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 binary logfiles generated by each of the BT routers
- Convert them into something more easily readable
- Scan the logfiles for anything that we think is a ‘threat’
- Create a frontend users can check for alerts

The briefing

Simple.

- Take binary logfiles generated by each of the BT routers
- Convert them into something more easily readable
- Scan the logfiles for anything that we think is a 'threat'
- Create a frontend users can check for alerts

Simple?

- Logs only contain 0.1% of network traffic
- Log we were given was for midnight to 1AM, a slow period
- Log was only for a couple of routers from BT's network
- The file was still nine million lines and 1.3 GB!

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
- TCP/UDP/ICMP Flooding
- Land attack
- Fraggle attack
- Smurf attack

The jobs

Threats

- Denial of Service
- 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
- 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 classes to make new analysis jobs simpler to write)

HBase

The Hadoop Database

- Non-relational database modeled after Google BigTable
- Data concatenated together and stored as a *key* and a *value*
- Forget everything you knew about SQL queries!
- Fast reads and writes, even on millions (or *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

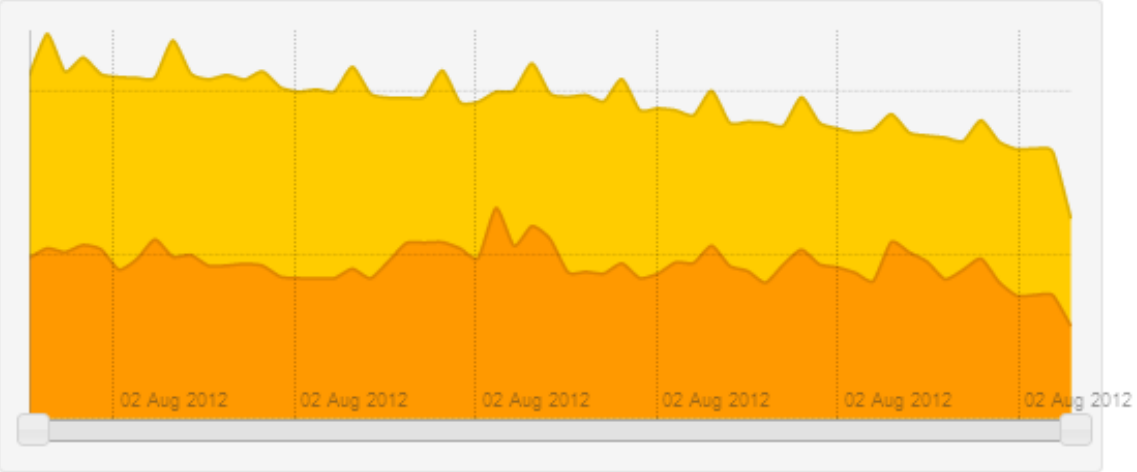
- Graphed statistics for each router in the network
- List of all active threat alerts
- Operator can mark as handled

Web UI (Dashboard)

Lima Web UI | Dashboard | Routers | Events

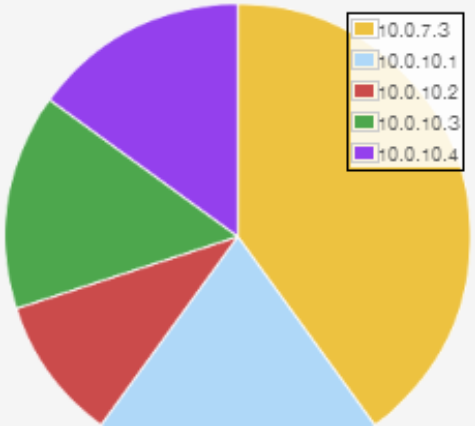
Latest Events

Router IP	Type	Time Submitted
10.0.7.3	portScan	26/2/2013 19:09:21
10.0.7.3	dos	26/2/2013 19:09:21
10.0.7.3	icmpFlooding	26/2/2013 19:09:21
10.0.7.3	pingPong	26/2/2013 19:09:21
10.0.7.3	fraggleAttack	26/2/2013 19:09:21



Jobs Running

Router IP	Last Seen	Progress
10.0.10.1	4/3/2013 02:06:52	
10.0.10.3	4/3/2013 02:06:52	
10.0.10.4	4/3/2013 02:06:52	



Status

Hadoop HDFS Name Nodes:	Online (2)
Hadoop HDFS Data Nodes:	Online (1/1)
Hadoop Map Reduce Managers:	Online (3)
HBase:	Online (Zookeeper up)
PostgreSQL:	Online
Thrift:	Online

Web UI (Routers)

Lima Web UI | Dashboard | Routers | Events

Router Information

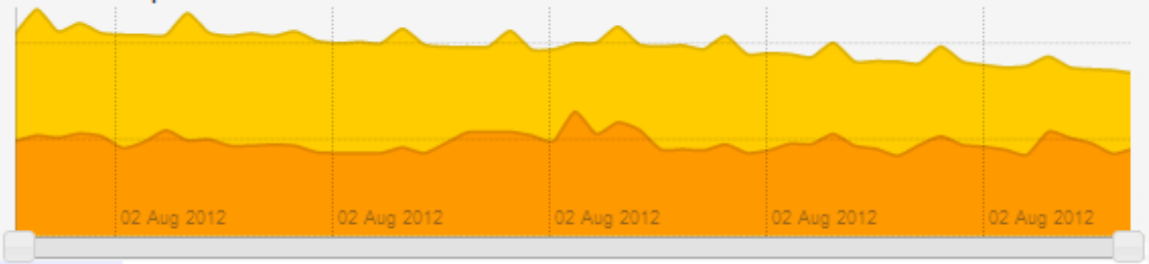
Total number of routers: 5
Last updated router: 10.0.7.3

Average flows per Hour: 9045410
Average packets per Hour: 2692945840
Average bytes per Hour: 1689231864532

Routers List

Search:

Router IP	Last Seen	Flows Per Hour	Packets Per Hour	Bytes Per Hour	JobTimestamp	JobStatus	JobMax
10.0.7.3	4/3/2013 02:06:52	9293160	2837463243	1766187564174	4/3/2013 02:06:52	0	0
Number of open events: 8							
10.0.10.1	4/3/2013 02:06:52	9483627	2457364323	1946253428546	4/3/2013 02:06:52	5	7
10.0.10.2	4/3/2013 02:06:52	8954375	2956473547	1574635243845	4/3/2013 02:06:52	0	0
Number of open events: 0							
10.0.10.3	4/3/2013 02:06:52	8293172	2305864567	1404346648355	4/3/2013 02:06:52	3	6
10.0.10.4	4/3/2013 02:06:52	9203720	2907563521	1754736253483	4/3/2013 02:06:52	2	10



Showing 1 to 5 of 5 entries

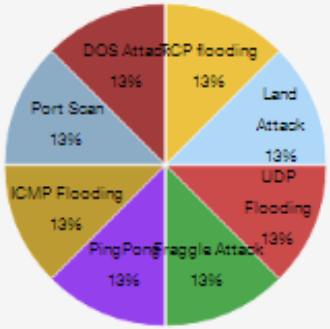
Web UI (Events)

Lima Web UI | Dashboard | Routers | Events

Alert! New attacks detected!

Event Information

Total number of active events: 8
Last threat: Port Scan - 10.0.7.3

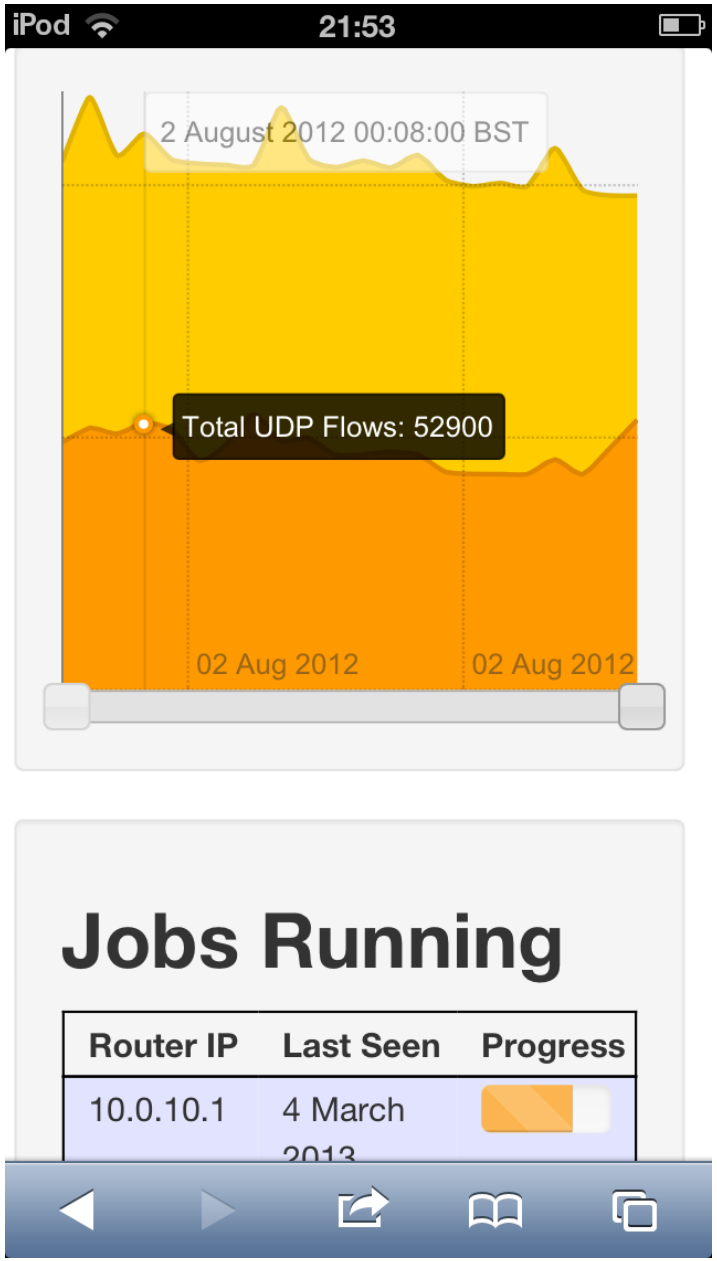


Events List

Search:

Event ID	Router IP	Type	Status	Message	Start Time	End Time	Time Submitted
12	10.0.7.3	tcpFlooding	open	tcpFlood1	2/8/2012 00:05:00	2/8/2012 00:08:20	26/2/2013 19:09:21
Destination IP: 23.43.21.43 **Source IP:** 1.53.2.12 **Flow Count:** 892745012 **Average Flow Data:** 883230776 **Total Flow Data:** 3070464000 **Total Packet Count:** 75241648							
13	10.0.7.3	landAttack	open	land1	2/8/2012 00:08:20	2/8/2012 00:11:40	26/2/2013 19:09:21
14	10.0.7.3	udpFlooding	open	udp1	2/8/2012 00:09:10	2/8/2012 00:11:40	26/2/2013 19:09:21
15	10.0.7.3	fraggleAttack	open	fraggle1	2/8/2012 00:33:20	2/8/2012 00:50:00	26/2/2013 19:09:21
16	10.0.7.3	pingPong	open	ping1	2/8/2012 00:50:00	2/8/2012 00:58:20	26/2/2013 19:09:21

Web UI (as seen on iOS)



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