Hashdoop: A MapReduce Framework for Network Anomaly Detection

Romain Fontugne, Johan Mazel, and Kensuke Fukuda

National Institute of Informatics Japanese-French Laboratory for Informatics

Problems

• Analysis of backbone traffic to prevent outages and maintain network resources available

• Detectors common approach:
- **Traffic discretization** (spatial/temporal aggregation)
- **Normal traffic modeling** (e.g. PCA)
- **Anomaly detection** (thresholding)

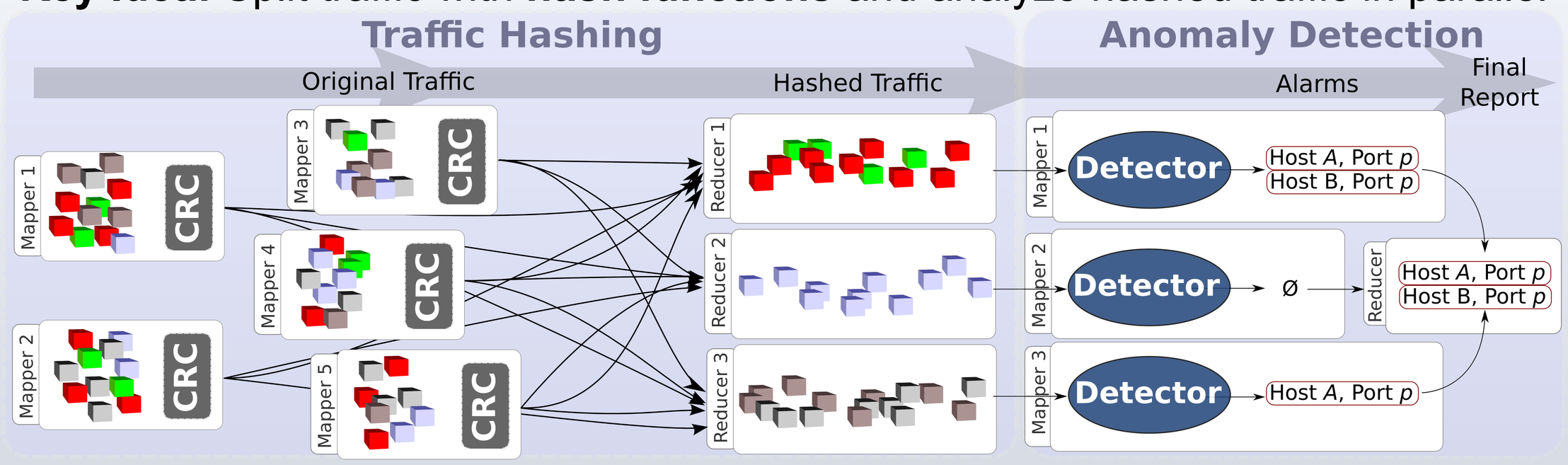
Motivation

•Cope with Internet traffic growth? Sampling?
⇒ Investigate **MapReduce** model

• Difficulties:
- MapReduce **splits the dataset**
- Detectors compute statistics from **spatial/temporal traffic structures**
- **Split traffic while preserving these structures?**

Proposal

Key Idea: Split traffic with **hash functions** and analyze hashed traffic in parallel

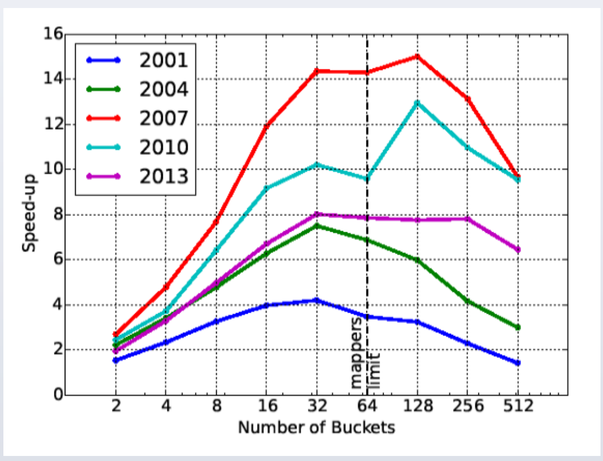


Evaluation

- **Dataset:** 15 traces from the MAWI archive collected in 2001, 2004, 2007, 2010, and 2013
- **Detectors:** Simple packet count based detector and Astute [Silveira et al. SIGCOMM'10]
- **Hadoop cluster:** 6 nodes, 128 mappers, 92 reducers, Hadoop 2.0.0

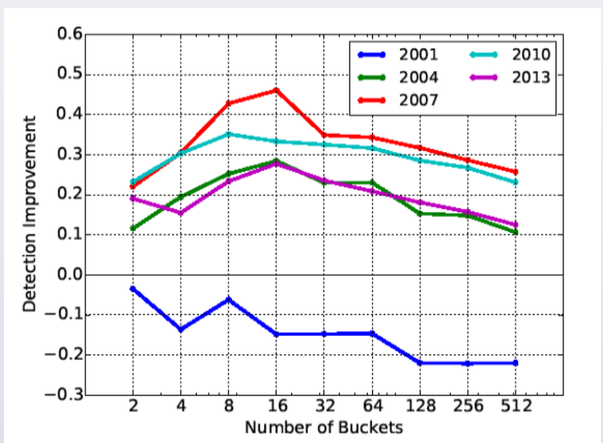
Processing Time:

Max. Speed-up x15



Detection Performance:

Improvements with large trace



⇒**Enable real-time detection:** Analysis of a trace in 2010 (900 sec. of traffic) takes **1296 sec.** on a single node but only **216 sec.** with Hashdoop!

Reference R. Fontugne et al. Hashdoop: A MapReduce framework for network anomaly detection INFOCOM Workshop, Big Security, 2014

Romain FONTUGNE,
romain@nii.ac.jp

NECOMA