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
Traffic Hashing

Anomaly Detection

Evaluation
• 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

⇒ 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