Internet Atlas:
A Geographical Database of the Internet

Ramakrishnan Durairajan, Subhadip Ghosh, Xin Tang
Paul Barford, and Brian Eriksson
Motivation

THE ARPA NETWORK

DEC 1969

4 NODES

FIGURE 6.2  Drawing of 4 Node Network  
(Courtesy of Alex McKenzie)
Objectives of our work

• Create and maintain a comprehensive catalog of the physical Internet
  – Geographic locations of nodes (buildings that house PoPs, IXPs etc.) and links (fiber conduits)

• Deploy portal for visualization and analysis

• Extend with relevant related data
  – Active probes, BGP updates, Twitter, weather, real-time probing capability, attack data, etc.

• Apply maps to problems of interest
  – Robustness, performance, security
Related work

• Many prior Internet mapping efforts
  – S. Gorman studies from early 2000’s
  – CAIDA
  – DIMES

• Commercial activities
  – TeleGeography
  – Renesys
  – Lumeta

• Internet Topology Zoo

rkrish@cs.wisc.edu
Compiling a physical repository

• Step #1: Identification
  – Utilize search to find maps of physical locations

• Step #2: Transcription
  – Multiple methods to automate data entry

• Step #3: Verification
  – Ensure that data reflects latest network maps

• Our hypothesis
  – Physical sites are limited in number and fixed in location
  – But the raw number is still large!
Challenges

• Accuracy
  – How accurate are the node locations?
  – How accurate are the link paths and connections?

• Completeness
  – How much of the physical Internet is in the catalog?

• Varying data formats
  – Requires varying approaches for processing

• Verification
  – Networks change, data entry errors due to manual annotations
Internet Atlas @ UW

• Effort began in September ’11
  – Capture everything from maps discovered by search
  – Use all relevant data sources (ISP maps, colocation, data centers, NTP, traceroute, etc.)

• Data extraction tools

• Comprehensive database
  – Developed using MySQL

• Alpha web portal – http://atlas.wail.wisc.edu
  – Includes ArcGIS for visualization and analysis
Current DB

- Number of networks: 320
- Number of tier 1 networks: 10 (all)
- Number of data centers: 2,179
- Number of NTP servers: 744
- Number of traceroute servers: 221
- Number and type of other nodes: IXP (358), DNS root (282)
- Total number of nodes: 13,734
- Number of unique locations of nodes: 7,932
- Maximum overlap at any one node: 90
- Total number of links: 13,228
Identifying relevant data

• Internet search reveals significant information
  – ISP’s and data center hosts routinely publish maps and locations of their infrastructure
  – Other elements such as NTP list precise locations

• Creating a corpus of search terms
  – Geography is important

• Timely representations require repetition
Example: Telstra world wide
Example: Sprint IP network (US)
Example: Regional fiber
Example: Metro fiber maps
Automating transcription

• Web pages contain Internet resource information in a variety of formats
  – Text, flash, images, Google maps-based, etc.
• Extract information and enter automatically into DB
  – Requires identification of relevant page
• Library of parsing scripts for various formats
• Sometimes manual annotation is necessary
Geo-coding node locations

• Physical locations of nodes from search
  – Lat/Lon
  – Street address
  – City

• All locations decomposed in DB to Lat/Lon
  – Google geocoder
  – http://maps.googleapis.com/maps/api/geocode/xml?address="+address+"&sensor=false
Geo-accurate link transcription

- Transcribing geographic information for links is much more challenging than for nodes
- Step #1: Copy images
  - Max zoom required for max accuracy
- Step #2: Image patching via feature matching
- Step #3: Link image extraction from base map
- Step #4: Geographic projection
  - Key step uses ArcGIS registration functionality
- Step #5: Link vectorization

rkrish@cs.wisc.edu
Structure in link maps
Image extraction
Geo-specific link encoding
Internet Atlas – Full View
Internet Atlas – Layers
Internet Atlas – Identify
Internet Atlas – Zoom
Internet Atlas – Search
Internet Atlas – Search
Internet Atlas – Hurricane Sandy
Next steps

• Continue to populate DB
  – Goal = 1K networks by May, ‘14
• Continue to enhance web portal
  – Expanded data (BGPmon)
  – Expanded analytic capability
• Verification with active measurements
• Focus on analysis for target applications
Thank you!

Acknowledgements
• Paul, Subhadip, Xin, Brian, Mike, Math
• And as usual, the mistakes are mine!

Questions?

Try Internet Atlas.
http://atlas.wail.wisc.edu

Email us for accounts:
pb@cs.wisc.edu
rkrish@cs.wisc.edu