Building a solution for active monitoring of the Swedish Internet.



RIPE 51, Amsterdam Rickard Dahlstrand rd@tilde.se

Background

- The Demeter concept was build for <u>distributed</u> measurement of availability, performance, services and infrastructure.
- 4 large installations running at Södersjukhuset,
 Sveriges Radio, Nic-se and Sitic (PTS). All these had specific need for distributed measurement.
- They are running measurements for TCP bandwidth, UDP jitter, ICMP RTT, service availability (DNS) and more..

Timeline

- 2003- A distributed dns-monitoring system at Nic-se.
- 2004- 2 installations of a distributed bandwidthmeasurement system at Södersjukhuset and Sveriges Radio.
- 2004- Generic Internet service testing platform launched for distributed monitoring of websites, mail-servers etc. (0 customer)
- 2005- Sitic phase 0 was launched, using the setup from the Internet-service.
- 2005- Sitic phase 1 in planning-stage.

The Sitic-project

What is Sitic's interest in this project?

- Everyone shall have access to a broad range of communications services.
- The services shall be easy to use, affordable and of good quality.
- The consumer shall have access to clear information about rights and obligations.
- The consumer shall be well informed in order to easily be able to choose among various services and various service suppliers.
- Sweden shall have a robust infrastructure for communications services and an active exchange of information in IT security issues.

Context

- Well, we are all living in Sweden. So the scope is quite local. =)
- Telia (Skanova) is the big player for broadband access due to it's previously tax-funded copperinfrastructure (DSL).
- Government has funded communal fiberinfrastructure. DSL is however still used in many cases for the "last mile".
- "Fiber to the home" is only available in large cities.

Internet Monitor - Phase 0

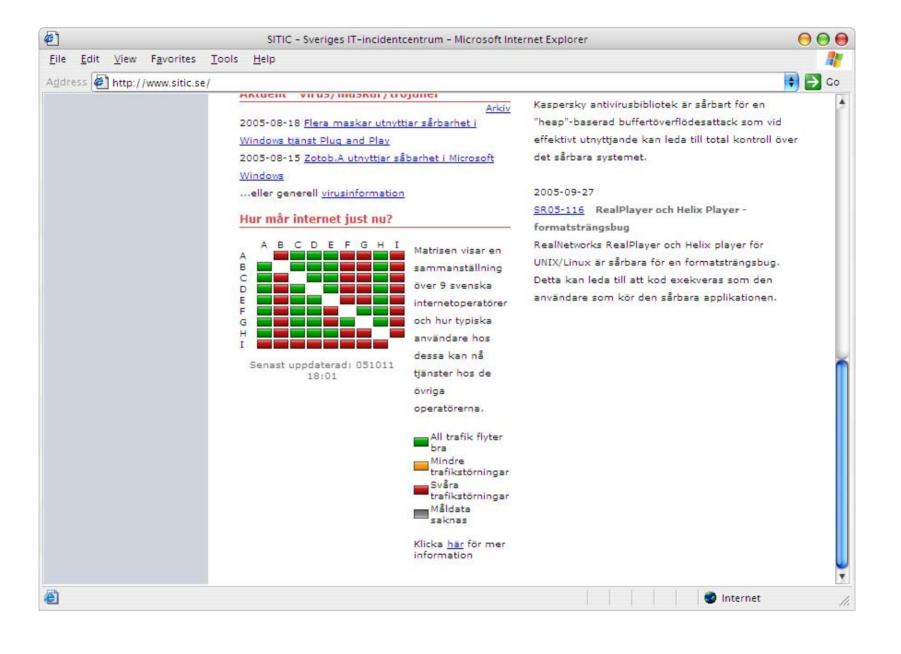
- Proof of concept!
- 9 measurement points at different Swedish internetoperators
- Low/no-cost internet connections are used (eq. to friends and family)
- Each point is pinging each other measurement point once every minute and reports back to a central server
- Sitic rented the system, Tilde runs it.
- Launched Mars 2005
- EOL Nov 2005

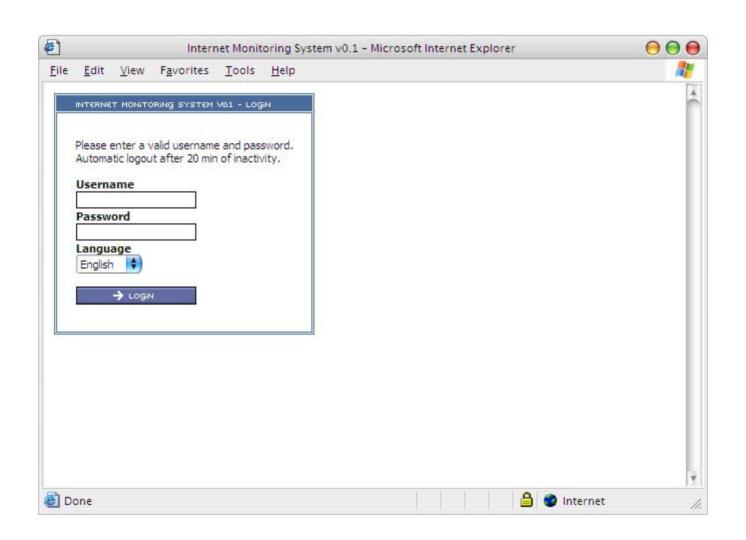
Goals for phase 0

- Does this type of measurement system work as a concept?
- Can we draw any conclusions from these tests?
- How is peering and transit affecting availability?
- Are there availability problems between operators?

The system

DISCLAMER! These measurements does not represent any real operator availability!





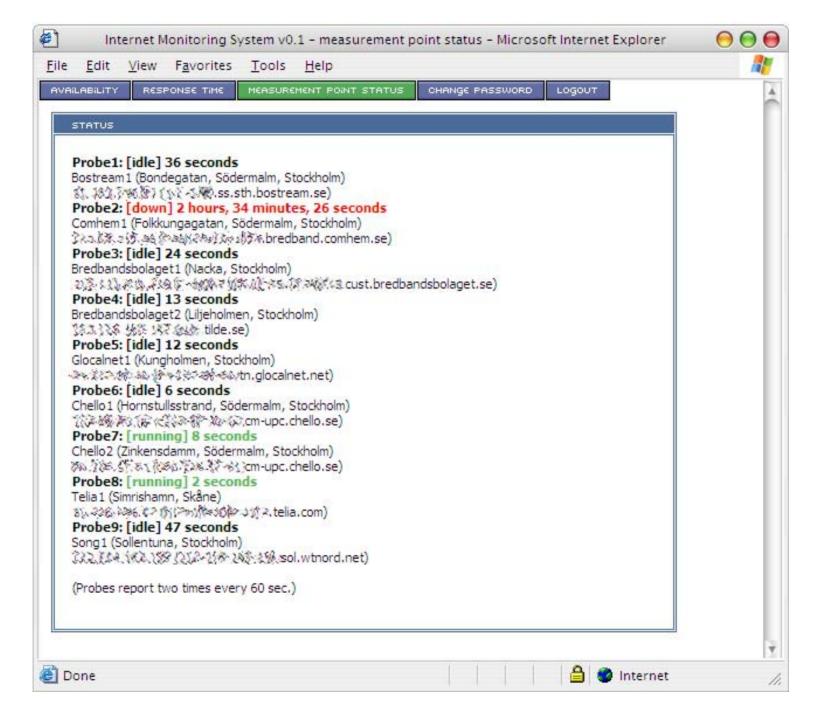
Measurement collection

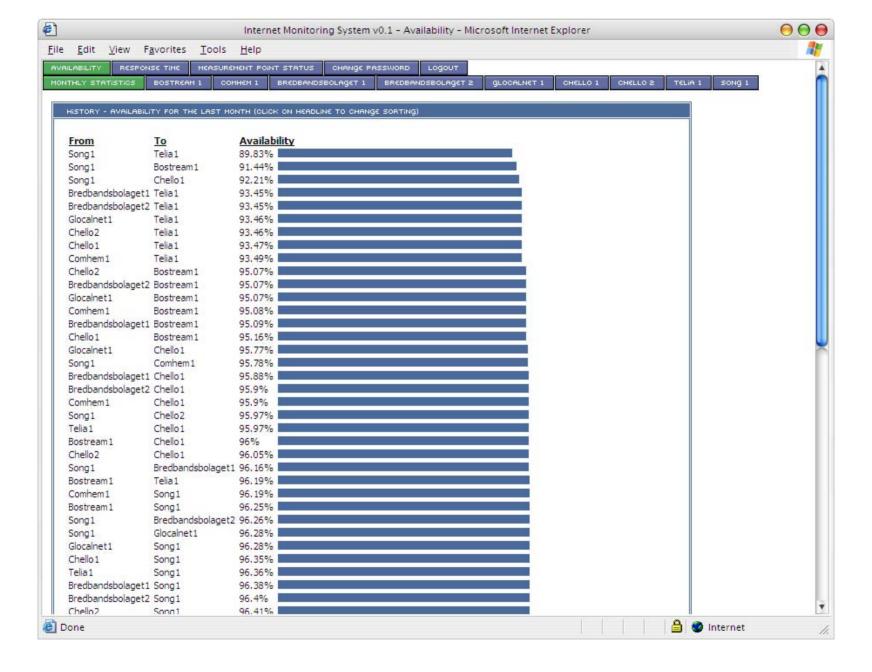
Each minute a script is run at the measurement point.

- The scripts start by sleeping a pre-defined number of seconds according to which probe is running on. (To allow for measurements to be done without affecting each other.)
- 2. The script connect to the server using https and gets a list of IP's to measure to. (Most ip's are dynamic)
- 3. 5 ICMP-messages are sent to each IP and RTT max,min,avg and stddev are collected.
- 4. All measurement-data is sent back to the server using https and added to the database.

```
#!/bin/sh
#
 Demeter Test System (c)Tilde 2004
# $Id: demeterechoperf.sh,v 1.1 2005/02/01 06:26:06 rd Exp $
PATH=/sbin:/bin:/usr/sbin:/usr/bin:/usr/local/bin
version="02"
myname='/bin/hostname' 2> /dev/null
ping="/sbin/ping"
awk="/usr/bin/awk"
wget="/usr/local/bin/wget"
case "$myname" in
probe1)
             del=0;;
probe2)
             del=6;;
probe3)
             del=12;;
probe4)
             del=18;;
probe5)
             del=24;;
probe6)
             del=30;;
probe7)
             del=36;;
probe8)
             del=42;;
probe9)
             del=48;;
*)
             del=0;;
esac
sleep ${del}
```

```
# Load servers from server-file
servers=`${wget} https://XXX.XXX.XXX.XXX/getservers?probe=${myname} --no-check
    certificate --tries=3 -timeout=3 -q -0-`
# Add the versionnumber, date and time first to the data.
demres=`date "+%Y%m%d;%H%M%S"`
demres="${version};${demres}"
# loop through all the servers in the list of servers
for serv in ${servers}; do
    pingsrv=`echo "${serv}" | awk -F \; '{print $2}'`
    serv=`echo "${serv}" | awk -F \; '{print $1}'`
    # Run ping and connect to serv using ICMP with a 5 queries
    # i=pause w=timeout -n -q =quite c=number of pings
    x= \{ping\} -i.1 -w2 -n -q -c 5 \{pingsrv\} 2> /dev/null | grep round-trip| awk
     '{print $4}' | awk -F / '{print $1"; "$2"; "$3"; "$4}'`
    # If successful add the server-ip and response time to the data.
    if [ "\{x\}" != "" ]; then
        demres="${demres};${serv};${x}"
    else
        demres="${demres};${serv};U;U;U;U"
    fi
done
resp=`${wget} "https://XXX.XXX.XXX.XXX/reportdta?probe=${myname}&result=${demres}"
     --no-check-certificate --tries=1 --timeout=30 -q -O-
exit 0
```

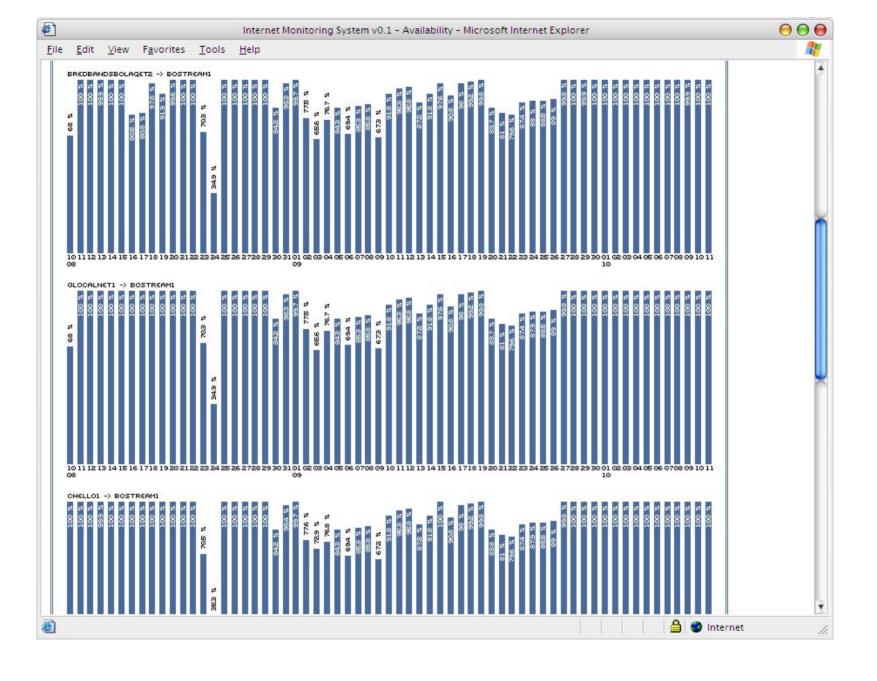


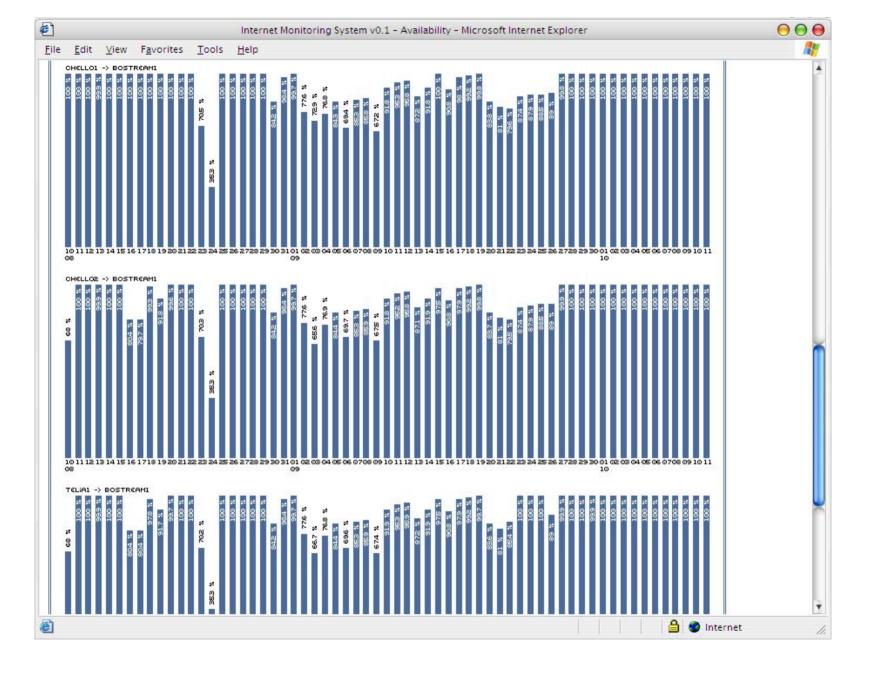


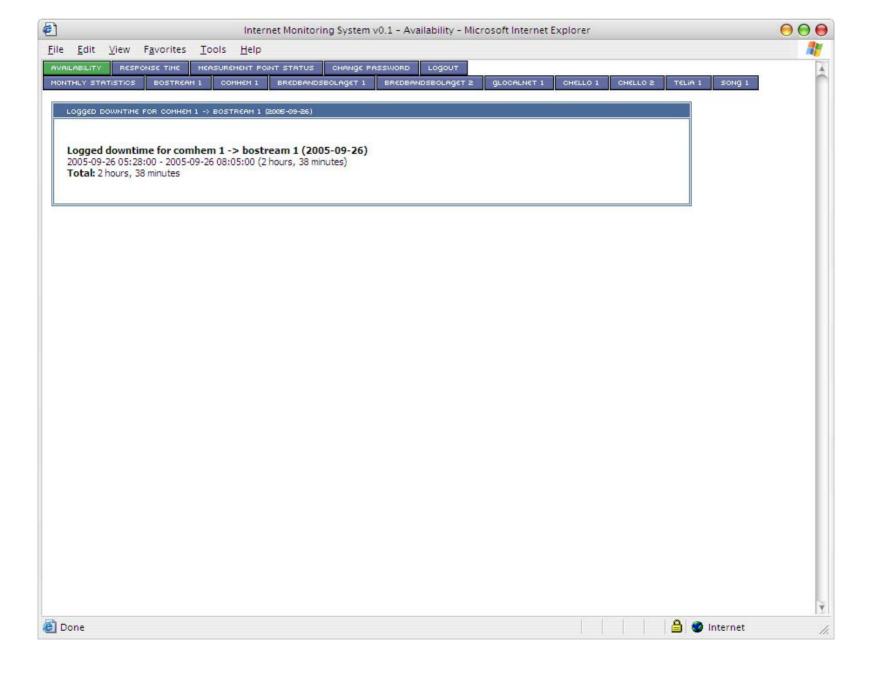
Client connectivity problems

- Crashing (ADLS/Cable)-modems
- "last mile" infrastructure problems (ADLS/cable mainly)
- Unscheduled maintenance (There are no scheduled!)
- Power outages
- Portal login-problems
- DHCP-problems
- Problems in the operators distribution-infrastructure
- Operator buy-out (Infrastructure moved to new core-infrastructure)
- Vacuum-cleaners that need the power outlet. =)



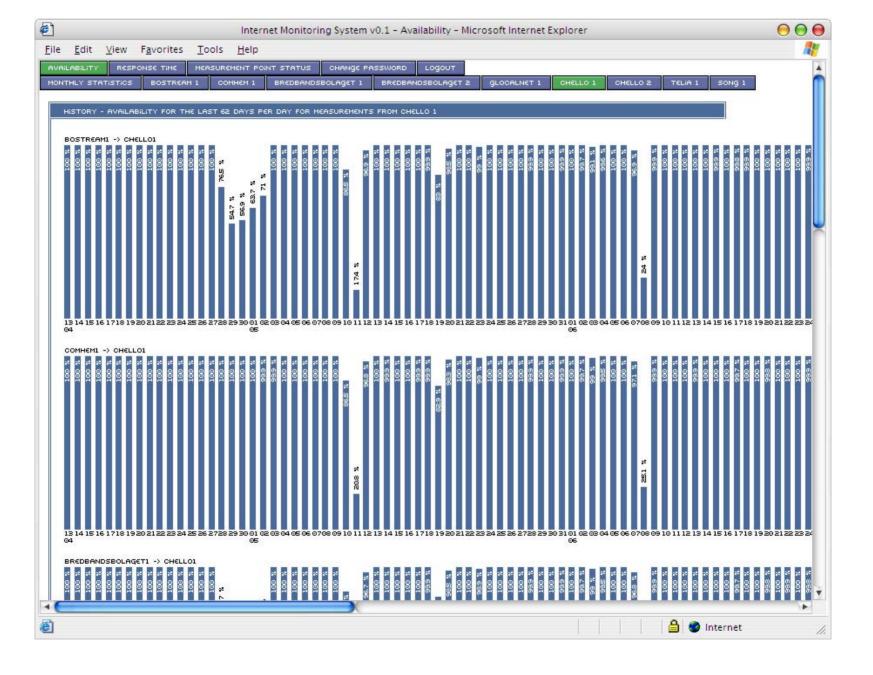




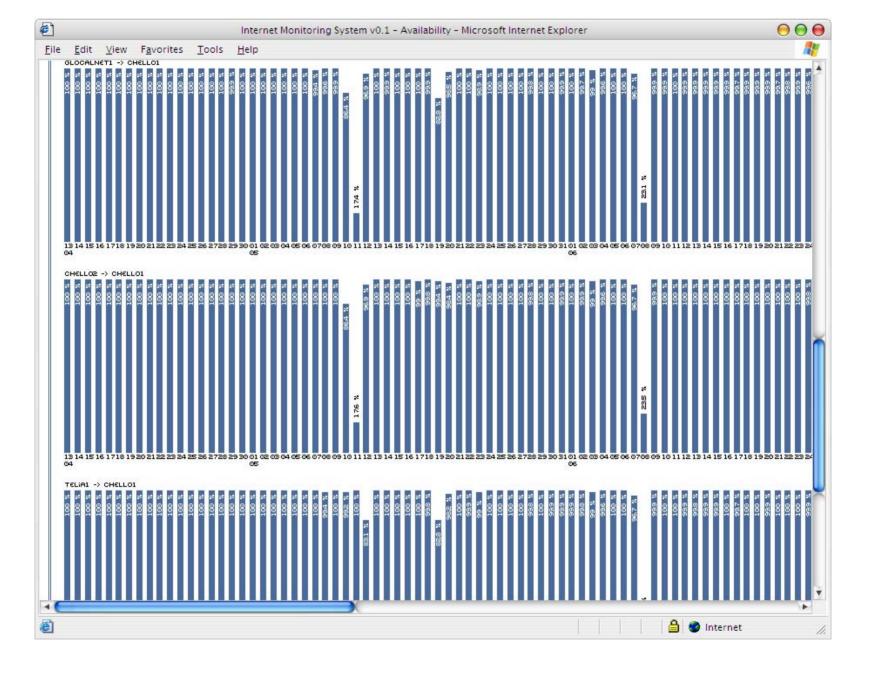


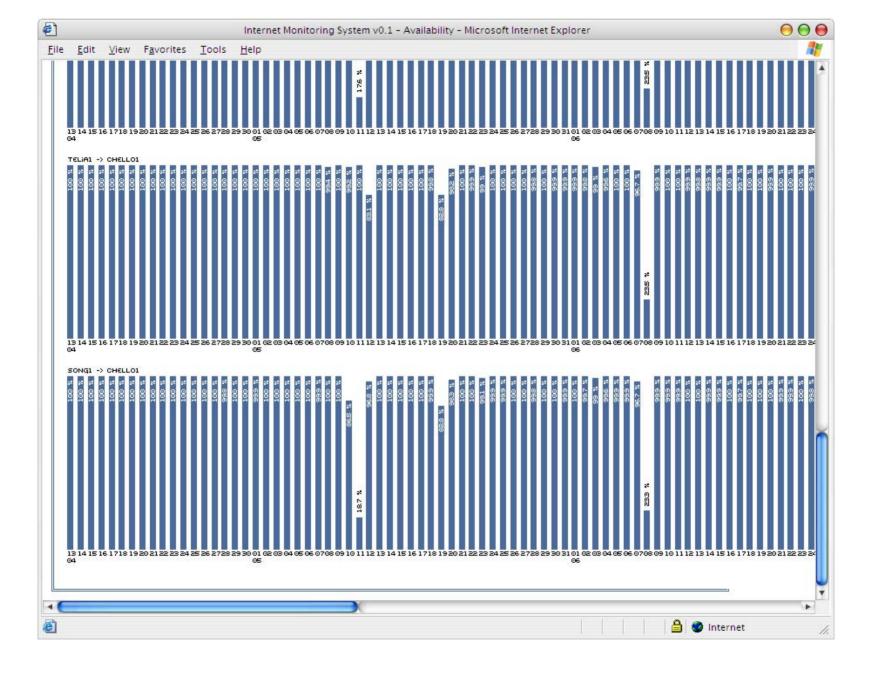
Routing problems between operators

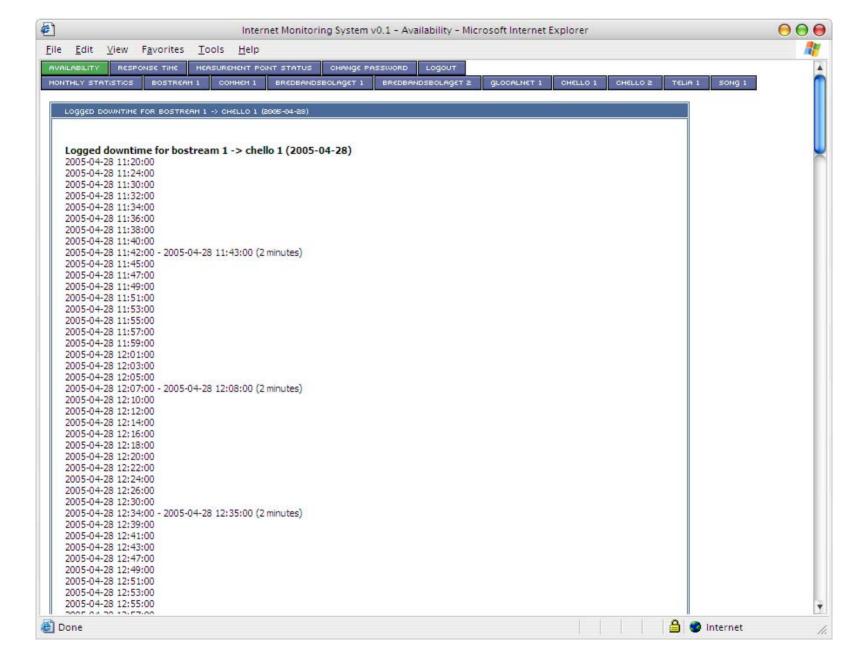
- Problems with peering and transit.
- Discovered when a measurement point has contact with some operators, but not others.
- These problems is most likely affecting a larger number of users.

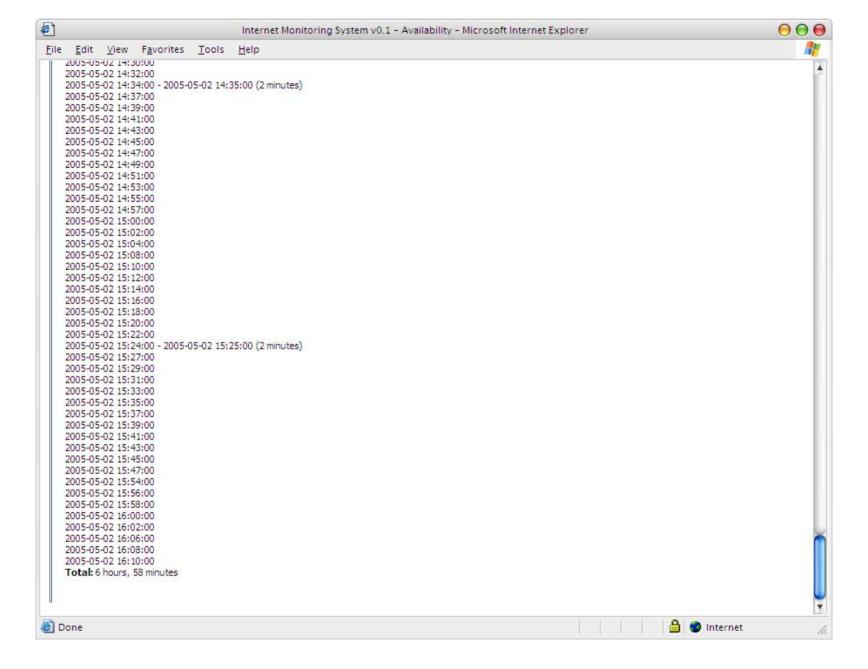








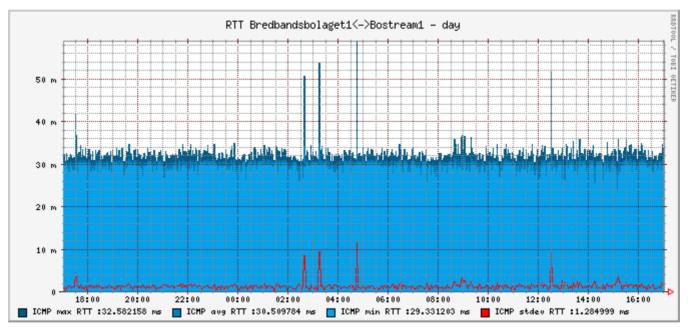




RTT/Jitter measurements

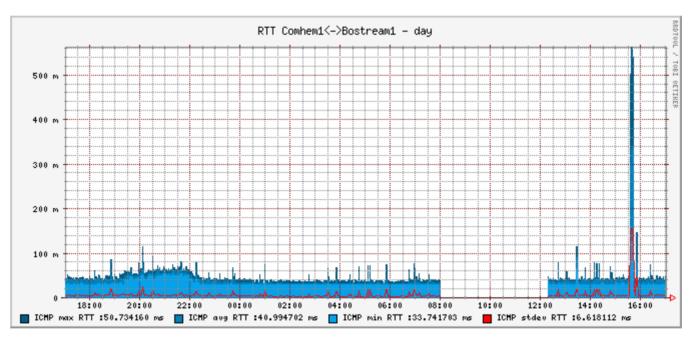
- Especially interesting for VoIP applications.
- Not a good idea for connections in use.
- Roundtrip times for what? ICMP, UDP, TCP??
- Tools, kernels and drivers?

Good RTT



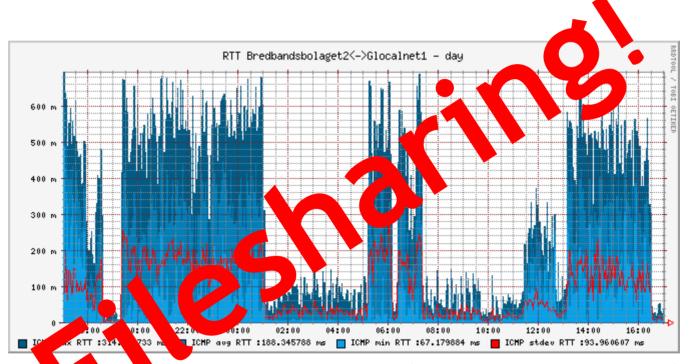
(Check the scale! 60 ms)

Bumb, downtime & peak



(Check the scale! 600 ms)

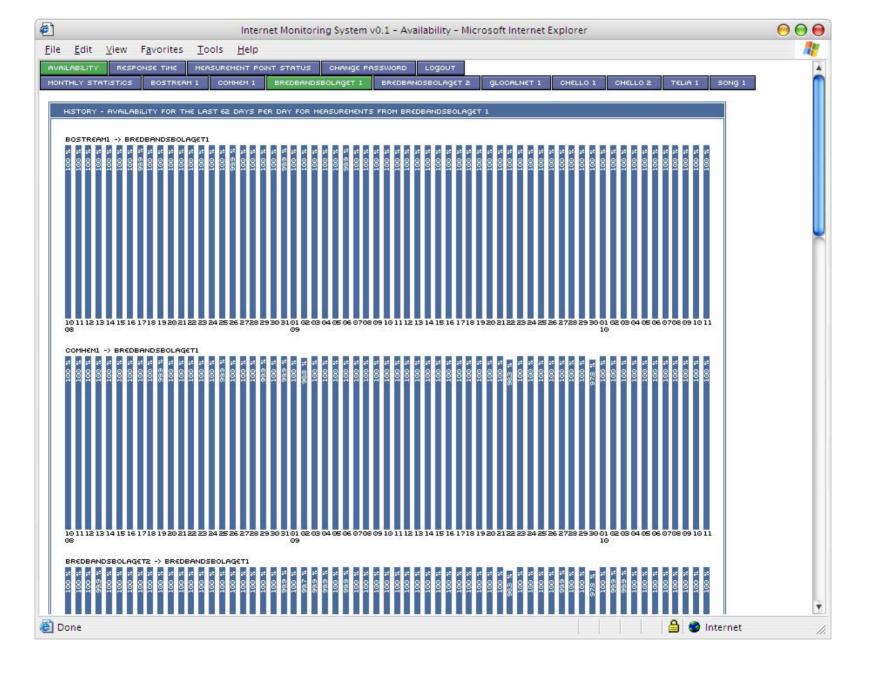
Bad RTT



(Check the scale! 700 ms)

No problems with

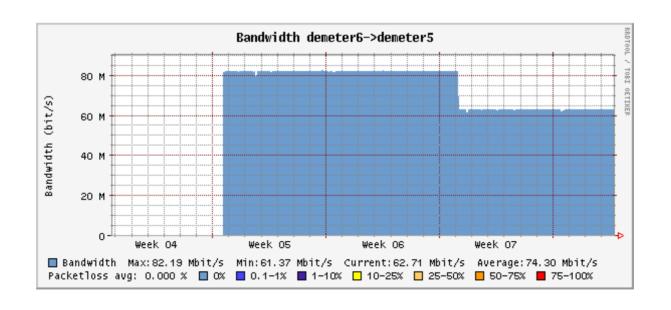
- Operators with fiber infrastructure (~100% availability)
- Measurement point hardware (Really!)
- Server
- Scripts
- Vacuum cleaners!! (REALLY!)

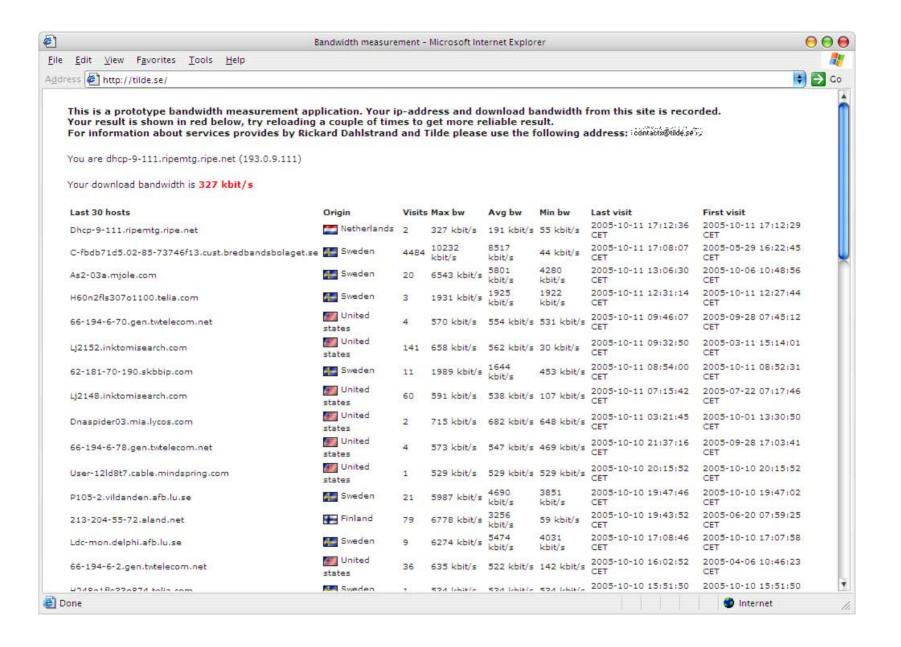


Future measurements

Bandwidth measurements

- REALLY not a good idea for connections in use.
- Netperf/iperf
- Http download bw





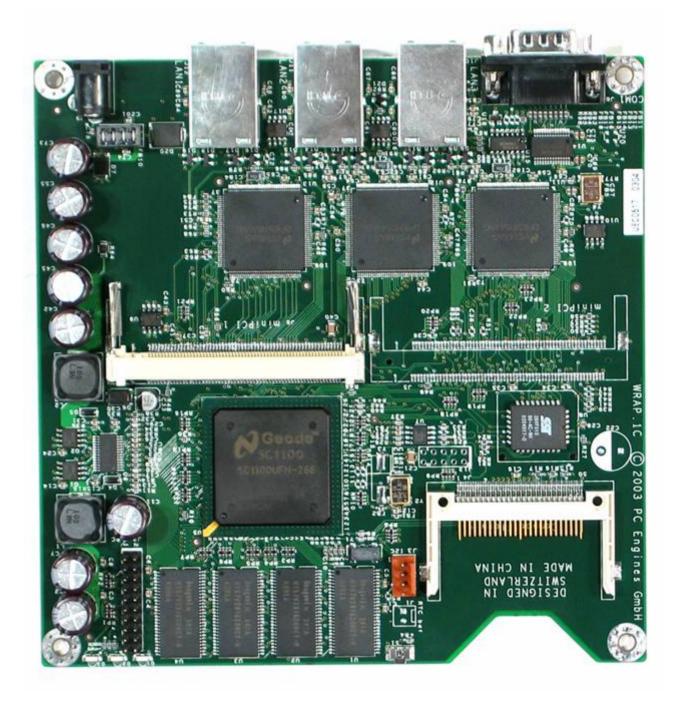
More crazy ideas

- RTT for other protocols? (UDP, TCP RTT)
- Hop-counts?
- Traceroutes? (Is this still a valid tool)

The neat stuff

Hardware

- PC Engines 6" WRAP system boards and boxes
- CPU: 266 MHz AMD Geode SC1100 CPU (x86)
- DRAM: 128 MB SDRAM
- Storage: 128 MB Compact Flash (Industrial grade are recommended!)
- OS: Any x86 OS are supported. (OpenBSD preferred)
- Power: 3-5W at 12 DC.
- LAN: 3 10/100 Ethernet port (National DP83816)
- Watchdog, LM77 temp monitor, serial console and 3 Leds you can flash and flicker!
- No keyboards, vga-ports, usb or power buttons!





Ramdisk-kernels

- Primary used for cd and floppy-based OSinstallation, but perfect for flash-disk based systems!
- Limits the number of writes and reads to the flash which means longer lifespan and less corrupted flash.
- One kernel-file contains the kernel and a ramdisk image containing the full distribution.
- Means that an entire distribution can be upgraded at once. If you replace one file on flash you replace not only the kernel, but the entire installation.

OS

- Based on OpenBSD 3.7.
- No patches, just a custom built process.
- Follows releases (distribution and kernel)
- Uses PF as a firewall.
- DHCP and NAT on internal interfaces for use as a firewall.
- NTP for timesyncing
- Ram based round robin Syslog + centralized logging.
- OpenBSD rocks!

Remote Management using SSH

- It's secure and can be scripted!
- All measurement points respond with SSH on a nonstandard port
- Only public-key authentication allowed (a master key)
- Using ssh & scp you can basically do anything!
- Including OS-upgrades!

To check the temperature

```
[rd@demeterupdate]# execall "sysctl hw.sensors"
-----Probe1-----
hw.sensors.0=lmtemp0, TEMP, temp, 38.50 degC / 101.30 degF
-----Probe2-----
hw.sensors.0=lmtemp0, TEMP, temp, 37.00 degC / 98.60 degF
-----Probe3-----
hw.sensors.0=lmtemp0, TEMP, temp, 45.50 degC / 113.90 degF
-----Probe4-----
hw.sensors.0=lmtemp0, TEMP, temp, 37.00 degC / 98.60 degF
-----Probe5-----
hw.sensors.0=lmtemp0, TEMP, temp, 39.00 degC / 102.20 degF
-----Probe6-----
hw.sensors.0=lmtemp0, TEMP, temp, 44.00 degC / 111.20 degF
-----Probe7-----
hw.sensors.0=lmtemp0, TEMP, temp, 40.50 degC / 104.90 degF
-----Probe8-----
hw.sensors.0=lmtemp0, TEMP, temp, 39.50 degC / 103.10 degF
-----Probe9-----
hw.sensors.0=lmtemp0, TEMP, temp, 39.00 degC / 102.20 degF
[rd@demeterupdate]#
```

To verify NTP-functionality

```
[rd@demeterupdate]# execall date
-----Probe1-----
Tue Oct 11 14:09:54 CEST 2005
-----Probe2-----
Tue Oct 11 14:09:55 CEST 2005
-----Probe3-----
Tue Oct 11 14:09:56 CEST 2005
-----Probe4-----
Tue Oct 11 14:09:57 CEST 2005
-----Probe5-----
Tue Oct 11 14:09:59 CEST 2005
-----Probe6-----
Tue Oct 11 14:10:01 CEST 2005
-----Probe7-----
Tue Oct 11 14:10:02 CEST 2005
-----Probe8-----
Tue Oct 11 14:10:03 CEST 2005
-----Probe9-----
Tue Oct 11 14:10:04 CEST 2005
[rd@demeterupdate]#
```

Or to upgrade

```
[rd@demeterupdate]# upgradeall bsd37.kernel
-----Probe1-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe2----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe3-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe4-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe5-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe6-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe7-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe8-----
bsd moved to obsd, new bsd moved to flash, rebooted
-----Probe9-----
bsd moved to obsd, new bsd moved to flash, rebooted
[rd@demeterupdate]#
```

For those who don't like SSH

- A webbased management and troubleshooting interface accessible over https.
- Contains logs, ifconfig, netstat, firewall-config etc.
- Open on local ethernet ports and via Internet using a shell debugcommand:

```
[root@probe3 root]# openhttps
External management interface open from 193.0.9.111, now goto https://xxx.xxx.xxx/
[root@probe3 root]# closehttps
External management interface closed!
[root@probe3 root]#
```



Log

inet6 :: 1 prefixlen 128

address: 00:0d:b9:00:7a:98

sis0: flags=8843 mtu 1500

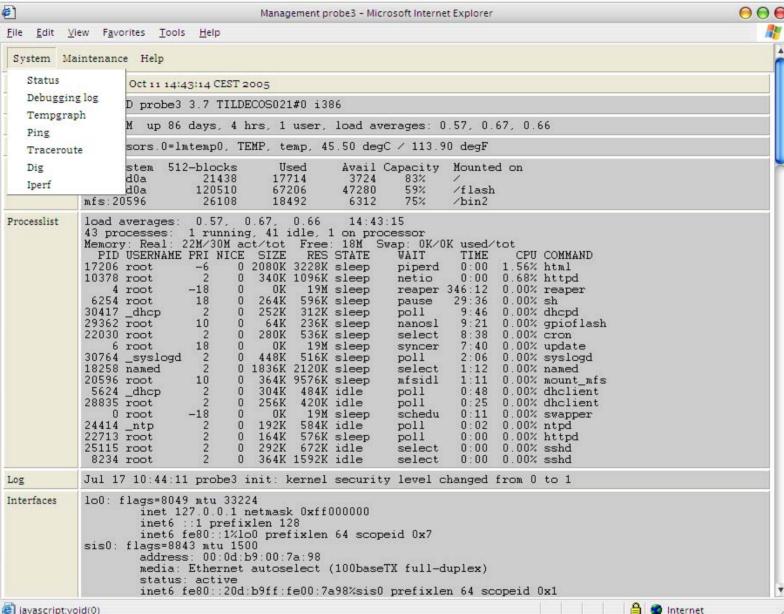
status: active

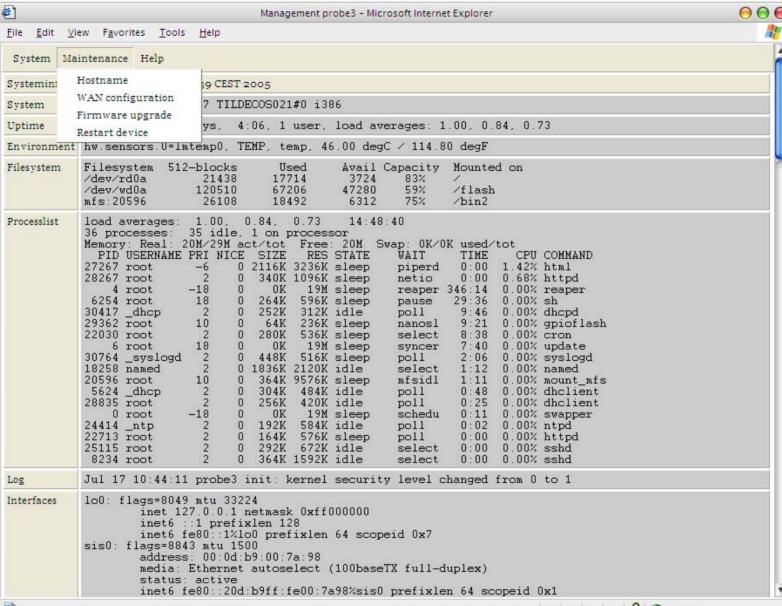
inet6 fe80::1%lo0 prefixlen 64 scopeid 0x7

media: Ethernet autoselect (100baseTX full-duplex)

inet6 fe80::20d:b9ff:fe00:7a98%sis0 prefixlen 64 scopeid 0x1

(







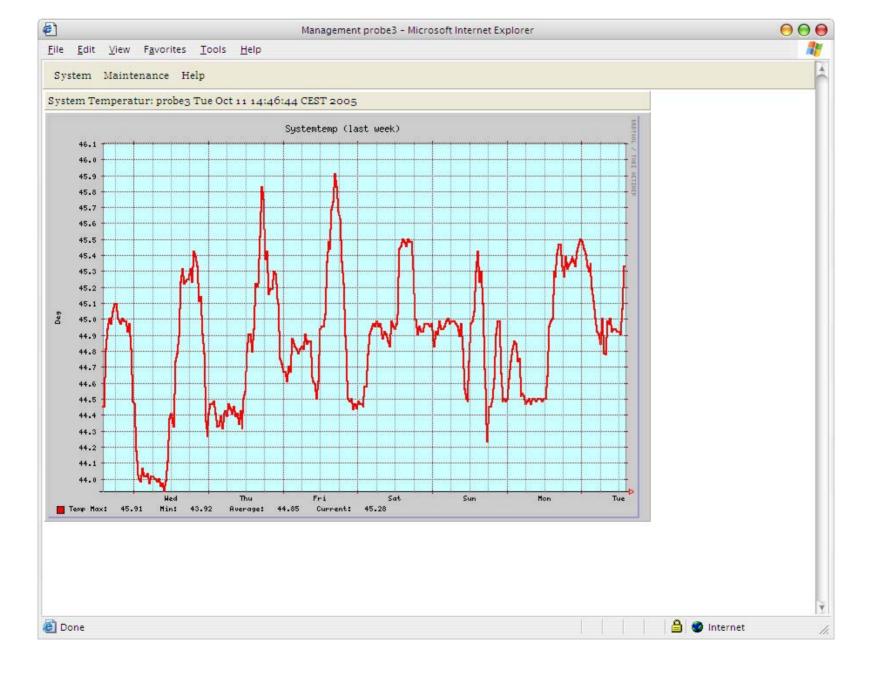


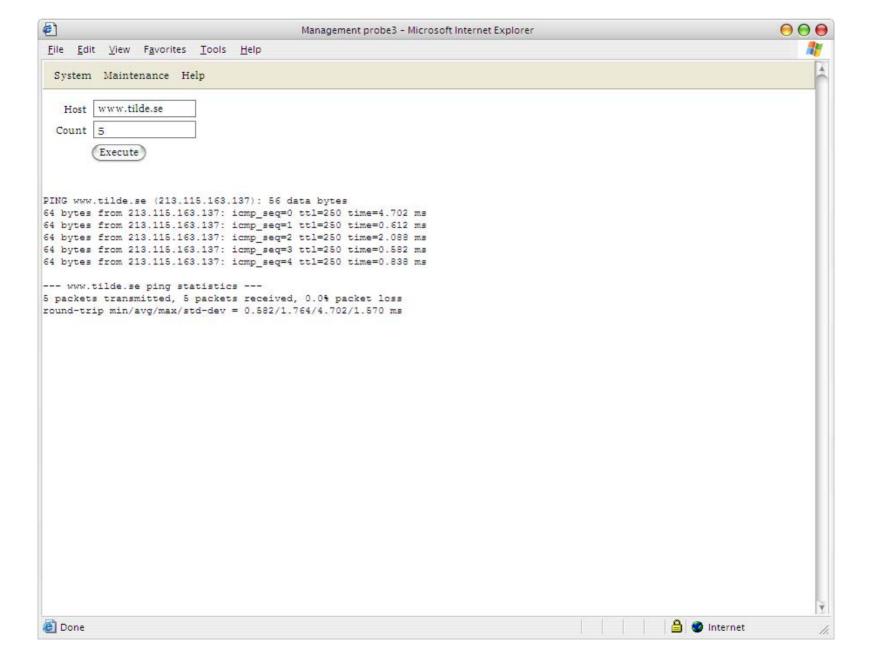


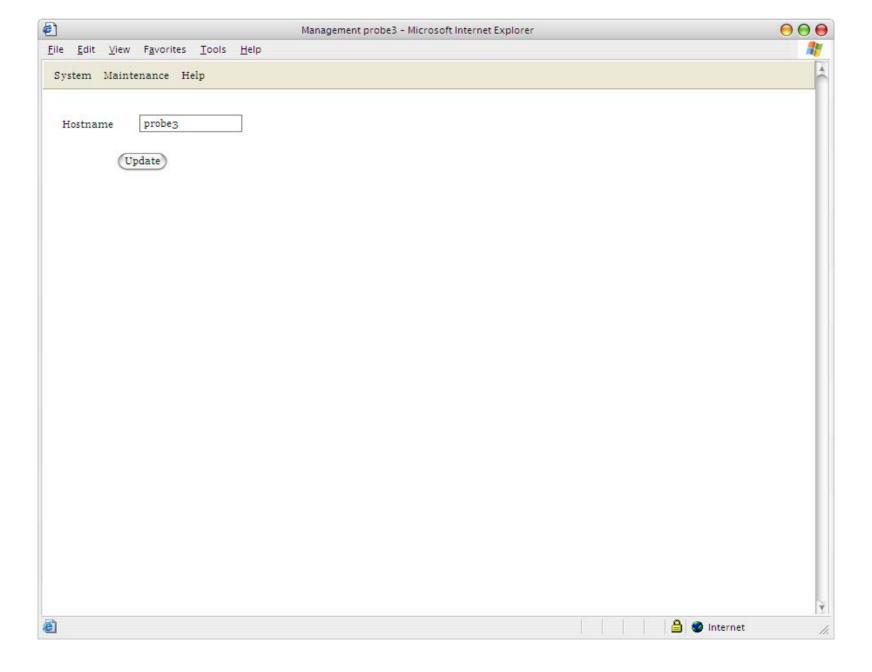
File Edit View Favorites Tools Help System Maintenance Help Debugging log: probe3 Tue Oct 11 14:46:12 CEST 2005 Oct 11 14:46:11 probe3 httpd[13773]: got request ``GET /longlog.html HTTP/1.1'' from host dhcp-9-111.ripemtg.r Oct 11 14:46:01 probe3 cron[9768]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:46:01 probe3 cron[28881]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:45:01 probe3 cron[5327]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:45:01 probe3 cron[9258]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:44:29 probe3 httpd[5157]: qot request 'GET /page.css HTTP/1.1' from host dhcp-9-111.ripemtq.ripe.n Oct 11 14:44:23 probe3 httpd[23035]: got request `GET /longlog.html HTTP/1.1'' from host dhcp-9-111.ripemtq.r Oct 11 14:44:01 probe3 cron[10048]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:44:01 probe3 cron[23931]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:43:27 probe3 httpd[6503]: got request ``GET /XulMenu.js HTTP/1.1'' from host dhop-9-111.ripemtg.ripe Oct 11 14:43:25 probe3 httpd[3296]: got request ``GET /ie5.js HTTP/1.1'' from host dhop-9-111.ripemtg.ripe.net Oct 11 14:43:20 probe3 httpd[3165]; got request 'GET /page.css HTTP/1.1' from host dhop-9-111.ripemtg.ripe.n Oct 11 14:43:16 probe3 crontab[30240]: (root) LIST (root) Oct 11 14:43:13 probe3 httpd[10378]: got request ``GET / HTTP/1.1'' from host dhop-9-111.ripemtg.ripe.net to p Oct 11 14:43:01 probe3 cron[27854]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:43:01 probe3 cron[14925]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:42:01 probe3 cron[537]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:42:01 probe3 cron[17816]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:41:55 probe3 httpd[4168]: got request `GET /XulMenu.js HTTP/1.1'' from host dhcp-9-111.ripemtg.ripe Oct 11 14:41:53 probe3 httpd[32733]: got request `GET /ie5.js HTTP/1.1'' from host dhcp-9-111.ripemtg.ripe.ne Oct 11 14:41:50 probe3 httpd[2249]; got request 'GET /page.css HTTP/1.1' from host dhcp-9-111.ripemtg.ripe.n Oct 11 14:41:43 probe3 crontab[2241]: (root) LIST (root) Oct 11 14:41:42 probe3 httpd[32439]: got request ``GET / HTTP/1.1'' from host dhcp-9-111.ripemtq.ripe.net to p Oct 11 14:41:01 probe3 cron[6817]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:41:01 probe3 cron[21979]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:40:51 probe3 httpd[11591]: got request ``GET /XulMenu.js HTTP/1.1'' from host dhcp-9-111.ripemtq.rip Oct 11 14:40:50 probe3 httpd[9039]: got request `GET /ie5.js HTTP/1.1' from host dhop-9-111.ripemtg.ripe.net Oct 11 14:40:48 probe3 httpd[15393]: got request `GET /page.css HTTP/1.1' from host dhcp-9-111 ripemtq ripe. Oct 11 14:40:46 probe3 crontab[1810]: (root) LIST (root) Oct 11 14:40:44 probe3 httpd[7559]: got request ``GET / HTTP/1.1'' from host dhop-9-111 ripemtg ripe net to po Oct 11 14:40:42 probe3 httpd[18487]: SSL ERROR OTHER Oct 11 14:40:31 probe3 httpd[6810]: got request ``GET /favicon.ico HTTP/1.1'' from host dhcp-9-111.ripemtg.rip Oct 11 14:40:24 probe3 httpd[22098]: SSL_ERROR OTHER Oct 11 14:40:01 probe3 cron[18231]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:40:01 probe3 cron[5086]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:39:01 probe3 cron[11343]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:39:01 probe3 cron[2229]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:38:01 probe3 cron[22669]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:38:01 probe3 cron[23272]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:37:01 probe3 cron[23689]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:37:01 probe3 cron[12473]: (root) CMD (/root/rrdcollectandgraph.sh >/dev/null 2>/dev/null) Oct 11 14:36:01 probe3 cron[3422]: (root) CMD (/root/demeterechoperflight.sh >/dev/null 2>&1) Oct 11 14:36:01 probe3 cron[1641]: (root) CMD (/root/rrdcollectandgraph sh >/dev/null 2>/dev/null)

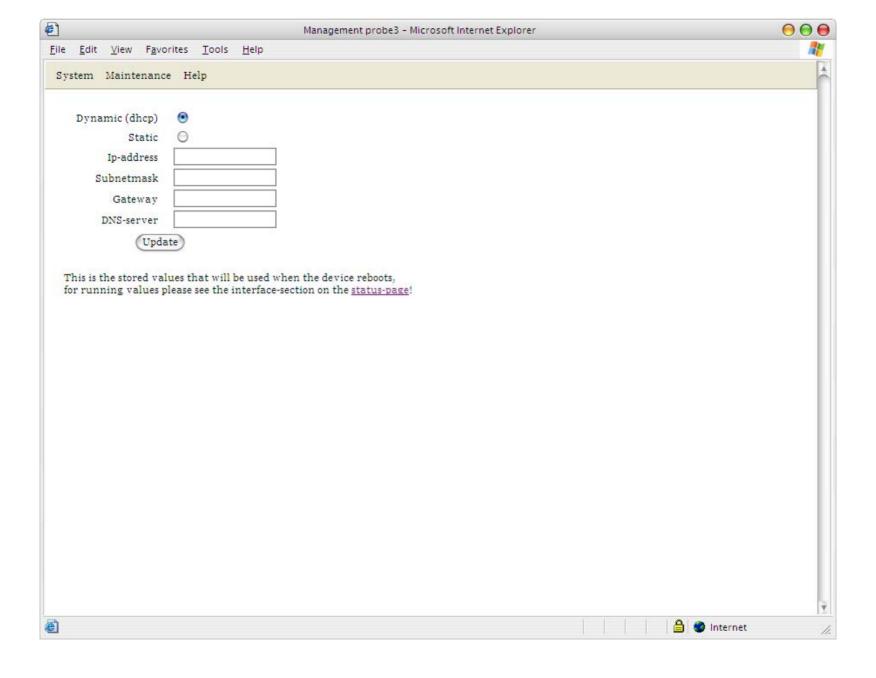












Conclusion

Is it any good?

- Well, not really. It doesn't really answer any questions clearly!
- And you can't really draw any conclusions what caused a problem and where the problem was located.
- 9 measurement points really doesn't represent the Swedish Internet very well.
- People in Sweden need resources from other parts of the world as well.
- However, it will never give you a false positive!
 If a connection is up, it's working!

Problems

- Multi-operator infrastructure. (Where network owners and service providers are separated)
- Not enough measurement points.
- Measurement are made from consumer connections.
- Measurements are only made from Swedish parts of the Internet.

Who needs this?

Consumers

- Is a problem with my operator or my connection?
- How is traffic working between my office and my homeconnection?
- Which operators/connection is the best?

PTS

- How is the Swedish part of Internet working?
- Operators (they just don't know it yet)
 - Check how well their peering and transit is actually working

Questions