Team Lima: Terabyte Threat Analysis

Using Hadoop to detect attacks travelling over BT's network infrastructure

Aaron Kirkbride
Simon Hollingshead
Matthew Huxtable
Alex Marshall
Jan Polášek
Ernest Zeidman

The briefing

Simple.

- Take logfiles generated by each of the BT routers
- Convert them into something more readable
- Search for unusual activity
- Report it to the end user

The briefing

Simple.

- Take logfiles generated by each of the BT routers
- Convert them into something more readable
- Search for unusual activity
- Report it to the end user

Simple?

- A subset of an hour of data became a 1.3 GB CSV!

Hadoop

- From the Apache Foundation
- Splits up a large input and distributes it over multiple machines
- Handles nodes that go down before they finish their work
- Jobs written as Java classes to *Map* and *Reduce* the data

The jobs

Threats

- Denial of Service (abnormal traffic to a single destination)
- TCP/UDP/ICMP Flooding
- Land attack
- Fraggle attack
- Smurf attack

The jobs

Threats

- Denial of Service (abnormal traffic to a single destination)
- TCP/UDP/ICMP Flooding
- Land attack
- Fraggle attack
- Smurf attack

Statistics

- TCP/UDP/ICMP packet counts per router
- Number of active flows over the period
- Overall estimated data flow per router

The jobs

Threats

- Denial of Service (abnormal traffic to a single destination)
- TCP/UDP/ICMP Flooding
- Land attack
- Fraggle attack
- Smurf attack

Statistics

- TCP/UDP/ICMP packet counts per router
- Number of active flows over the period
- Overall estimated data flow per router

(plus helpers so BT can implement their own alongside)

HBase

The Hadoop Database

- Non-relational database modeled after Google BigTable
- Each row made of a *key* and a *value*
- Fast reads and writes, even on millions (or even *billions*) of records

Web UI

- Small Python-based server core leveraging Flask
- Uses standard HTML5, CSS and JavaScript
- Libraries like jQuery, Twitter Bootstrap, and RickshawJS
- Pushes data to client, no polling needed!
- Easy to integrate with BT's current systems

Web UI

- Small Python-based server core leveraging Flask
- Uses standard HTML5, CSS and JavaScript
- Libraries like jQuery, Twitter Bootstrap, and RickshawJS
- Pushes data to client, no polling needed!
- Easy to integrate with BT's current systems

- Graphed statistics for each router in the network
- List of all active threat alerts.
- Operator can mark as handled

Web UI

There would be an image here.

Web UI

And here.

Web UI

Also finally here.

Team Lima: Terabyte Threat Analysis

Using Hadoop to detect attacks travelling over BT's network infrastructure

Aaron Kirkbride
Simon Hollingshead
Matthew Huxtable
Alex Marshall
Jan Polášek
Ernest Zeidman