DIBBS EI: Platform for Applied Network Data Analysis (PANDA)

What do we need to know about the Internet?

CAIDA

PANDA

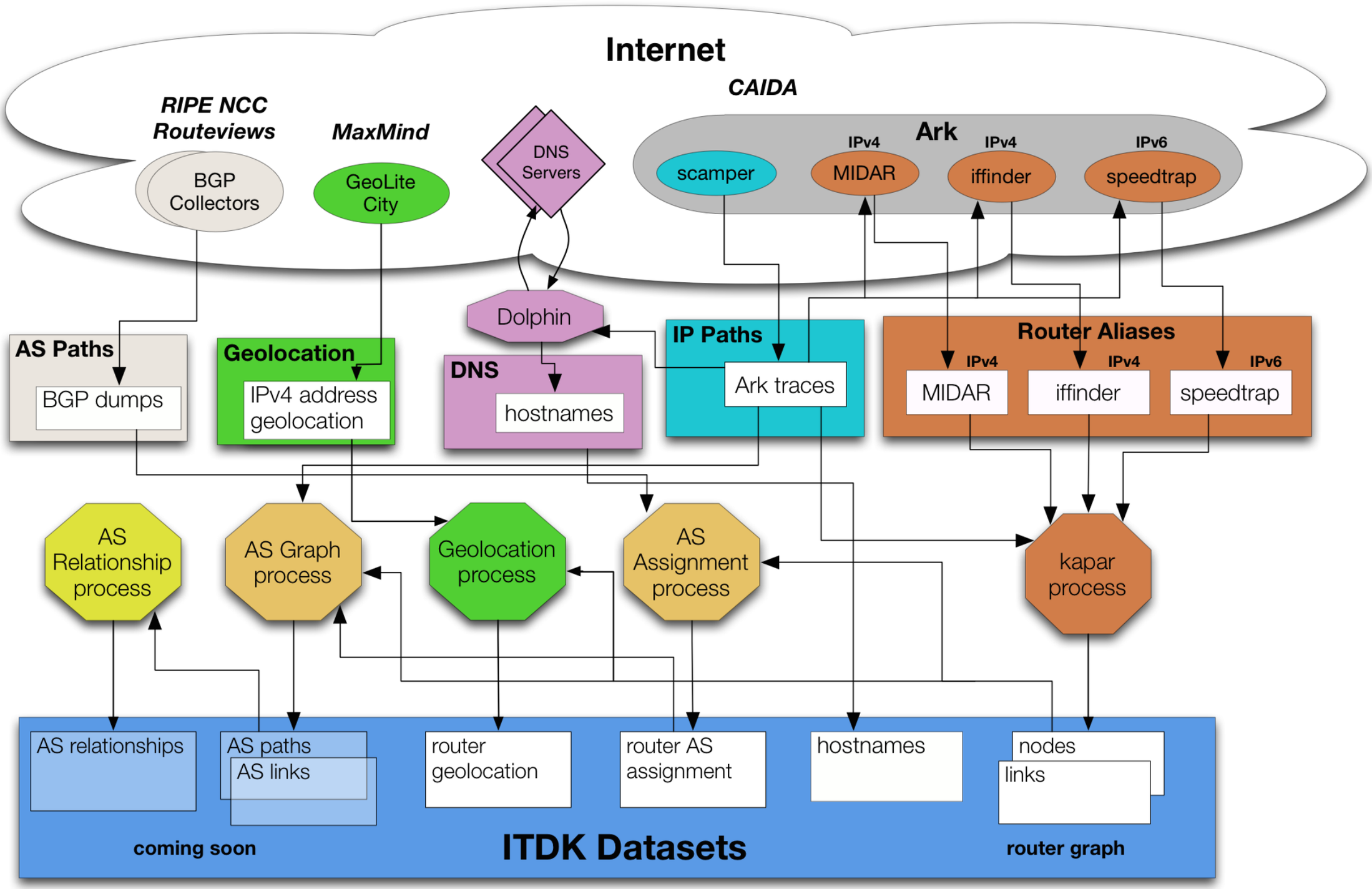
- Existing Data Building Blocks to Integrate
  - Archipelago Active Internet Measurement Platform and Supporting Components and Derivative Data (AS Relationships/Links/Types; ITDK)
  - ASRank: Comparison of routing and economic relationships among ISPs
  - BGPStream: efficient framework for routing (BGP) data analysis
  - Periscope: Extending measurement coverage by leveraging operational infrastructure
  - MANIC: Mapping and Analysis of Interdomain Congestion
  - Spoofer: Assessment of IP address validation best practices (Waikato)

ITDK: *Internet Topology Data Kit Process*

IP Paths | Alias Resolution | DNS Hostnames | Geolocations

BGP | AS Assignment | AS Relationship | ITDK Datasets

data collectors | data processes | data files



3 Tasks

- Task 1: PANDA: Platform for Applied Network Data Analysis
    - Software development to scale performance and functionality for community use
        - Create software modules to link components to each other and external software
        - Increase community accessibility of unified platform and underlying components
- Task 2: Support for and collaboration with multiple disciplines
- Task 3: Extensibility and adaptation to new opportunities

Overview

PANDA
Platform for Applied Network Data Analysis

existing ——— planned - - - - - -

Measurement Software and Hardware Infrastructures to Support Data Collection

Resulting Databases and Datasets

caida

CAIDA Platforms

Archipelago [Ark]
Collection of active monitors
- **Vela [Vela]** on-demand measurement
- **Henya [Hen]** topology query system

ASRank [AR]
ISPs' Autonomous System (AS) ranking and relationships

BGPStream [BS]
Collector/Database for processing BGP data

Spoofer [Spfr]
[Ark,User] run code to check for "spoofed" addreses

Periscope [Per]
Interface to public looking glass interfaces

Hosting Organizations

Letter Of Collaboration (LOC)
- **Waikato University [WaiU]**
- **RIPE NCC [RIPE]**
- **University of Twente [UTwe]**
- **University of Pisa [UPisa]**

DigitalElements [DE]
MAXMIND [Max]
Hurrican Electric [HE]
Federal Communication Commission [FCC]
Regional Internet Registry [RIR]
Route Views[RV]
Amazon [Amzn]
Cloudscene[CS]
PeeringDB[PDB]
Packet Clearing House [PCH]

Users [User]
Volunteers running measurement code

Operators [Op]
ISP Operators running measurement code

Security assessments
testing network vulnerability

- **[Ark] Spoofer** traces → **[Spfr] Spoofer** DB (detect false address filtering)
- **[User,WaiU] netstinky** (checks protocol compliance)
- **[User,UPisa] home traffic** (not yet, evaluation phase)

Topology measurement IP level
path and performance measurement: IP level

- **[Ark] servers** (traceroutes)
- **[Ark] MIDAR** (router aliases)
- **[RIPE] RIPE Atlas** (traceroute,ping)
- **[Op] Looking Glass Servers** (third party traceroute/ping)
- **[PDB] IX** DB (Internet eXchanges)
- **[CS] IX** DB (Internet eXchanges)
- **[PCH] IX** DB (Internet eXchanges)
- **[HE] IX** DB (Internet eXchanges)

→ **[Ark] Ark traceroutes** files (IP paths)
→ **[Ark] ITDK** files (router topology)
→ **[Hen]Henya** DB (10 years of traceroute data)
→ **[Vela] Vela** (IP paths)
→ **[Per] Periscope** DB (traceroute/ping/BGP)
→ **IX** DB (Internet eXchanges)

Topology measurement AS Level
routing measurement data : AS Level

- **[Ark] ISP-level traceroute** (IP paths to AS paths)
- **[RIPE,RV] BGP** data (AS's paths and prefixes)

→ **AS Relationships** files (ISP business types)
→ **Prefix2AS** files (AS's prefixes)
→ **AS Link Geo** files (inter-AS link with geolocation)
→ **AS Geolocation** files (location of ASes)
→ **Customer Cone** files (AS's customers)
→ **[AR] AS Rank** (AS info and ranking)

- **[RIR] WHOIS** data (Internet ID ownership)

→ **AS2Org** files (Organization's AS)
→ **[BS] BGPStream** DB (AS and prefix paths)

Meta-data to support analytics
geographic location of Internet resources

- **[Max] Maxmind Lite** (IP geolocation)
- **[DE] Netacuity** (IP geolocation)
- **DROP** (hostname geolocation)
- **[UTwe] OpenIntel** (DNS Database)

→ **DDec** (hostname geolocation)

Performance measurements
quality of experience assessments

- **[Ark] border mapping** (ISP border mapping)
- **[Ark] TSLP** (time-series latency probing)
- **[Ama] Mech Turk** (crowdsourcing QOE assissment)
- **[FCC] MBA** (latency/performance)

→ **inter-domain_links** DB (ISP border IPs)
→ **congestion** DB (ISP border delay)

Passive traffic analytics
measuring internet traffic

PANDA GATEWAY

Research Questions

- **Network mapping interdomain Internet topology**
  - tralXeroute: integrate IXP, BGP, IP traceroutes
  - detect MPLS/middle box behavior (Benoit/Yves)
  - identify root causes of routing events (Crovella/BU)
  - reactive measurement experiments using PEERING testbed (USC)
  - path prediction (USC,ETH)
  - geolocation of Internet infrastructure
  - grey market transfers (detecting anomalous changes in topology, DNS, and BGP data to infer address transfers.)

Research Questions

- **Security**
  - types of networks involved in attacks?
  - other paths through affected networks?
  - detect and mitigate route hijacking
  - mapping suspcious activity at IP-level to DNS names (Twente)
  - censorship
  - more sophisticated analysis of spoofer or other security data (correlating with region, type, and size, other security properties)
  - SSL certificate notary deployment
  - DNSSEC-readiness
  - content-based access control

Research Questions

- **Economics**
  - regional variation in modes of interconnection (public vs. private);
  - trends in firm growth (using routing table coverage as a proxy)
  - correlate firm size with interconnection congestion
  - relation of network infrastructure development & economic growth
  - integrate traffic monitoring: where is traffic going? is it all shifting to private interconnects 1-hop up?
  - trends in to which (replicated) sites are only reachable regionally

Research Questions

- **Policy**
  - policy implications of network and measurement results in formats accessible to economists and lawyers.
  - interconnection
    - show me all observed interconnection links with the highest average persistent congestion in the U.S. over the last month
    - label interdomain links with geolocation, facility, AS relationships
    - help policymakers interpret it
    - help develop better measurements (future FCCMBA program)