The future of the RIPE NCC Hostcount

Henk Uijterwaal

RIPE NCC

DNS-WG, RIPE49, 22 September 2004

Introduction

- RIPE NCC hostcount:
    - Count the number of hosts on the Internet based on DNS
    - Long running project (1990), one of the longest time series around
    - Focus changed from absolute size to growth over time

- Has been running "as is" for a long time

- Time to revisit this activity

Why is this interesting?

- Absolute number:
  - Coverage of the Internet
- Trends
  - First derivative of absolute number or difference/time:
  - Growth of the net
  - New technology being picked up
  - Planning purposes

- Split out over regions or countries to see regional effects

Problems

- Note: everything said about the absolute number also applies to the trend

- Main problem:
  - Zone data transfers are blocked by more and more sites;
  - An unknown number of hosts is not counted

Problems

- No comparison with other data sources that we have
  - BGP, Allocation data, inverse tree, …
  - No estimate on the number of addresses used but not in the DNS
  - No consistent error estimates
  - Procedure for retroactively fixing problems not statistically analyzed

Other questions one can ask

- Who is using the data?
  - 1,500 visits website/month, 300 go deeper
  - No further insight
- What applications use the data?
- If not, why not?
- Is the project getting the exposure we want?

Other approaches

- Allocated addresses by the RIR's
- BGP routing table
- DNS host count
- Inverse Tree
- Ping (skitter)

Overestimate

Underestimate

- No approach is accurate, they are all estimates
- Various other projects have used some of these techniques

Correlations

- These 5 data series are correlated
  - How, exactly?
- They are probably shifted in time:
  - Apply for address space
  - Announce block in BGP
  - Populate block
- Topic to be studied

How to proceed from here?

- Draft document available
  - http://www.amsterdamned.org/~henk/hostcount.pdf
  - Longer version of this talk

- Comments: dns-wg@ripe.net before 22/10

- Final plan and schedule expected 1/11

Implementation

- ## Phase 1
  - Investigate research questions
  - Prototype
  - Estimate effort to maintain this in the long run

- ## White paper for community discussion
  - What should we do

- ## Phase 2
  - Implementation of production version

Questions, Discussion