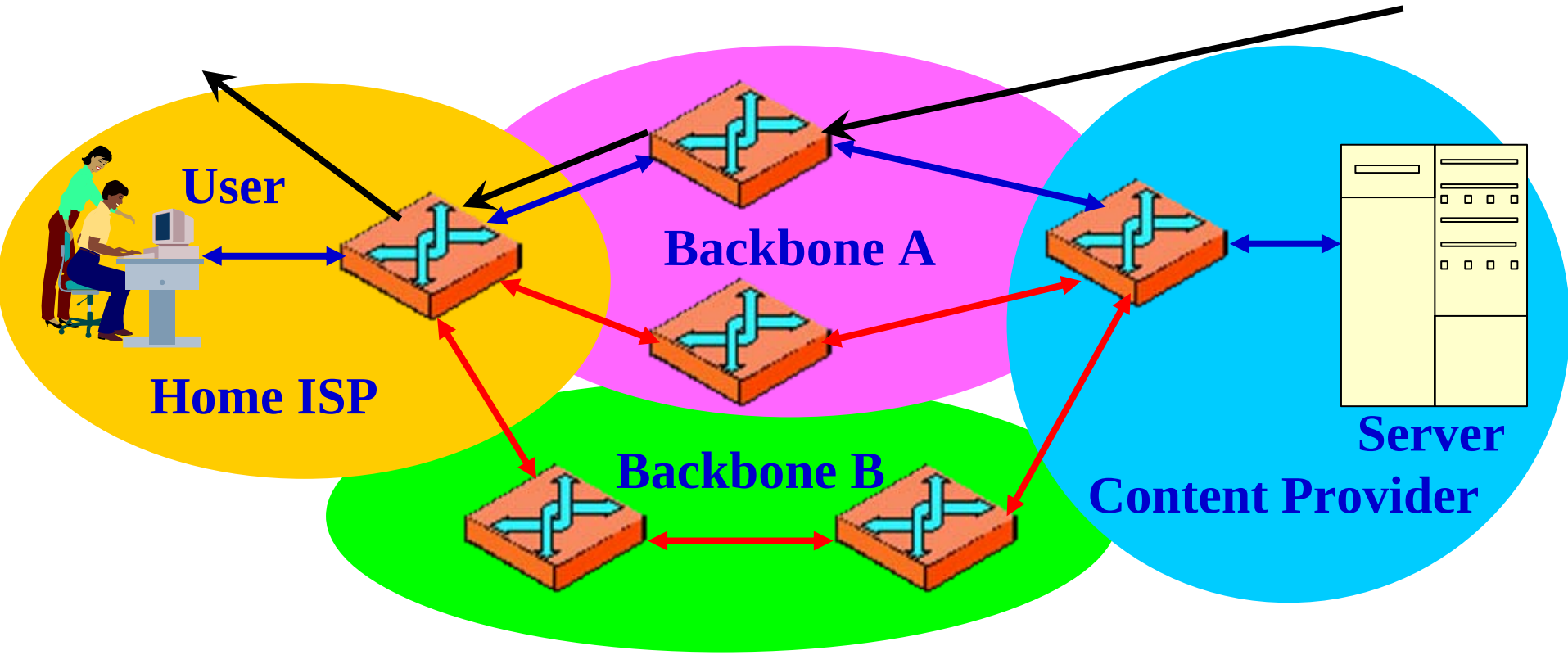
But why are we doing this?

User: the Internet has a problem



- User reports a problem
- Operator looks at devices and fixes it
- Right?

Real situation



- Different providers: What is actually installed?
- How to get access to PoPs and devices?
- Lot of work for each ISP
- History and long term trends are interesting too

Why are we doing this?

- Operators need a 3rd party to do measurements
  - Neutral and trusted
  - Access to sites of competitors
  - Load sharing, one organization doing the work for everybody
- The RIPE NCC meets all these criteria
- The result
  - Correct data without a commercial bias
  - Available for all

General requirements for measurements

- We are funded by network operators
- This has some consequences for what we can do:
  - Primary users are operators
  - Data should be applicable to day to day operations
  - Need tools that can be used in NOC's
  - Focus on practical measurements
  - General interest in long term trends

This talk

- Introduction
- Measurement services we offer
  - TTM: Test Traffic Measurements
  - DNSMON: DNS Monitoring Service
  - RIS: Routing Information Service
- Do try this at home
- The world changes
- Conclusions

TTM: Performance measurements

- Main features:
  - One way end-to-end measurements
  - Dedicated measurement devices and infrastructure
    - RIPE NCC Test Box
    - Black box (PC + GPS)
    - Look at results only
  - Active measurements, “Real traffic”
  - Follows well defined IETF standards
  - Focus on external networks

TTM: Things we measure

- *Routing information:*
  - Router/Interface level
  - AS-level
- *Delay or Latency*
- *Packet Loss*
- *Bandwidth*
- *Derived quantities:*
  - IPDV or Jitter
  - Packet reordering
  - Protocol specific performance
  - …

Full analysis for each Test Box

TTM Summaries for ttXX.ripe.net

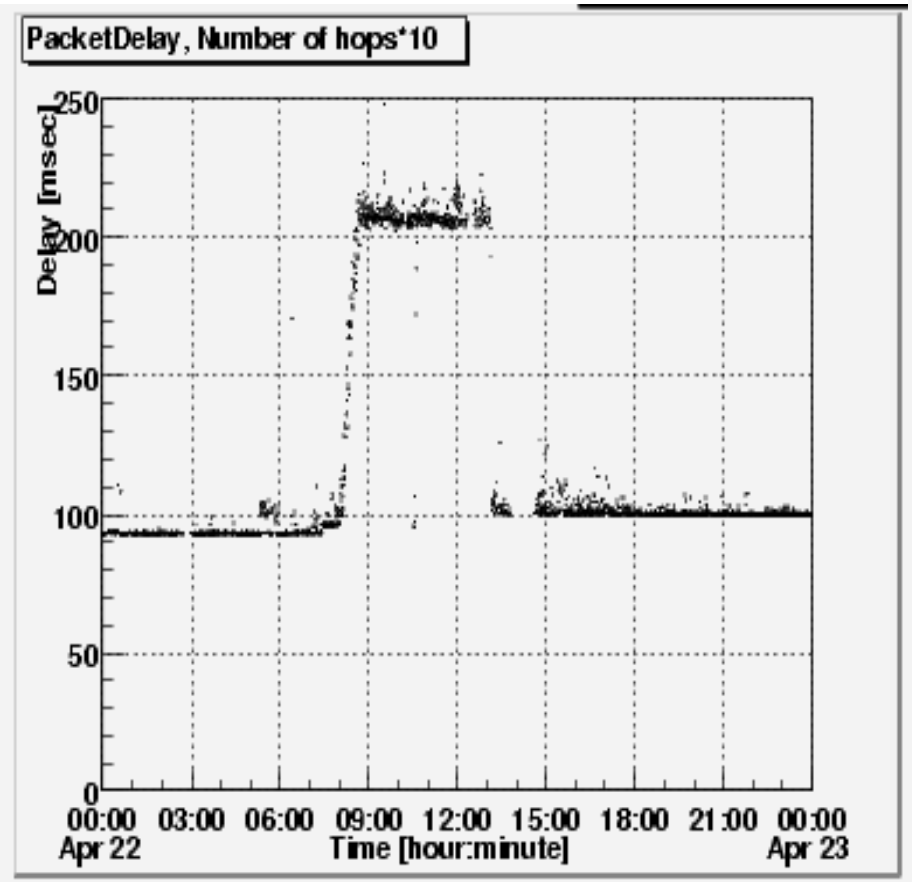
Monday 22 April 2002

The table below displays the TTM delay & loss parameters [...]

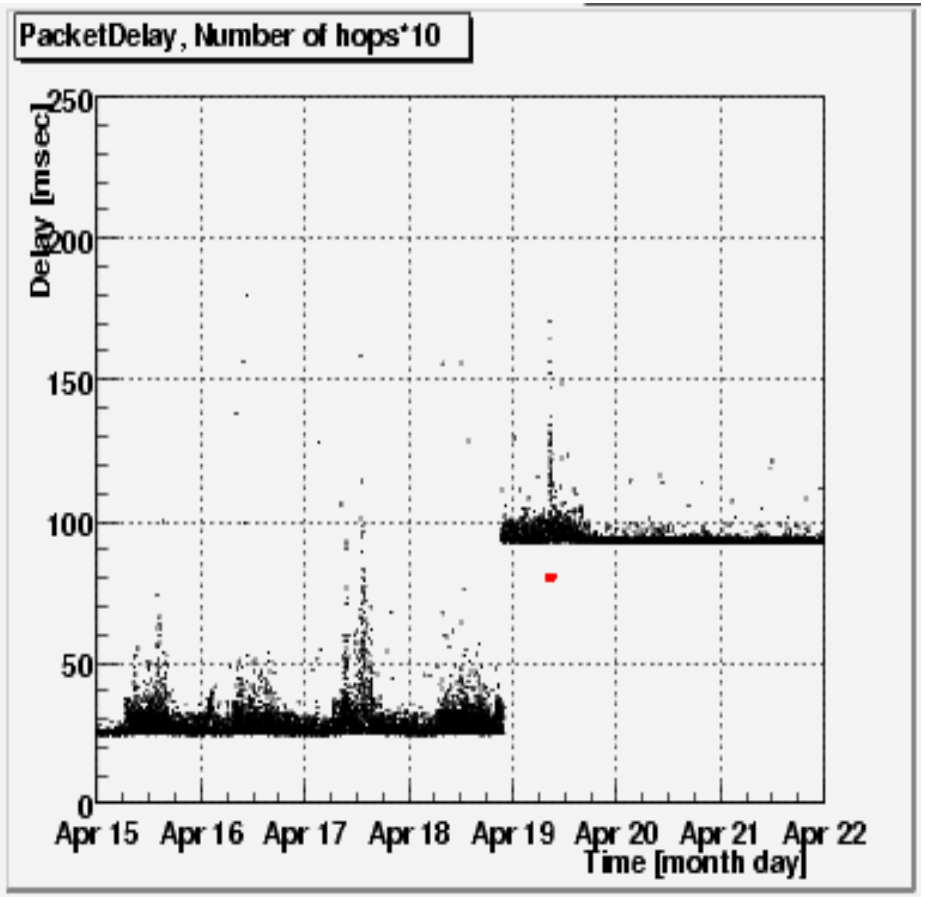
	Incoming delay and loss				Outgoing delay and loss			
	Minimum	Median	Maximum	Loss	Minimum	Median	Maximum	Loss
tt01	22.59 (27.92)	24.08 (28.57)	60.35 (54.21)	0.432 (0.770)	27.96 (28.03)	28.38 (28.46)	32.17 (32.17)	0.206 (0.422)
tt07	23.32 (23.32)	23.47 (23.52)	32.79 (40.84)	0.104 (0.104)	109.0 (110.7)	112.5 (112.5)	136.4 (136.1)	0.0698 (0.0710)
tt08	19.31 (19.38)	19.51 (19.56)	30.06 (30.06)	0.0349 (0.104)	19.49 (19.27)	19.74 (20.03)	26.49 (30.60)	0.0349 (0.0348)
tt12	12.83 (12.90)	13.27 (13.30)	23.22 (25.47)	0.0347 (0.0695)	12.88 (12.66)	13.27 (13.82)	20.53 (24.65)	0.00 (0.105)
tt13	11.50 (11.57)	11.64 (11.70)	21.78 (21.80)	0.210 (0.210)	11.72 (11.48)	11.90 (12.02)	18.63 (22.40)	0.245 (0.175)
tt25	92.95 (24.89)	100.0 (26.85)	211.7 (94.31)	0.454 (0.0699)	27.42 (27.11)	27.81 (28.19)	95.39 (41.72)	0.174 (0.0694)
tt26	18.10	18.87	29.68	0.672	95.90	96.15	103.0	1.15

Zoom in

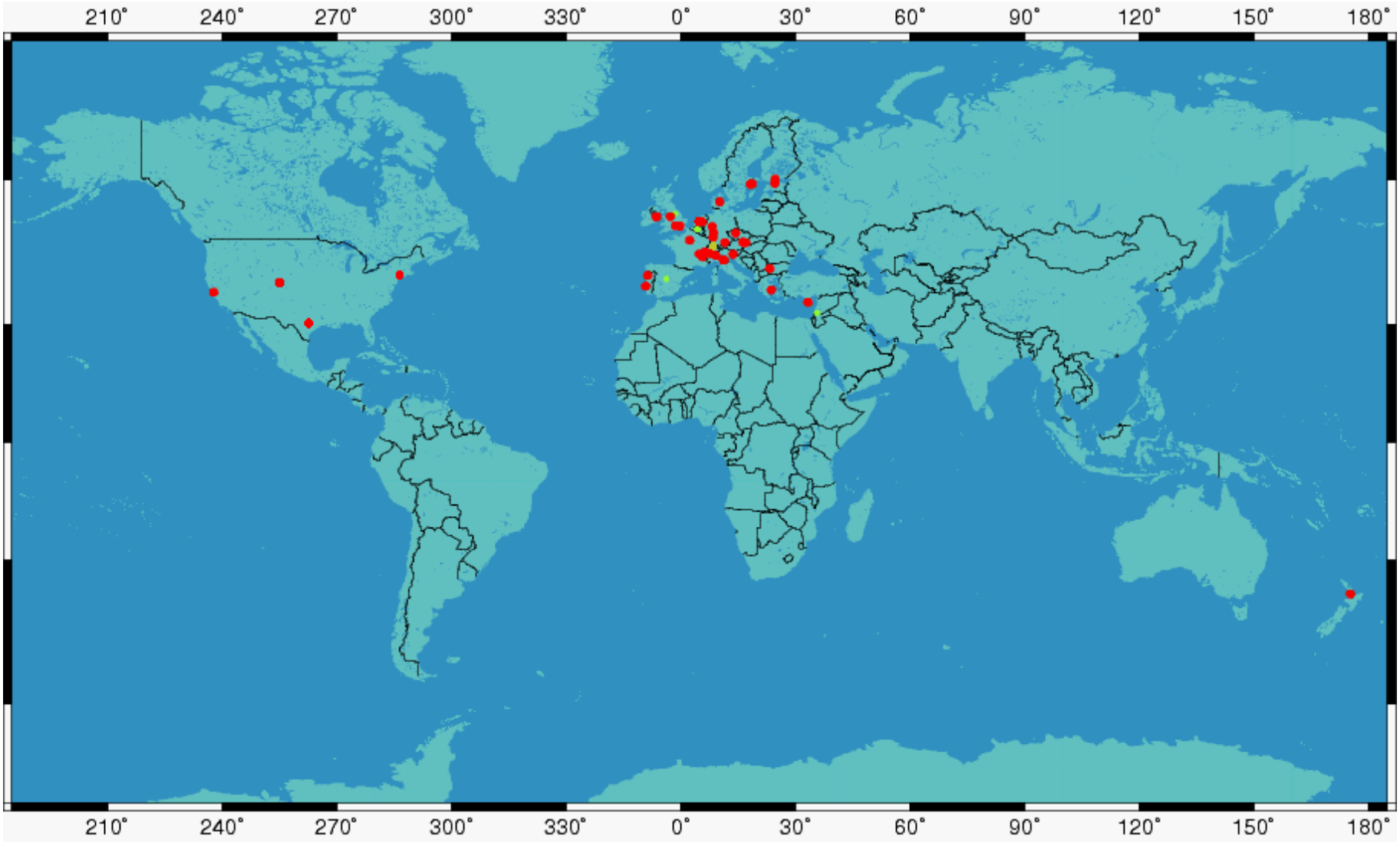
Last Day



Last Week



Where are the test-boxes located?

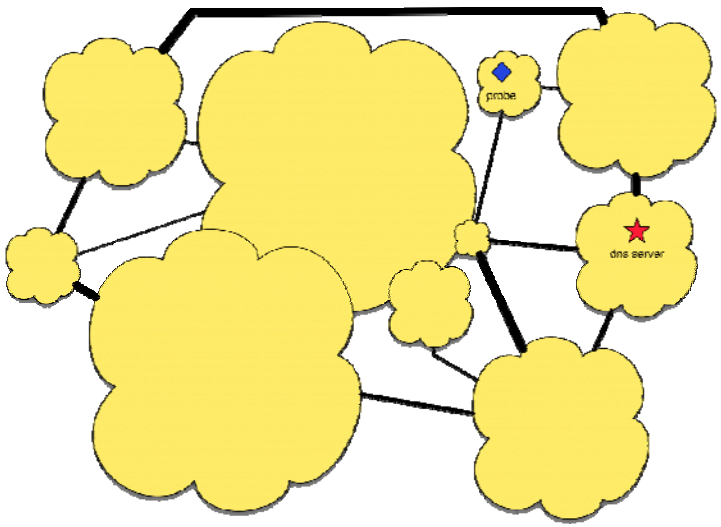


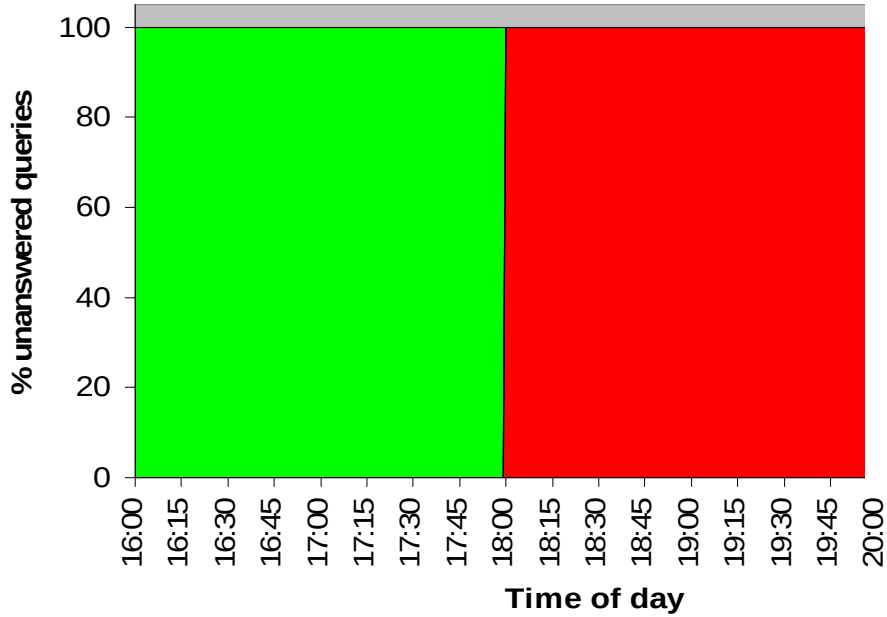
DNSMON: Monitoring of DNS

- We have this distributed measurement platform
- Can we use it for other purposes?
- Newspaper article on reachability of root servers

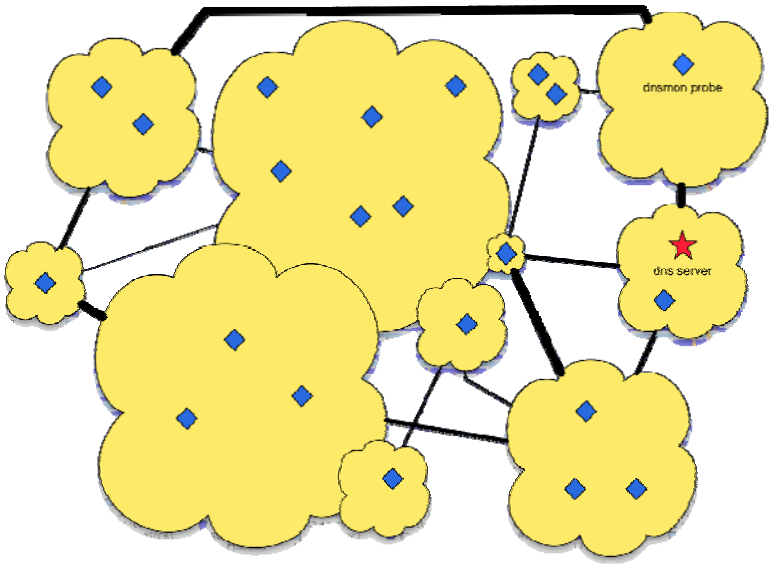
What the article did…

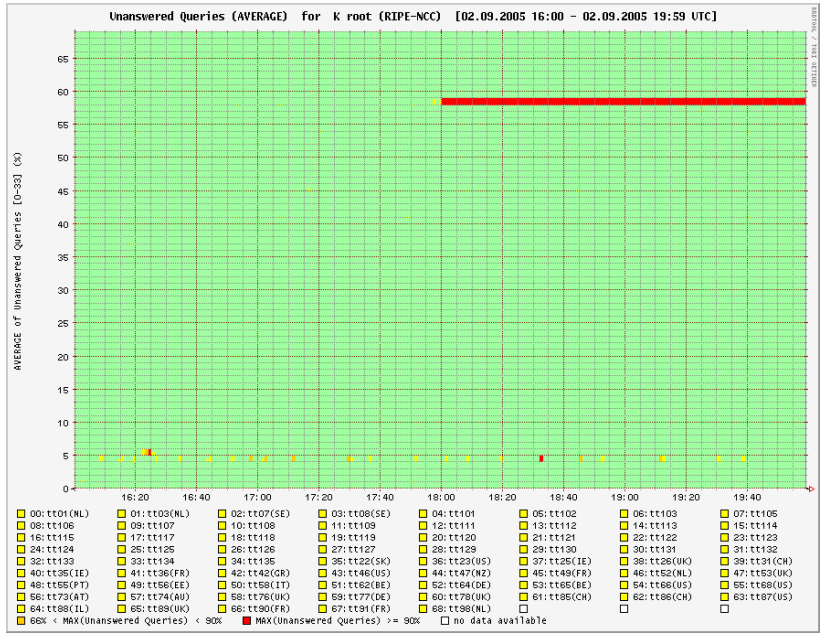
- User queries a root server
  - From his home machine
  - 100% loss after 18:00
  - Where is the problem?
- What he did:



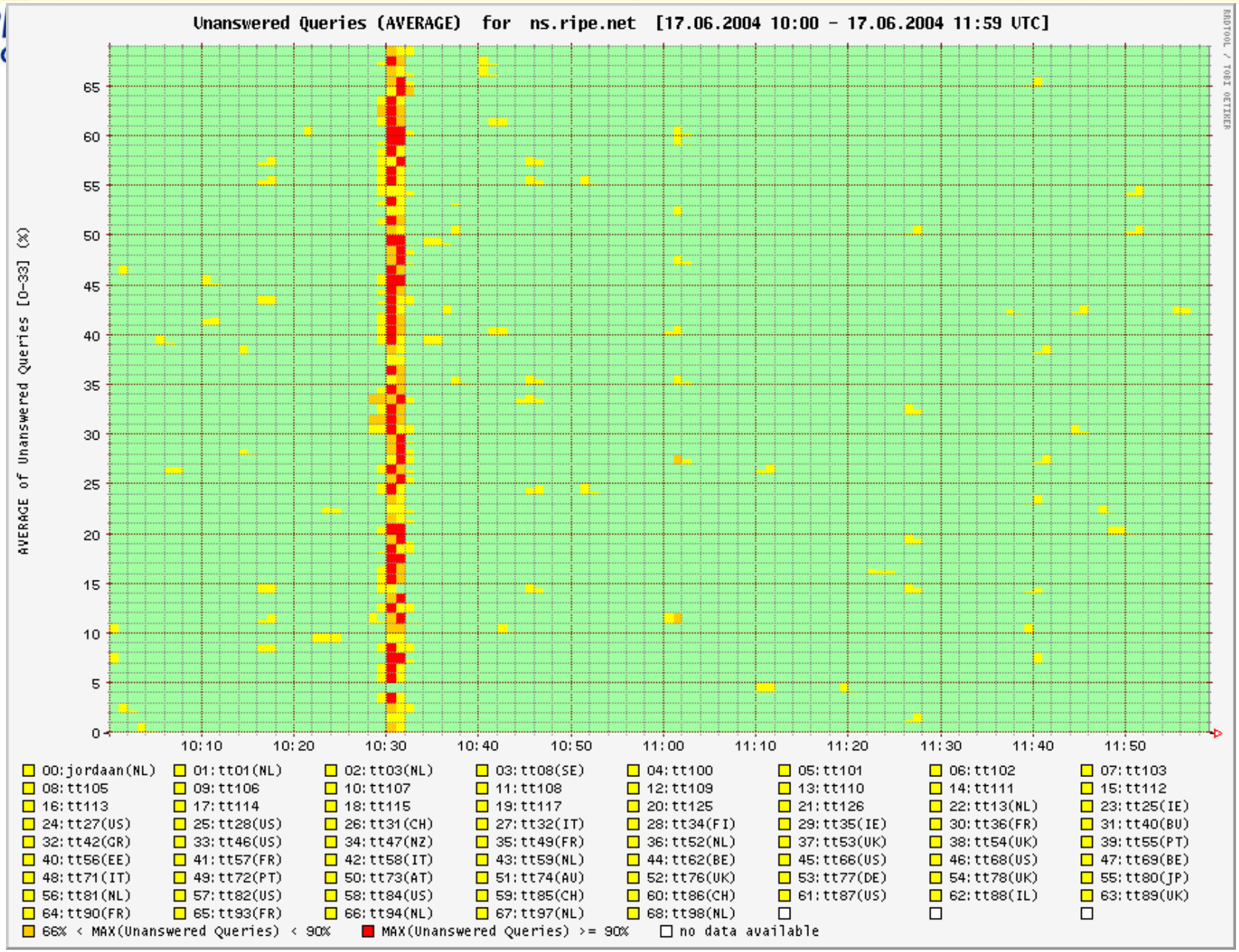


Measure from multiple locations





- We can test from 100 sites
- Stack the plots
- This is clearly a local problem
- Counter example on the next slide


Unanswered Queries (AVERAGE) for ns.ripe.net [17.06.2004 10:00 - 17.06.2004 11:59 UTC]
AVERAGE of Unanswered Queries [0-33] (%)
0
5
10
15
20
25
30
35
40
45
50
55
60
65
10:10
10:20
10:30
10:40
10:50
11:00
11:10
11:20
11:30
11:40
11:50
00:jordaan(NL)
01:tt01(NL)
02:tt03(NL)
03:tt08(SE)
04:tt100
05:tt101
06:tt102
07:tt103
08:tt105
09:tt106
10:tt107
11:tt108
12:tt109
13:tt110
14:tt111
15:tt112
16:tt113
17:tt114
18:tt115
19:tt117
20:tt125
21:tt126
22:tt13(NL)
23:tt25(IE)
24:tt27(US)
25:tt28(US)
26:tt31(CH)
27:tt32(IT)
28:tt34(FI)
29:tt35(IE)
30:tt36(FR)
31:tt40(BU)
32:tt42(GR)
33:tt46(US)
34:tt47(NZ)
35:tt49(FR)
36:tt52(NL)
37:tt53(UK)
38:tt54(UK)
39:tt55(PT)
40:tt56(EE)
41:tt57(FR)
42:tt58(IT)
43:tt59(NL)
44:tt62(BE)
45:tt66(US)
46:tt68(US)
47:tt69(BE)
48:tt71(IT)
49:tt72(PT)
50:tt73(AT)
51:tt74(AU)
52:tt76(UK)
53:tt77(DE)
54:tt78(UK)
55:tt80(JP)
56:tt81(NL)
57:tt82(US)
58:tt84(US)
59:tt85(CH)
60:tt86(CH)
61:tt87(US)
62:tt88(IL)
63:tt89(UK)
64:tt90(FR)
65:tt93(FR)
66:tt94(NL)
67:tt97(NL)
68:tt98(NL)
66% < MAX(Unanswered Queries) < 90%
MAX(Unanswered Queries) >= 90%
no data available

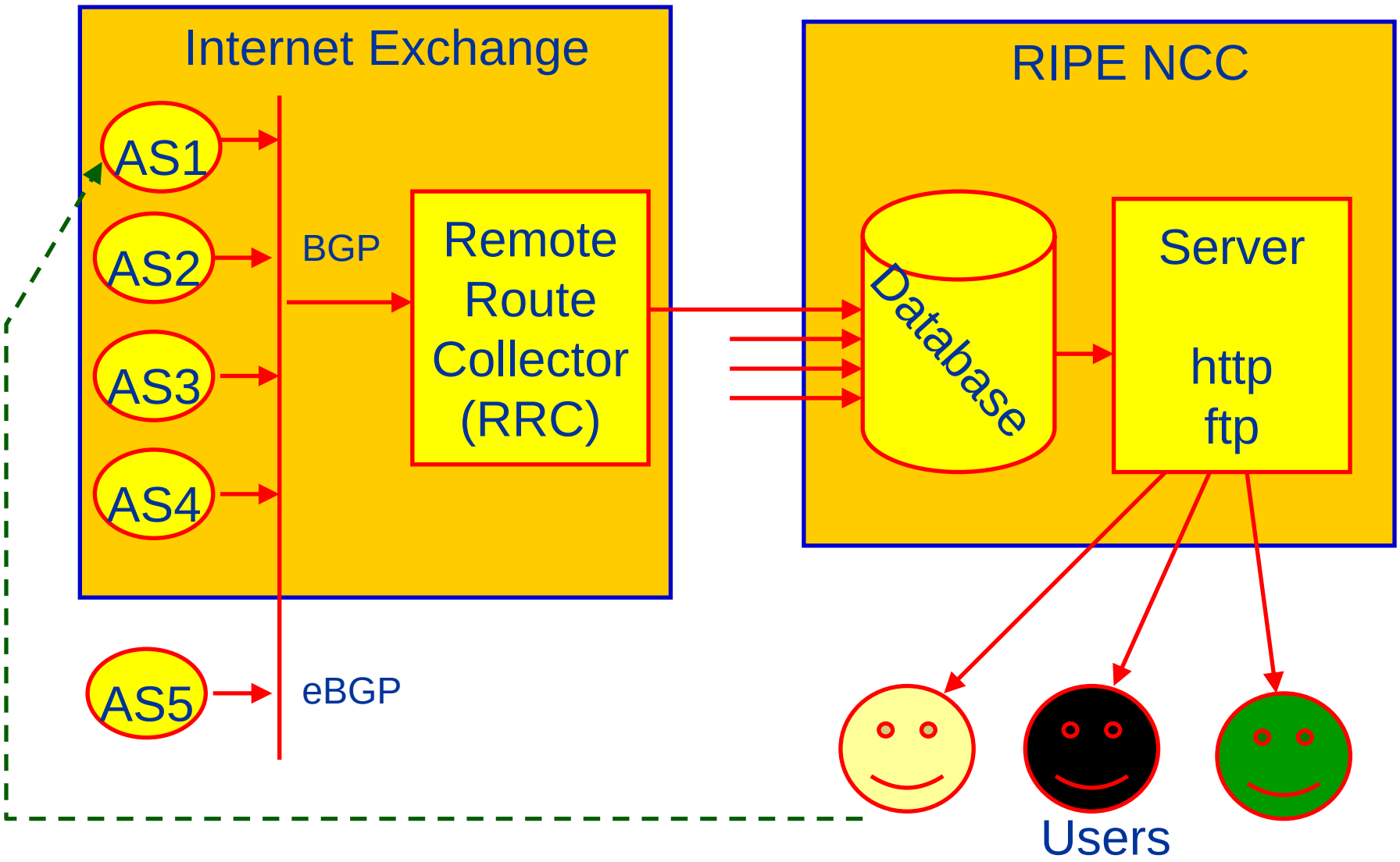
Routing Information Service (RIS)

- BGP routing can be very complex: 40000 AS's
- Classic tool: looking glass
  - Shows BGP status of a router
  - Problems:
    - Access often limited
    - If routing doesn't work, you cannot reach the LG either
    - Have to query multiple LG's in order to find the problem
    - No history mechanism
- Can this be improved?

Yes, the RIS

- A device that collects BGP information at multiple points on the Internet
- Includes history information
- Available to the entire community
- Set of tools to access the information suited to various audiences

RIS: Basic Architecture



Applications

- Daily operations in a NOC
  - Query DB
  - Graphics
  - myASn
  - …
- Statistics
  - RISreport
  - Hot Spots
  - Martians
  - Black holes
  - …
- Research @ RIPE NCC
  - Debogon
  - Beacon prefixes
  - Activity Index
  - …
- Other
  - Map IP to AS
  - …
- *Research outside NCC*

This talk

- Introduction
- Measurement services we offer
- Do try this at home
  - Short term
- The world changes
- Conclusions

So, you want to try this at home?

- Do a measurement on the Internet
  - Multiple machines
  - Long term (>1 year)

- Common factors in our setup
  - Distributed setup
  - Sit in PoPs of 3rd parties
  - Little operator support
  - Collects data about infrastructure
  - Offered as a regular service

Things to consider

- Hardware
- Time keeping
- Software
- Practical issues

Buying hardware

- PC's are cheap these days
- Very rapid changes in PC hardware
  - What you find in a box changes very rapidly
  - A 905A isn't a 905B
  - Compatible doesn't always refer to Unix
  - Often needs an OS upgrade to keep things running
- Recommendations:
  - Make software kernel independent
  - Do not rely on specific features

Distributing hardware

- Approach one: Buy centrally, install and ship
  - Configuration and quality control is easier
  - Cheaper due to volume discounts
  - Can pre-configure boxes, plug and play
- Approach two: Have site buy hardware according to specs
  - Requires international vendor
  - Is the hardware really the same?
  - No issues with import/export procedures (outside EU)
- It is a trade-off you have to make

Hardware doesn't live forever

- Our initial model:
  - If a box breaks, return it to us and we'll fix it
  - Return it if there is a problem
- This was horribly impractical:
  - Box breaks
  - Shipping takes time and costs money
  - We're not hardware shop either
  - Ended up calling the manufacturer for support under warrantee anyway
  - Shipping equipment back costs time and money again
  - Needed reinstallation

Hardware doesn't live forever (2)

- A much better model:
  - A number of vendors sell identical hardware world-wide
  - Buy from them
  - Get a service contract and transfer it to the hosting site
  - If something breaks, they come out and fix it locally
- Advantages:
  - Faster
  - Cheaper

Hardware needs to be replaced

- Think about a replacement strategy
  - We didn't…
- The older the hardware, the more problems
  - Things break
  - No spare parts
- Need an upgrade path
  - Think about this in advance
  - Otherwise your experiment will slowly break down
  - Details depend on financial arrangements

GPS

- GPS is great for time keeping
  - Accurate
  - Works everywhere
- GPS is hard to get deployed
  - Needs a clear view of the sky. Computer rooms are often in the basement
  - Building maintenance
  - Costs of cabling
  - Electrical issues
- Do you really need it?

GPS: Attempt 1

- Motorola Oncore
  - RF Receiver Module on a PCI board
  - Separate antenna
  - Coax cable between the two, 1.5 GHz
- Cheapest solution…
- … but not the simplest
  - Coax cable: not present in standard office buildings, hard to pull, cannot be bent
  - Cable length is limited to 10-30m, this is often not sufficient

GPS: Attempt 2

- Trimble Acutime
  - GPS unit with integrated receiver
  - Output is RS422
  - 8 wires
- Advantages
  - RS422 is low frequency, cables can be 250m long
  - Can use existing CAT5 cabling
  - Major improvement
- However, it is still hard to get this deployed

Remote support

- Remote support is hard to get
  - Installation usually OK
  - It helps if boxes are pre-configured
  - After that: operators or volunteers
    - Operators are usually busy
    - Volunteers tend to disappear

Remote support (2)

- Do not rely on this
  - Make boxes run with as little operator support
  - Reboot and restart automatically
  - Check the BIOS settings for no keyboard
  - Test things in advance on a local machine, nothing is worse than having to call lots of people to press the reset button

S/W upload

- We use a central repository for all our machines
  - General area
  - Site/machine specific extension
- Rsync to distribute software
  - rsync is an open source utility that provides fast incremental file transfer.
- Provide a boot floppy or CD
  - Install a basic O/S from scratch (after a disk failure or so)
  - Easy to recover a box
  - Available for most O/S.

System administration

- Think about scalability
- Process checking
  - cfengine
  - Generic tool to check if processes run (and restart)
  - Boxes just continue to measure
- Log files
  - Necessary to check health of systems
  - compare against “standard” log
  - automate as much as possible
  - operator can focus on unusual situations

Data Quality Monitoring (DQM)

- With lots of boxes, you'll probably automatically process the data
  - Lots of plots and results
- Really useful to do a sanity check
  - Boring to check by hand
- Automate this:
  - Check that there is data in all plots
  - Check that things are within limits
- Worth automating
  - Evolves over time

Data disclosure policy

- You are collecting data about providers
- Discuss in advance how you will publish the data
- Anonimize
  - Simple is usually sufficient
- In general this is not a problem IF you discuss in advance what you plan to do with the data

Security

- Boxes sit in the middle of a network
  - Operators don't like "black boxes" in their network
- Some solutions:
  - Strip boxes to the bare minimum you need
    - Easier to maintain, less risk of exploits
  - Close all ports that you don't use
  - Recent versions
    - Monitor relevant mailing lists
  - Security audit

Using the output

- Think about your audience
  - Tools should fit their needs
  - Tools should be understandable to them
  - Most people have very little time
- Provide feedback loops on your work
  - “User group”
  - Actively solicit feedback, don’t wait for people to speak up

This talk

- Introduction
- Measurement services we offer
- Do try this at home
- The world changes
  - Long term
- Conclusions

The world changes

- The Internet is a rapidly changing environment
- Developing a measurement service takes time
  - Make sure that your experiment will be relevant when it gets deployed
  - Make sure that you can adapt to an ever changing playing field

Things that changed over the years

- Some things that changed over 10 years

- Network capacity is an expensive resource
  - 1997: Buy as little as possible, only buy when capacity 100% used, wait for delivery
  - 2007: Prices have dropped dramatically, readily available, buy when you need it
- PoP to end-user traffic is slow
  - 1997: 28k modem was state of the art
  - 2007: ADSL and Cable are cheap and standard

Things that changed over the years (2)

- Measurements of applications are not interesting as the link to the user is the bottleneck
  - 1997: Last mile was slow
  - 2007: No longer true
- Internet is sparsely connected, routing has not been optimized
  - 1997: We see traffic going across the Atlantic between two sites a few km apart. Delays indicate non-optimal routing
  - 2007: Well connected Internet

Things that changed over the years (3)

- Real Time Traffic Engineering based on performance measurements cannot be done
  - 1997: It is acceptable to measure and have numbers tomorrow
  - 2007: Requirement for measurements with results on the spot
- IPv6
  - 1997: Only experimental
  - 2007: Production

Is this a problem?

- No
- However: developing a measurement setup takes time
- Make it flexible enough so it can adjust to changes

This talk

- Introduction
- Measurement services we offer
- Do try this at home
- The world changes
- Conclusions

Conclusions

- Overview of issues when building and maintaining a measurement infrastructure
- No clear solutions, but mainly food for thought
- Think about these things before deploying your own setup

URLs, email

- Services
  - TTM
    - http://www.ripe.net/ttm
    - ttm@ripe.net: TTM crew @ NCC
  - DNSMON:
    - http://dnsmon.ripe.net
    - dnsmon@ripe.net: DNSMON crew @ NCC
  - RIS
    - http://www.ripe.net/ris
    - ris@ripe.net: RIS crew @ NCC
- General
  - www.ripe.net
  - henk@ripe.net
- Products and tools: www.google.com


Questions?