Internet Traffic Measurement Research @ DNLAB

Youngseok Lee

lee@cnu.ac.kr

cnu.lee@ucdavis.edu

Introduction

- Dept of Computer Engineering in Chungnam National University, Korea
  - http://cse.cnu.ac.kr
- Data Networks Lab (http://networks.cnu.ac.kr)
  - People: 1 Prof, 1 Phd student and 7 master students
  - Research interests
    - Internet measurement and big data analysis

Previous Research Projects

- Traffic engineering
  - "Traffic Engineering in Next-Generation Optical Networks”, IEEE Com munication Survey and Tutorial, 2004.
  - "Traffic Engineering-Aware Shortest-path Routing and its Application in IP-over-WDM Networks", OSA Journal of Optical Networking, 200 4.
- IP packet measurement with Hadoop
- IP geolocation database

Internet Measurement

- Challenges
  - Scalability, fault-tolerant system, extensibility
- CAIDA data
  - Capture, Curation, Storage, Search, Sharing, Analysis, and Visualization
    - Ark topology: 1.8 TB
    - Telescope: 102 TB
    - Packet headers: 18.8 TB

Josh Polterock, "CAIDA: A Data Sharing Case Study,"
Security at the Cyber Border: Exploring Cybersecurity for International Research Network Connections workshop, 2012

Harness Distributed Computing and Storage ?

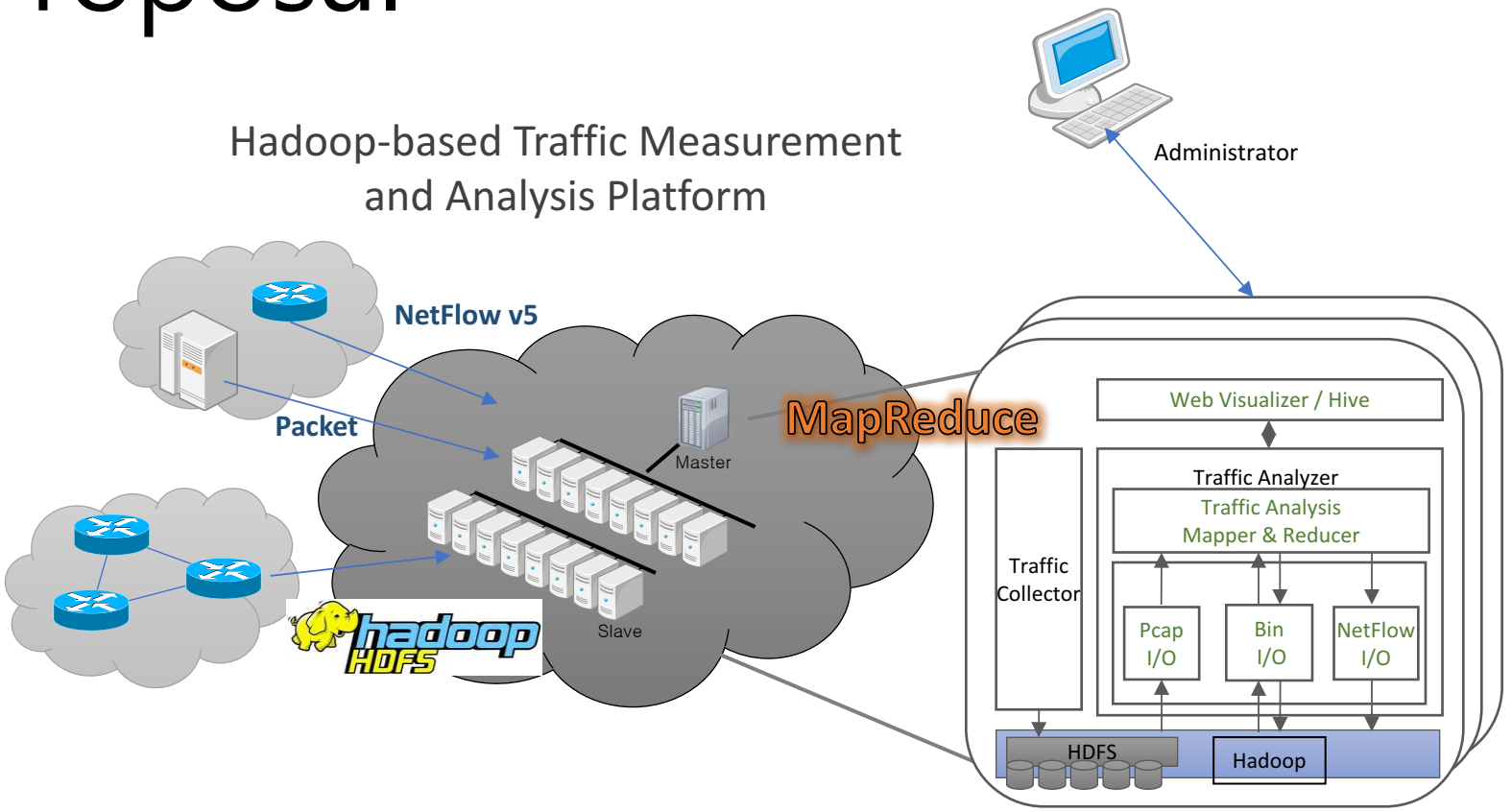
Google MapReduce, 2004

- 1 PB sorting by Google
  - 2008: 6 hours and 2 minutes on 4,000 computers
  - 2011: 33 minutes on 8000 computers
  - 2011: 10PB, 8000 computers, 6 hours and 27 minutes

Google

Apache Hadoop project

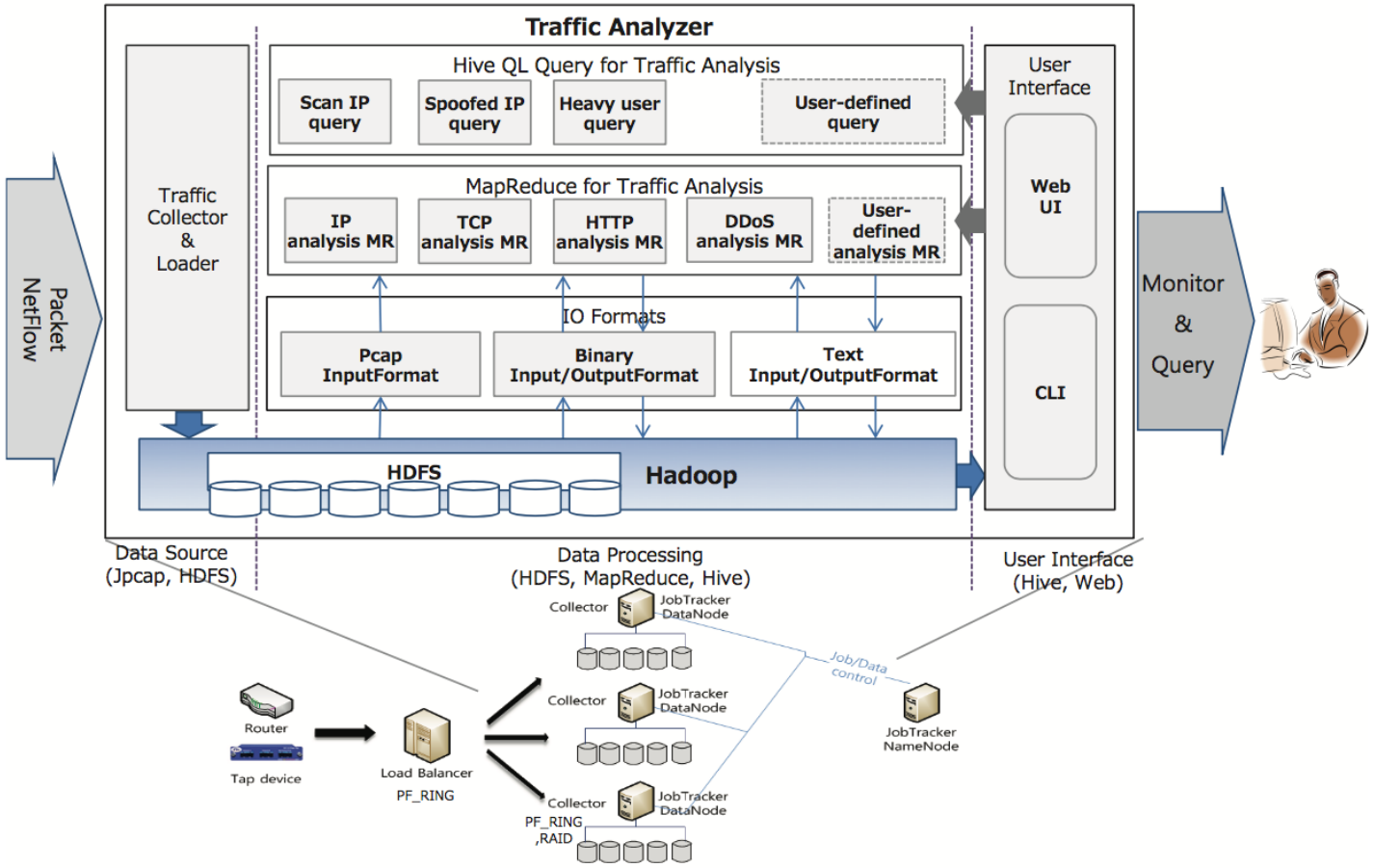
Our Proposal



1. *Yeonhee Lee and Youngseok Lee, "Towards Scalable Internet Traffic Measurement and Analysis with Hadoop," ACM SIGCOMM Computer Communication Review (CCR), Jan. 2013*
2. *Yeonhee Lee and Youngseok Lee "A Hadoop-based Packet Trace Processing Tool" , TMA, April 2011*
3. *Yeonhee Lee and Youngseok Lee, "Detecting DDoS Attacks with Hadoop", ACM CoNEXT Student Workshop, Dec, 2011*

Toward Scalable Internet Traffic Measurement and Analysis with Hadoop, ACM SIGCOMM CCR, Jan 2013

- To develop IP packet measurement system with Hadoop



Our Tool

	Traffic Analysis Job	Hadoop Tool Command	Description
IP Analysis	Total traffic and host/port count statistics	*PcapTotalStats –r[source dir/file] –n[reduces]*	Computing byte/packet/flowcounts regarding IPv4/v6/non-IP and the number of unique IP addresses/ports
	Periodic flow statistics	*PcapTotalFlowStats –r[source dir/file]*	Computing bytecount, packetcount per each interval, and periodic flow statistics regarding byte/packet/flowcounts
	Periodic simple traffic statistics	*PcapStats –r[source dir/file] –n[reduces]*	Computing periodic bytecount/packetcount regarding IPv4/v6/non-IP per interval
	Total count grouping by key	*PcapCountUp –r[source dir/file] –n[reduces]*	Computing total bytecount/packetcount by key (e.g., packetcount per each source IP address)
TCP Analysis	TCP statistics	*TcpStatRunner –**jt** –r[source dir/file] –n[reduces]*	Computing RTT, retransmission, out-of-order, and throughput per TCP connection
Application Analysis	Web usage pattern	*HttpStatRunner –**ju** -r[source dir/file] –n[reduces]*	Sorting Web URLs for user by timestamp
	Web popularity	*HttpStatRunner –**jw** -r[source dir/file] –n[reduces]*	Computing user count, view count for Web URL per Host
	DDoS analysis	*HttpStatRunner –**jd** -r[source dir/file] –n[reduces]*	Extracting attacked server and infected hosts
Flow Analysis	Flow concatenation and print	*FlowPrint [source dir/file]*	Aggregating multiple NetFlow files and converting flow records to human readable ones
	Aggregate flow statistics	*FlowStats [source dir/file]*	Computing total traffic of sIP/dIP/sPort/dPort/srcAS/dstAS/srcSubnet/dstSubnet per inbound/outbound
-	Top N	*TopN [source dir/file]*	Sorting records by key and emitting N numbers of record from the top.

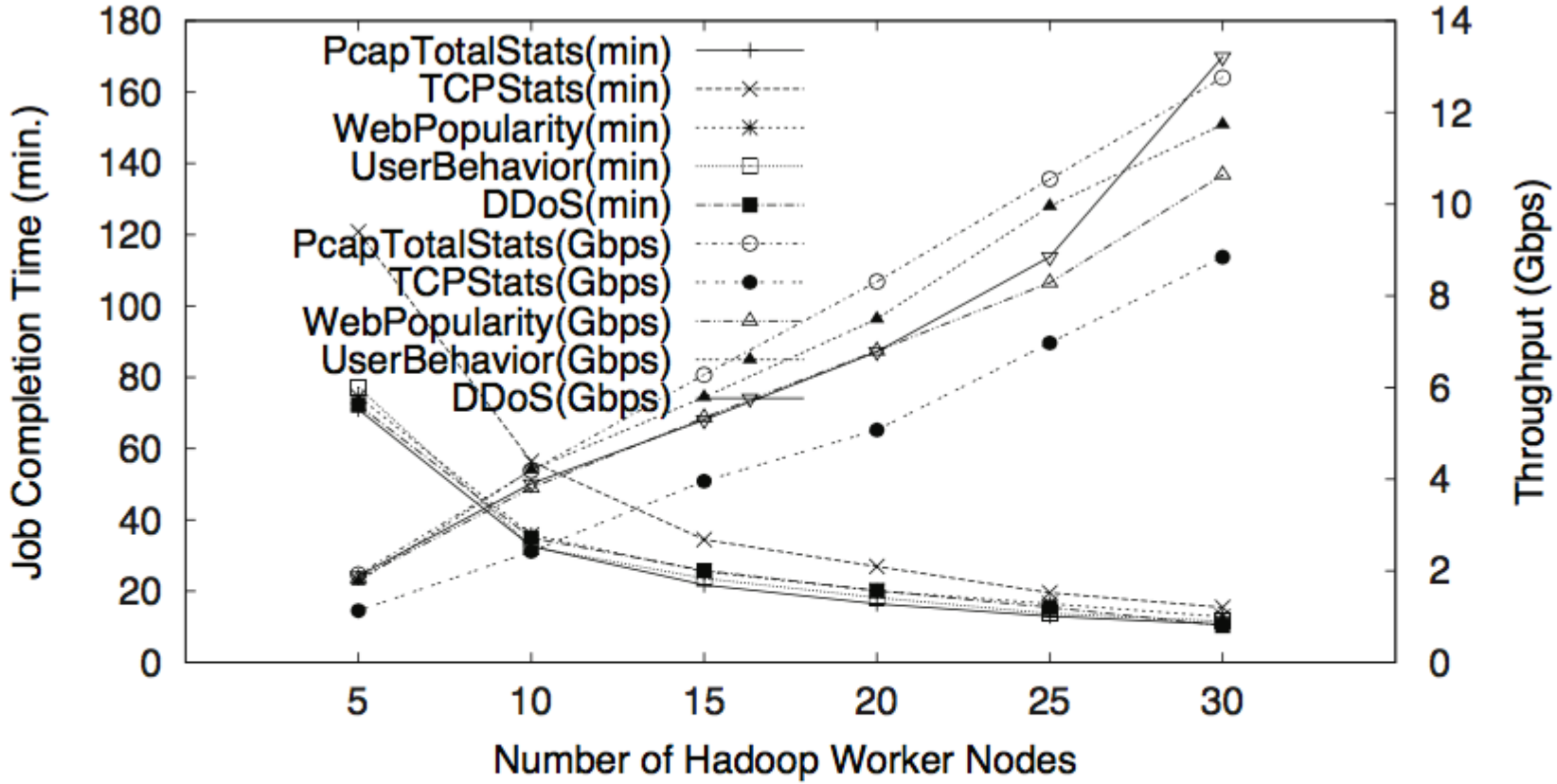


Figure 6: Completion time and throughput of traffic analysis MapReduce jobs on the 30-node Hadoop testbed for 1 TB `libpcap` files.

IP Geolocation with a Crowd-sourcing Broadband Performance Tool, ACM SIGCOMM CCR, Jan. 2016

- Build Korean IP geolocation DB with crowd-sourcing broadband performance data over 6 years

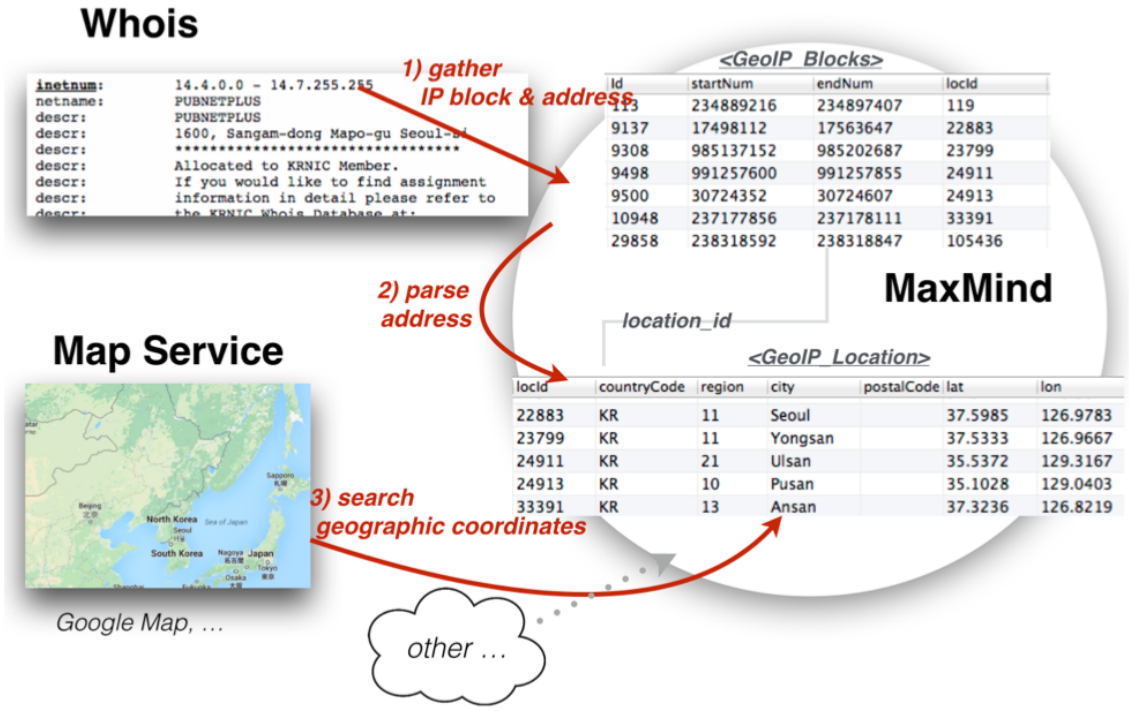


Figure 13: Method for building an IP geolocation DB in MaxMind.

How to build IP geolocation DB

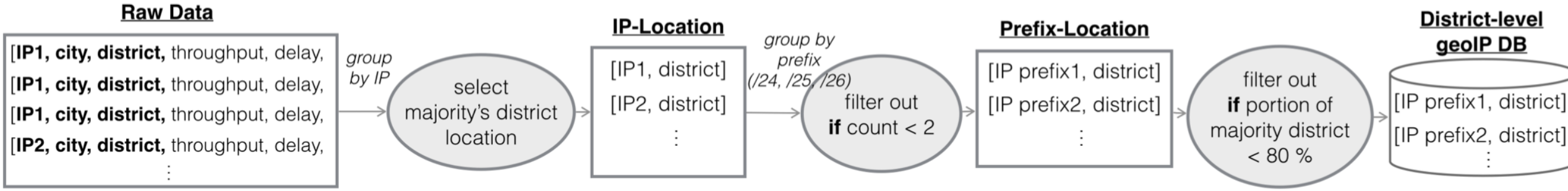


Figure 3: Process for extracting district-level IP geolocation from crowd-sourcing performance test data.

Table 2: The number of IP blocks and location entries of IP geolocation DBs in Korea.

DB	#of blks	IPs/blks	locations	coords
G_{2012}	619 K	64	233	231
MaxMind	24 K	7,387	1,543	313
IP2Loc	74 K	3,021	130	130

Accuracy

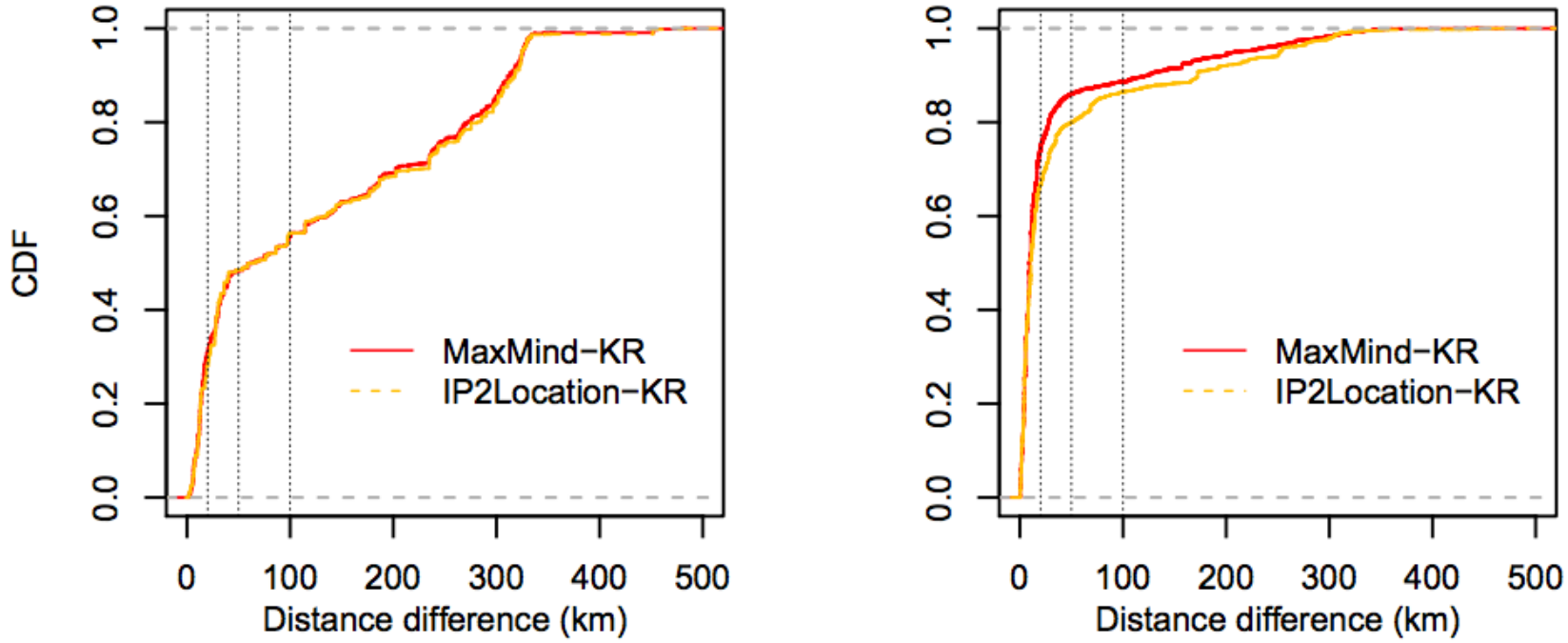


Figure 11: Distance difference of commercial geoIP DBs against G_{2012}_d26: capital vs. non-capital regions.

Ongoing Research Projects

- **Mobile app performance speed measurement platform**
- IP packet measurement with Android VPN
- ToR crawling
- Privacy examination in the web community

Mobile App Speed?

- Quality of Experience (QoE) for Mobile Apps

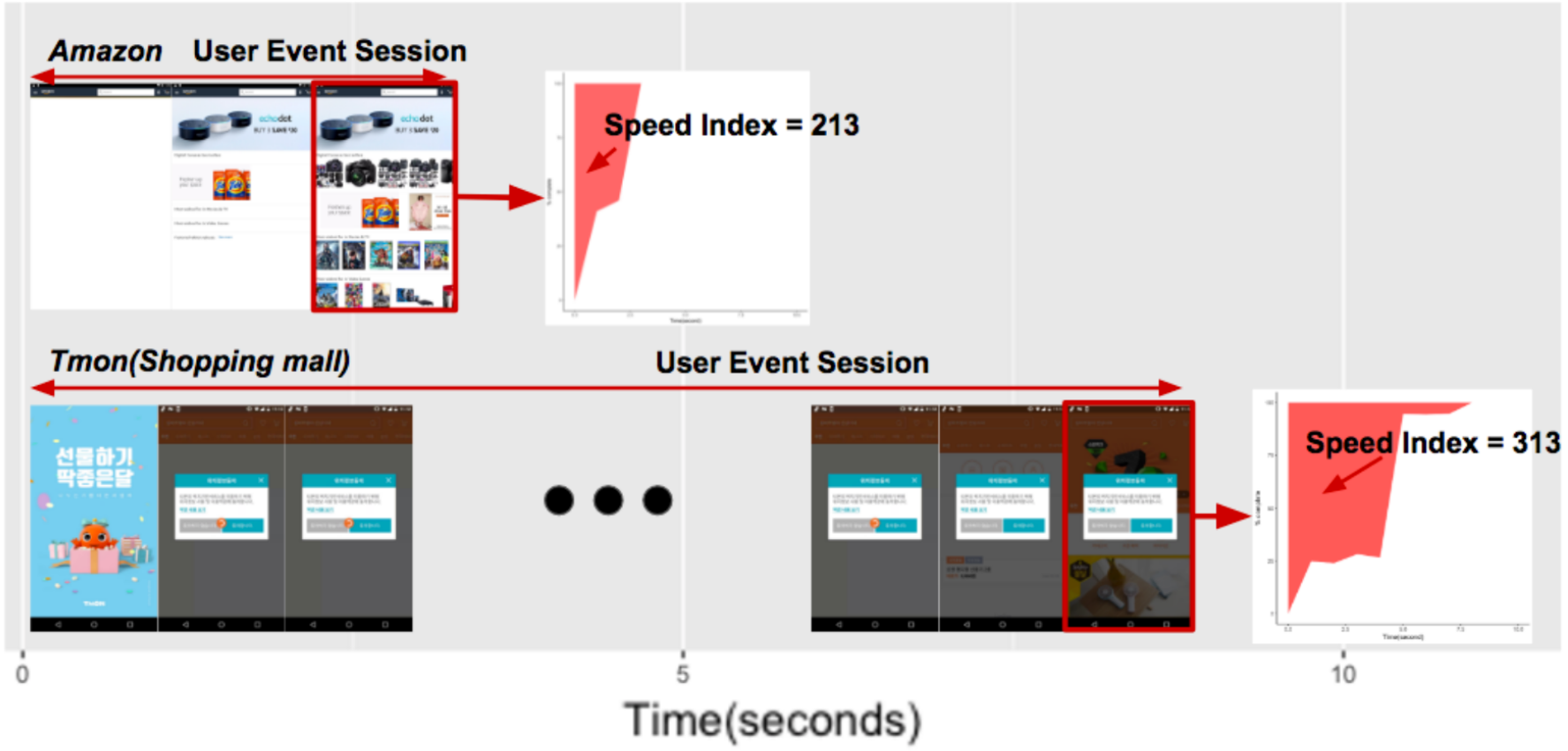


Figure 1: Speed Index of two mobile apps: Amazon vs. Tmon (Shopping mall) apps

App Speed Measurement System

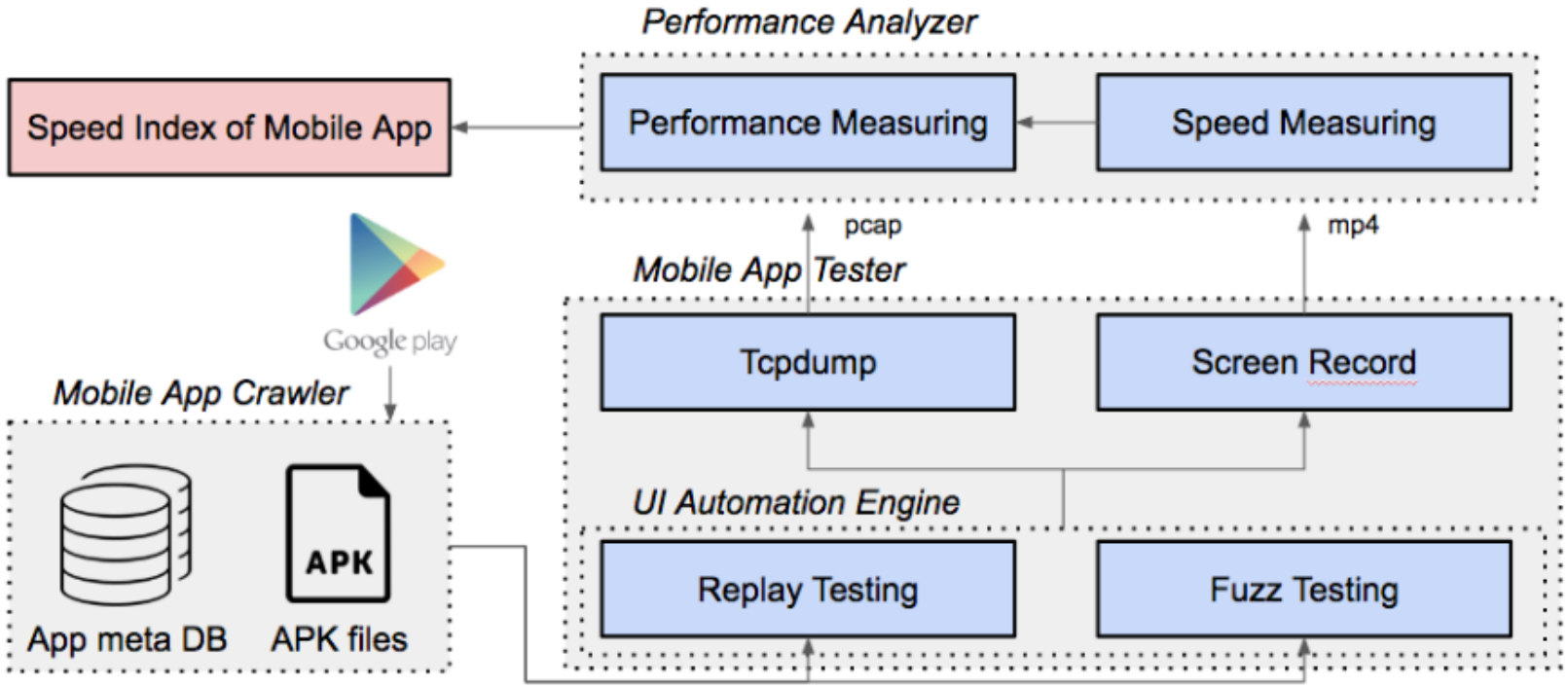
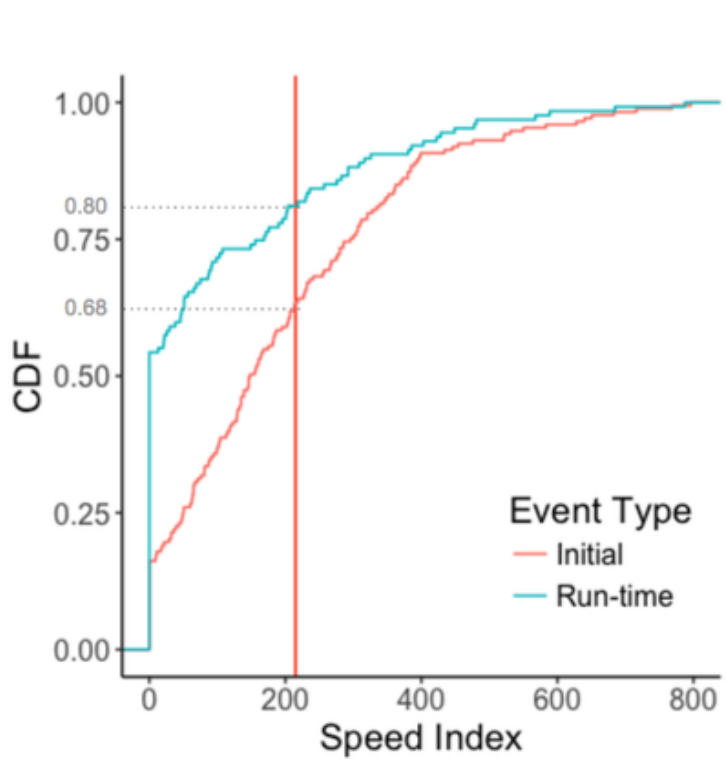
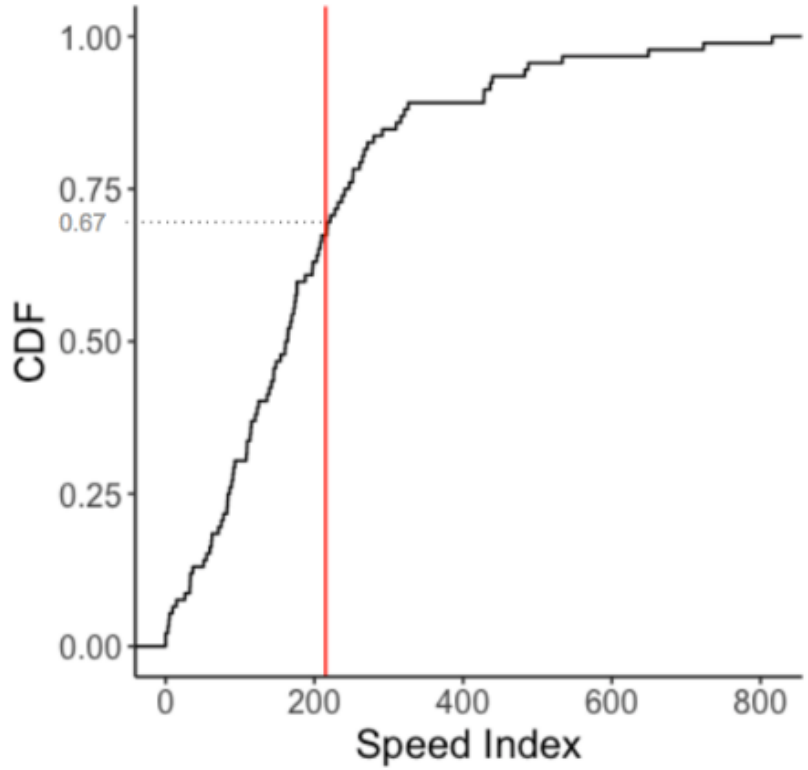


Figure 2: The mobile Speed Index measurement system architecture.

5.2 Case study of 100 sample apps

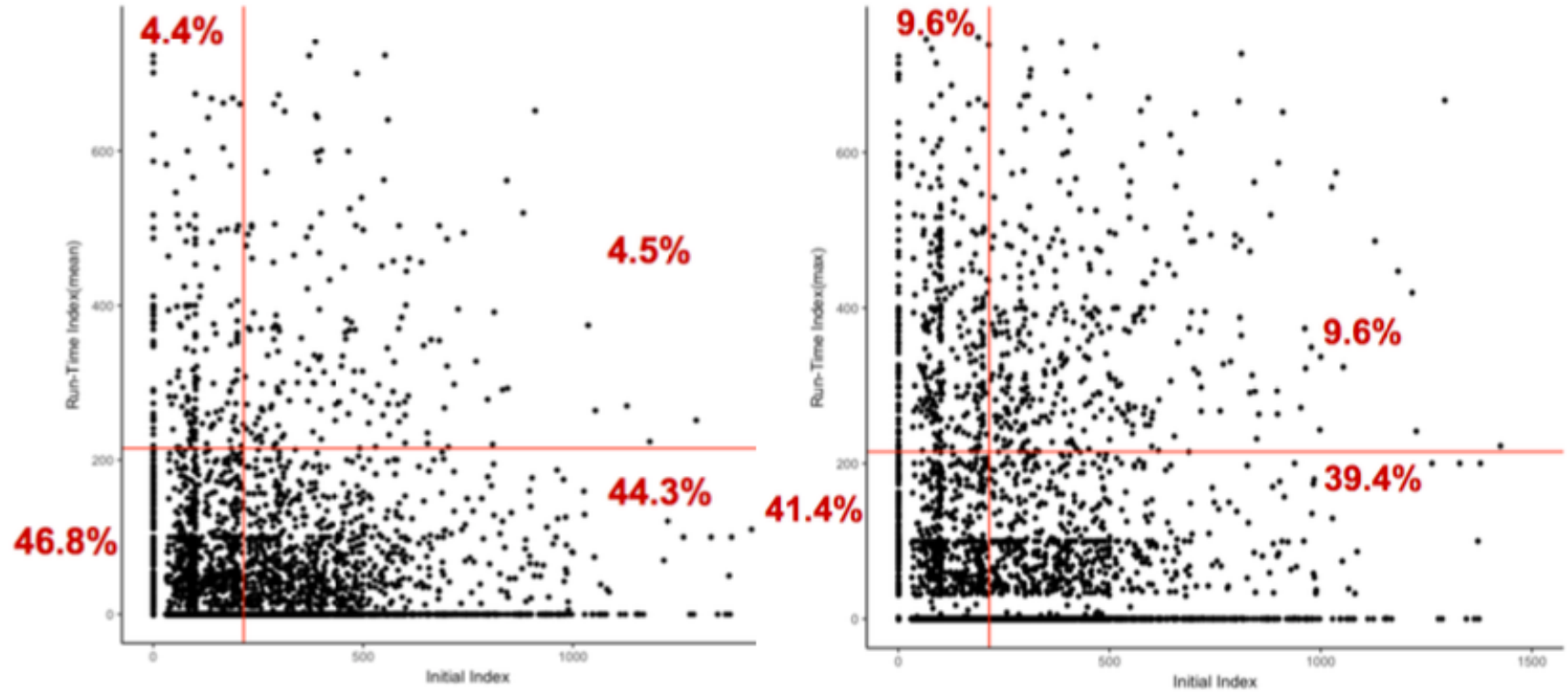


(a) Initial and run-time Speed Index of user events



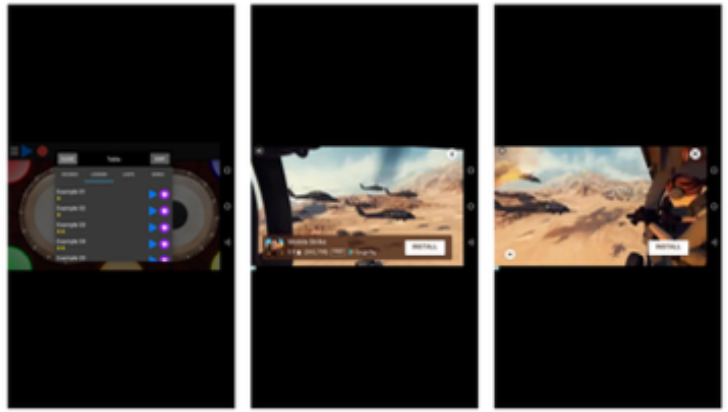
(b) Average Speed Index of mobile apps

5.3 Speed Index of 15,846 mobile apps

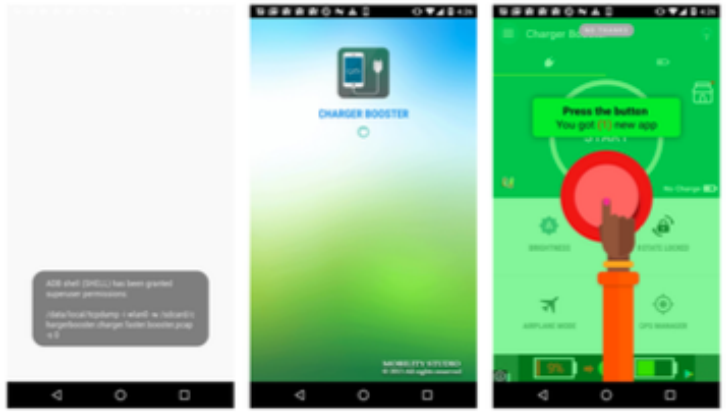


(a) Average Speed Index (b) Maximum Speed Index

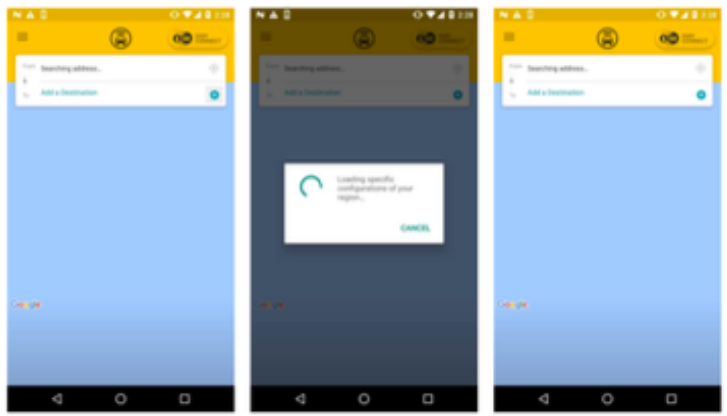
Figure 9: Initial and run-time Speed Index of 15,846 mobile apps.



(a) Advertisement



(b) Interrupting contents



(c) Popup

(d) Video contents

Figure 7: Example displays of mobile apps with the high mobile Speed Index in qualitative features.

Discussion

- Accuracy of mobile app speed index
- Usages of mobile app speed measurement
  - Know which factors make mobile apps slow/fast?
- Large-scale experiment with automated app files
  - login

Big Data Analysis of Computer Networks and Security

Date : August 4 2017 **Venue : Dept. of Computer Science, University of California, Davis**

CHUNGNAM NATIONAL UNIVERSITY FINTECH SECURITY RESEARCH CENTER

UCDAVIS NETWORKS RESEARCH LAB University of California, Davis