PSC
PITTSBURGH SUPERCOMPUTING CENTER

Automated Discovery of Failing TCP Flows with XSight

Chris Rapier [rapier@psc.edu]
Bryan Learn [blearn@psc.edu]
Pittsburgh Supercomputing Center

Internet2 Tech Exchange September 27, 2016

PSC
PITTSBURGH SUPERCOMPUTING CENTER

The Problem

- Identifying failing flows as they are happening is difficult.
  - We rely on
    - User feedback
    - Aggregate throughput
    - Synthetic benchmarks
    - Logs, &c
  - Generally speaking, single flow information is either lost in the noise or comes too late to be of high value.
- Finding a failing flow while it is happening leads to faster problem resolution and improved workflow.

A Solution: XSight

- XSight is a cross domain solution that gathers and analyzes individual flow data.
- Designed to instrument DTNs, collect and collate data, and provide actionable analysis in near real time.
- Works across domain boundaries in order to create a holistic view of TCP flow health.

The Building Blocks

- Listener Agent
- Data Store
- Analysis Agent
- User Interface

The Listener Agent

- Lightweight Web10g based client.
  - Small memory and CPU footprint under normal conditions.
- Monitors all TCP flows.
  - No need to instrument individual applications
- Can filter out flows by application, ip, or network.
  - Helps maintain confidentiality and ensures only relevant data is collected.
- Can send different data to multiple stores.
  - Internal flow data can be sent to a local store while R&E data is sent to central collector.

The Database

- Time series database using InfluxDB
- Stores
  - Metadata (application, src/dst, etc)
  - Path information
  - Metric data
- Data is tagged with origination information.
  - Allows NOC (c.f. GRNOC) to see all data.
  - Participating organizations can only see their data.
    - Useful for reporting, local analysis, etc.

Analysis Agent

Collecting data is useless unless you do something with it.

- Periodically reviews new and ongoing flows and subjects data to heuristic analysis.
  - Failing flows are tagged in the database
  - Tag corresponds to which test the data failed
- Still a work in progress.
  - Currently tests for duplicate ACKs and timeouts.
  - Will soon incorporate macroscopic model, TCP options, excessive jitter, etc.
  - Working with a collaborator on automated causal analysis using machine learning.

User Interface

- Dashboard
  - Navigate through flow data using discovery graph
  - Failing flows flagged by analysis agent
- Data View
  - Lists all flagged flows
  - Isolate individual flows and extract detailed metrics

Status

- XSight is the result of a 12 month NSF EAGER grant.
  - Much accomplished but much left to do
    - UI, Analysis agent, etc.
  - New grant currently pending
  - Looking for partner institutions and organizations

webpage coming to http://www.psc.edu/

email: rapier@psc.edu

Demo Time!