Shared Darknet Development

David A. J. Ripley
daripley@anml.iu.edu

Indiana University Advanced Network Management Laboratory
Who?

- Under the auspices of the REN-ISAC
  - [http://www.ren-isac.net](http://www.ren-isac.net)
- Policy guidance and co-ordination provided by SALSA-CSI2
  - [http://security.internet2.edu/csi2/](http://security.internet2.edu/csi2/)
- Logistic, development and technical support provided by the IU’s Advanced Network Management Laboratory.
  - [http://anml.iu.edu](http://anml.iu.edu)
- Data, requirements (and additional development) provided by the community
  - For certain values of “community.”
  - Participation in development potentially driven by requirements.
    (Scratching your own itch.)
What?

• A darknet reporting, querying and analysis system.
  – spanning multiple IP spaces
  – distributed across multiple continents, countries, institutions (Multiple sensors per institution.)
  – Specific to the .edu world for now(?)
    • No technical reason for that.
What Part 2 / Why?

- "Wide-Aperture" sensor
- Large, distributed IP space
  - More representative data.
  - Better statistics.
  - More difficult for bad people to avoid being seen.
- More potential for sharing
  - Sharing of infrastructure.
  - Sharing of aggregated information (reports)
  - Sharing of site-specific information.
How?

Four components:

1. Site-specific sensors
   - Optionally/preferably with data storage, if only short-term

2. Data collection and aggregation

3. Analysis and reporting

4. Direct data access and querying (maybe)
Site Sensors and Data Sources

• Sites have their own sensors
  – “Site” is open to interpretation
  – May have multiple sensors per site
  – Sensors can be anything that generates appropriate data
    • Netflow exports from routers
    • UMich Internet Motion Sensor
    • Firewall logs (pf, netfilter)
    • Anything else…
Data Collection and Aggregation

- Central collection & Analysis host(s)
  - Collect data from individual sites
  - **Batched** or real time
  - Pushed or **pulled**
  - Protocol agnostic
  - **Batched:**
    - ssh, http(s) ftp(!) - authenticated or not.
  - **Real-time**
    - GRE, ssh port forwarding.
Data Collection and Aggregation 2
Reporting

- Report definitions consist of scope, recipients, information to be generated
  - Modular architecture for reported information
- Automatic generation of reports
  - Based on REN-ISAC registry or similar information base.
- Manual report definition
  - Defined by individuals?
Example of report definition generation:
Registry ties people to netblocks, netblocks to institutions, people to institutions.
People for a given institution receive a standard set of reports for their netblocks.
- Per port counts, total traffic, etc.
Depending on information in registry, all kinds of reporting relationships can be defined.
Reporting 3

• Custom reports possible.
  – There are going to be data-sharing/administrative/technical relationships not defined by the registry.
  – Some may be temporally limited.
  – “Hey, can Bob see my stuff for the next two months?”
Reporting 4

• Report definition (auto or manually generated) is likely to contain:
  – Recipients
  – Netblocks of interest
  – List of reporting modules
    • e.g. top ten destination ports
    • top ten hosts by # of flows
    • % change in total traffic vs. last day/week/month.
  • (This is all pluggable.)
Querying?

• Possibility of access to raw data
  – Data sharing policies?
  – Authentication & access control
  – Masking/Anonymization (explicitly provided for already.)
The Awful Truth

- Difficult to implement, but not just for technical reasons.
- People are generally willing, but…
  - Everyone is busy
  - Hardware, operating system, data format, architecture, firewall, policies etc are all different.
  - Surely some sites have something in common?
    - (Not yet they don’t)
The Awful Truth

• Do we:
  – Send people hardware?
  – Try and gain access to their systems?
  – Persuade them to do more work?
  – Persuade them to change their infrastructure?
The Awful Truth Part 2

• How can we make participation desirable?
  – Even better, how can we make it compelling?

• It’s not enough for it to be easy
  – Although that’s a big part of the problem.
The Awful Truth Part 3

- Making it easy to join in isn’t enough.
- It has to be easy to keep participating.
  - Everything has to be reliable; maintenance is a drag.
    - Like I already said – everyone is busy.
The Awful Truth Never Ends

- It’s hard to make projects that rely on the participation of large numbers of people to deliver any value a compelling prospect for the early participants.
Questions? Comments? Want to Participate?

daripley@anml.iu.edu

This could be you!