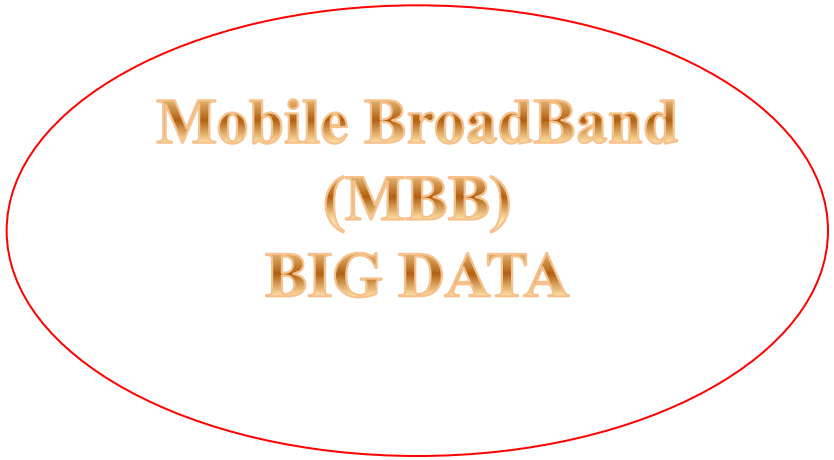
MBB Big Data Intelligence

HE, Cheng
Data Mining Group
Noah's Ark Laboratory
Central Research Institute
Huawei Technologies

Security Level:

Big Data

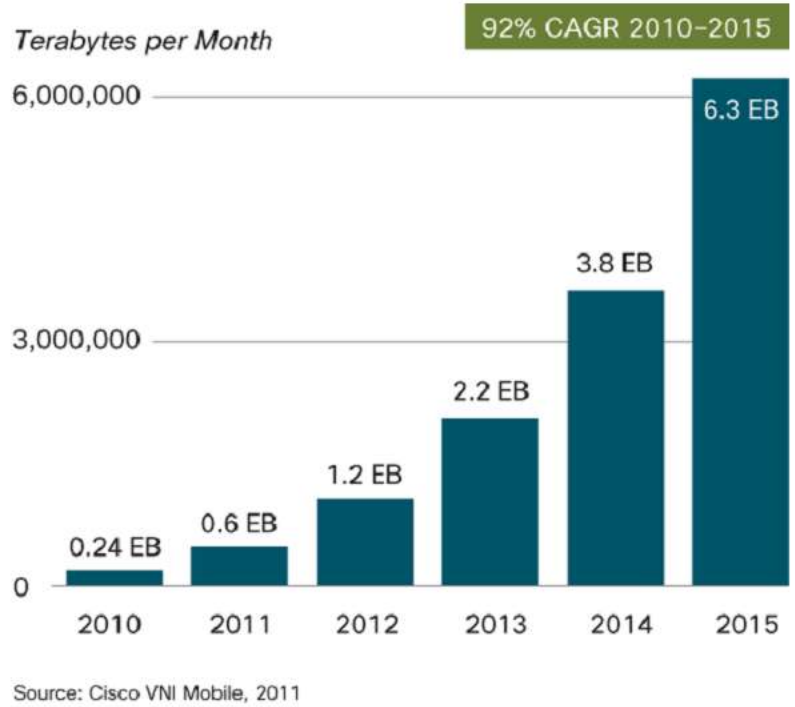
- **Definition:**
  - Big data are datasets that grow so large that they become awkward to work with using on-hand database management tools.
  - Difficulties include capture, storage, search, sharing, analytics and visualizing.
  - *Ref: http://en.wikipedia.org/wiki/Big_data*



Contents

MBB Big Data – 3V Features

Volume



Source: Cisco VNI Mobile, 2011

Velocity

`**Access rate:**
384Kb -> 1000Mb
` **Core Network rate:**
100Mb - > 100Gb

Mobile BroadBand (MBB) BIG DATA

Variety

ıousand apps. in app store;

425+ thousand apps. in android platform;

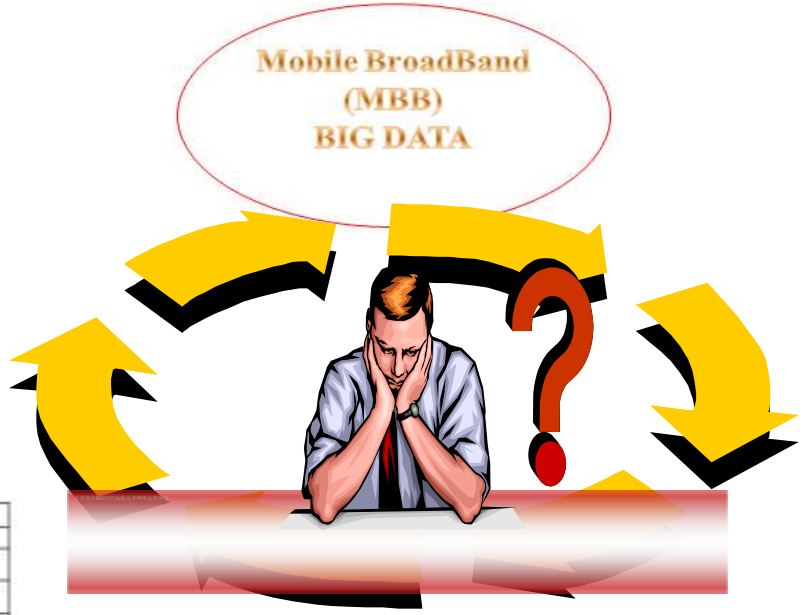
Sys. logs, Equipment logs, Signaling packets;

MBB Big Data - Challenges

Streaming Data Analysis

·Ultra-high Speed Pkt Flows

·On-line Processing;



Heterogeneous Data

·Data, Signals, Logs, Statistical Reports;

·Semi-structured / Unstructured;

Position	Name		Position	Name	
0	Payload length		22	Inner IP	Header length
1	UTC1		23		Total length
2	UTC2		24		id
3	Total Length				
4	Outer IP	Header length	25		Frag_flag
5		Total length	26		offset
6		id	27		Protocol type
7		Frag_flag	28		Source IP
8		offset	29		Destination IP
9		Protocol type	30	Inner TCP	Source Port
10		Source IP	31		Dest. Port
11		Destination IP	32		length
12	Outer TCP	Source Port	33		Flag
13		Destination Port	34	Inner UDP	Source Port
14		length	35		Dest. Port
15		Flag	36		length
16	Outer UDP	Source Port	37	*Payload*	
17		Destination Port	38	Signaling mark	
18		Total length	39	App Class	
19	GTPU	Type	40	App Software	
20		length	41	Protocol	
21		TEID	42	.....	

High / Multi Dimensional Data Analysis

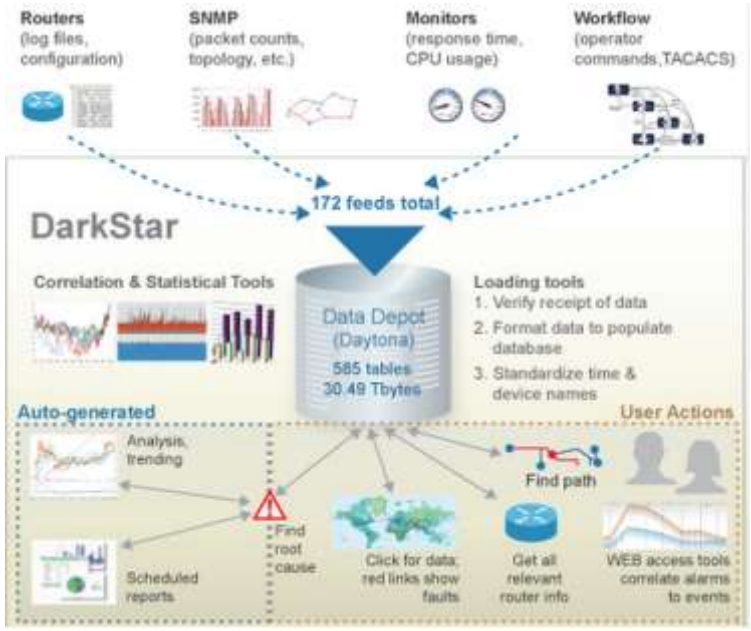
·From hundreds to 10s of thousands dimensions;

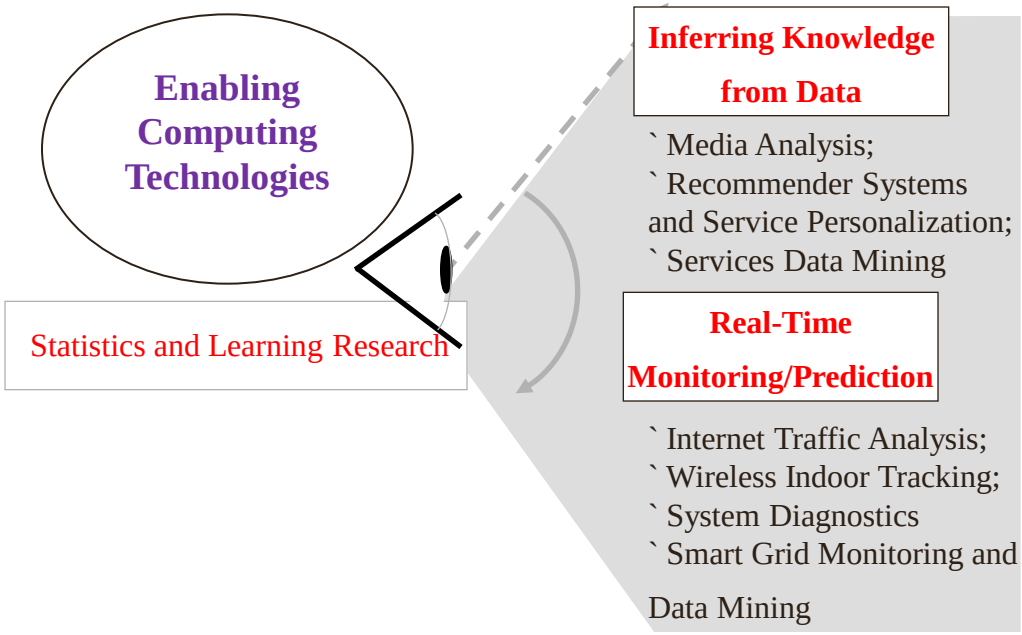
·Correlation Analysis;

State of the Art – Industrial Survey

- **AT&T Research**
  ·DarkStar Project;

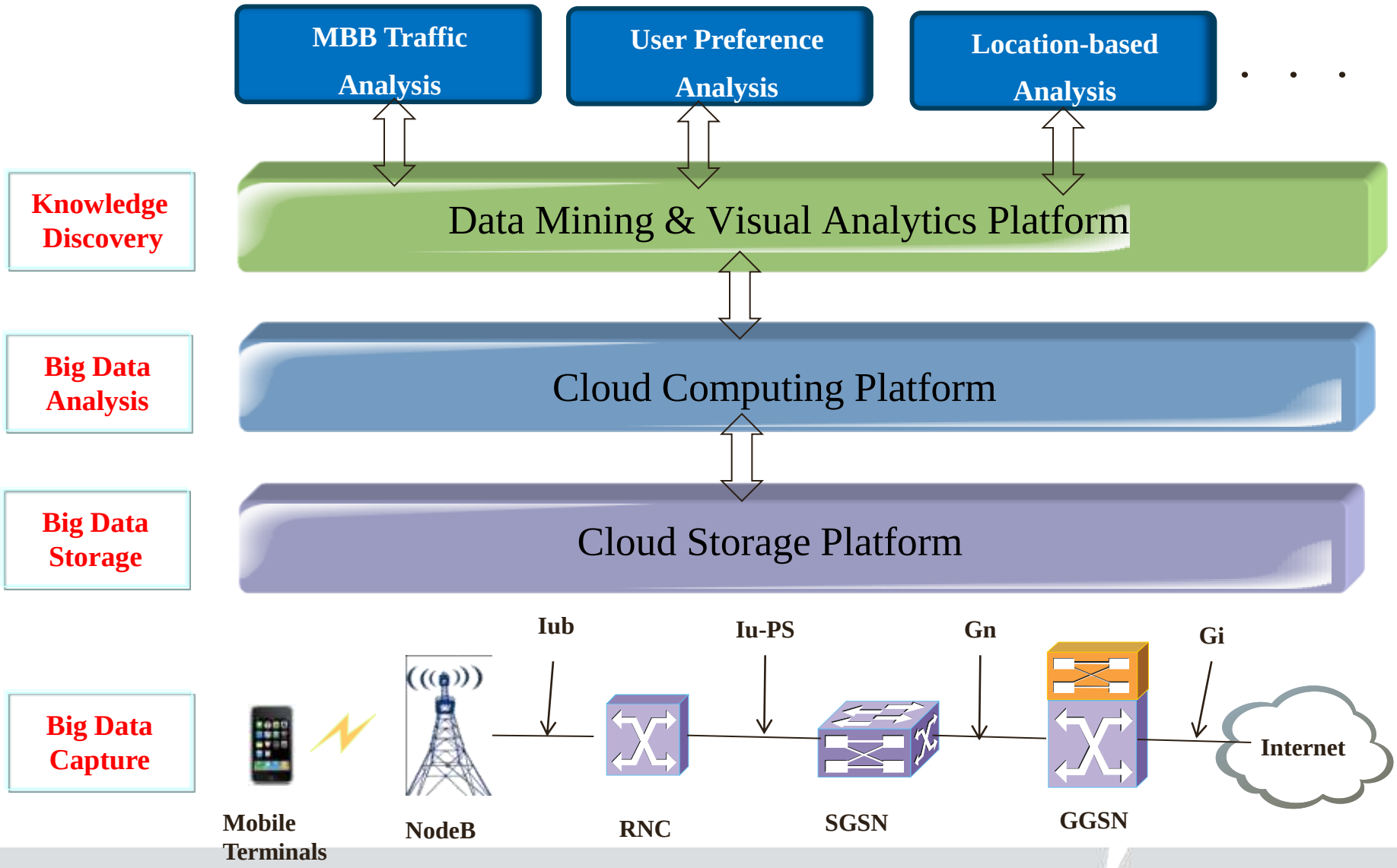
- **Bell Lab**
  ·ECT Research Domain;





Contents

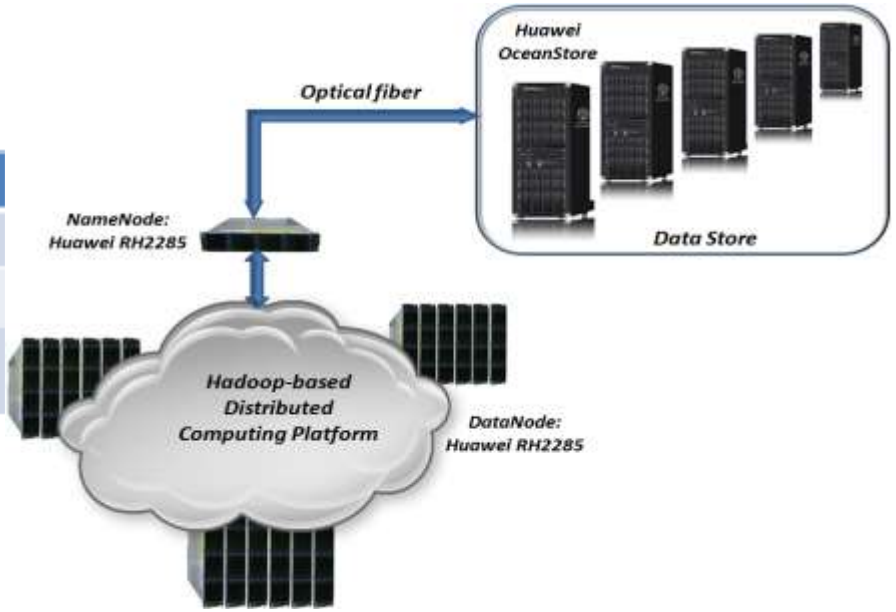
Systematic Overview



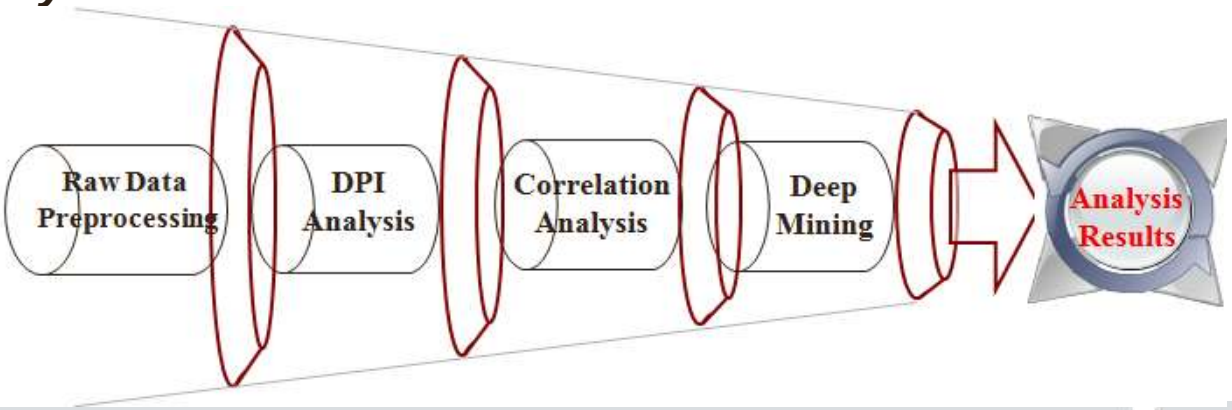
Cloud Computing Platform & Procedure

- *Hadoop based platform*

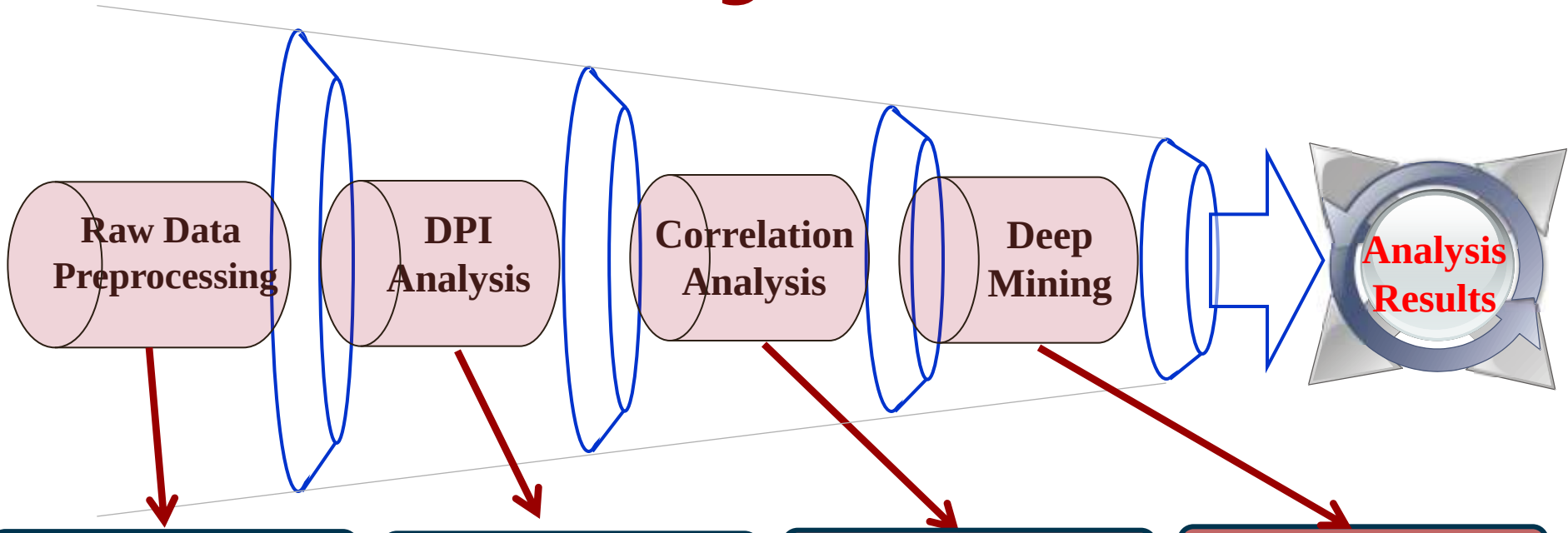
Total capability	
Storage	Near 1PB
Memory	>1600 GB
CPU	> 1600 cores Intel-Xeon@2.53GHZ



- *Big Data Analysis Procedure*



Issues & Challenges



Challenges

- Big Data Traces :
- Heterogeneous data transformation

Challenges

- Process the data sequentially;
- Identify miscellaneous apps by payload & flow based methods;
- Memory & Computation Intensive;

Challenges

- Correlate heterogeneous data;
- Lots of Data Join processes;
- Need time synchronization

Challenges

- Reveal insight pattern or models of the aggregate traffic, as well as different typical applications;
- Get more info. about users' behavior patterns;

Contents

Use Case - Traffic Analysis

◆ **What's impact the specific app. bring into the network**

` Story: CMCC & Tencent; 2010;

◆ **The fact is …**

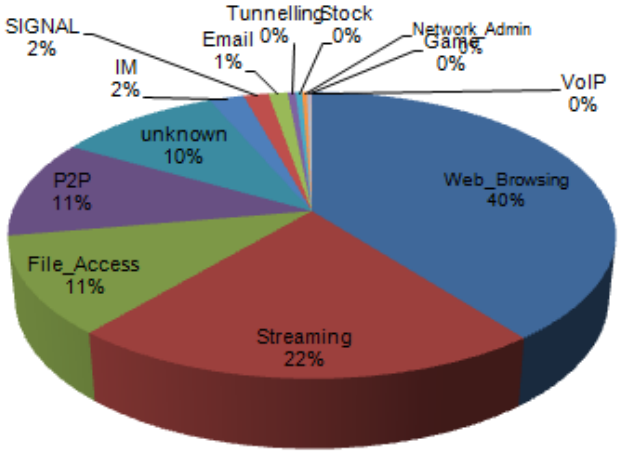
` There is only 2% of IM traffic, but

` Triggers more RRC connections than Web.

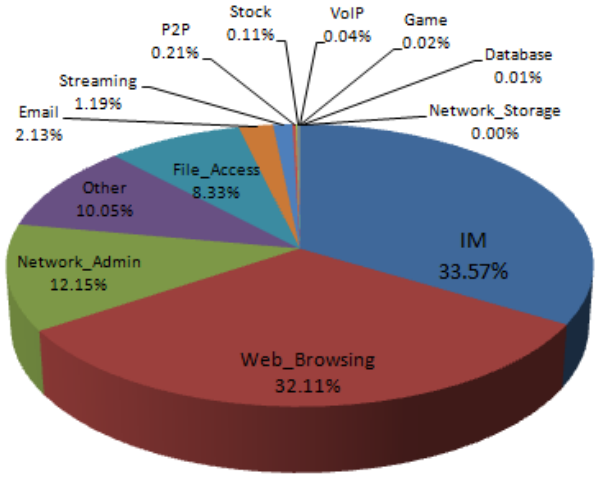
◆ **Potential Solutions…**

` A new IM protocol design for MBB;

` A new pricing package with IM items included in the tariff;



Traffic Volume Distribution of Apps.



Signal Distributions Triggered by Apps.

Use Case - Traffic Analysis

◆ **What's impact the specific type of terminal bring into the network**

` Story: AT&T & Apple Devices;

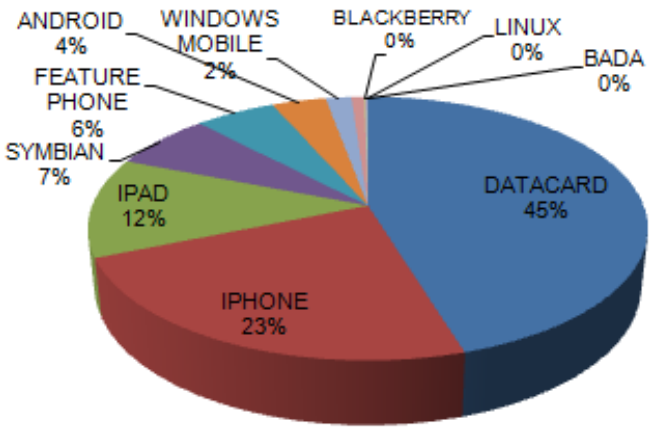
` BusinessWeek, The New York Times;

◆ **The fact is …**

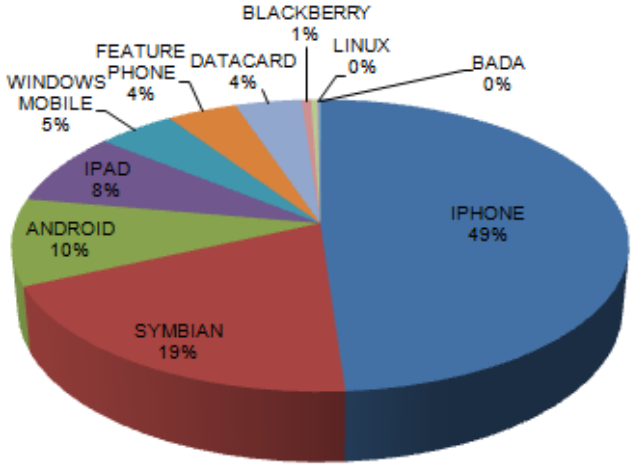
` Apple devices contribute 35% of total traffic;

` The figure will increase to 65% if only consider smart phones;

` 57% of the calls are initiated by Apple devices;



Traffic Volume Distribution of Devices.



App. Calls Distribution Initiated by Devices

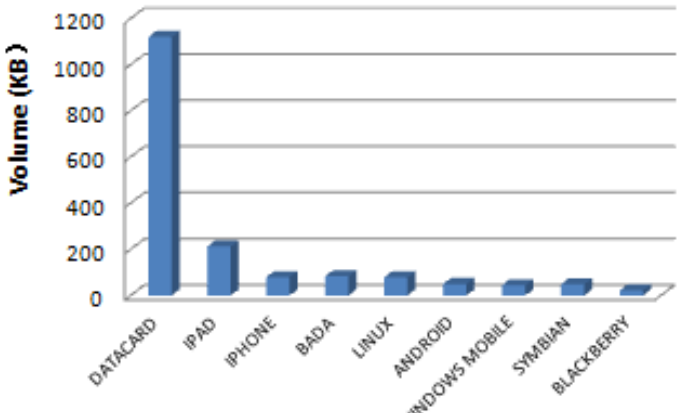
Use Case - Traffic Analysis

◆ **The Good news is …**

` Apple devices have relative higher resource utilization rate;

`In one Call;

`In one Unit Time (second);

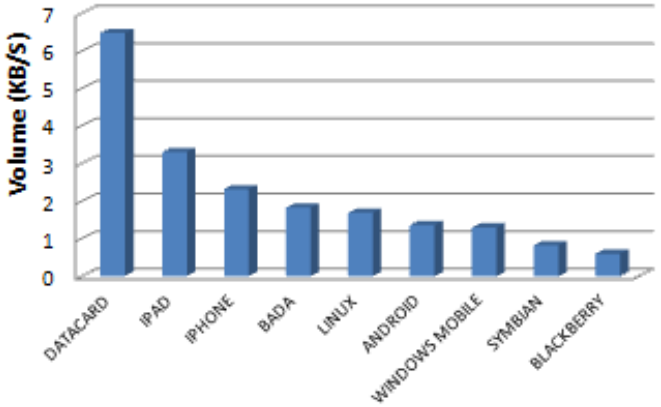


Avg. volume per call over devices

◆ **Potential solutions…**

` New Device Design takes Network Configurations into consideration;

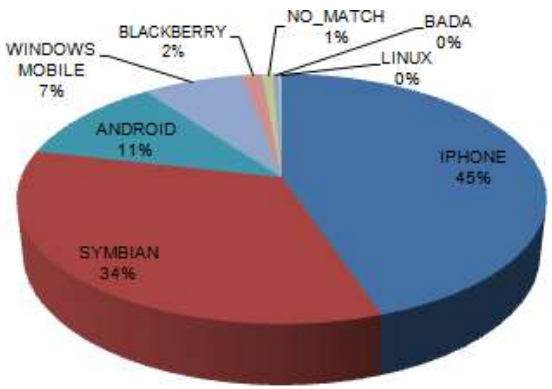
` New Pricing Package definition;



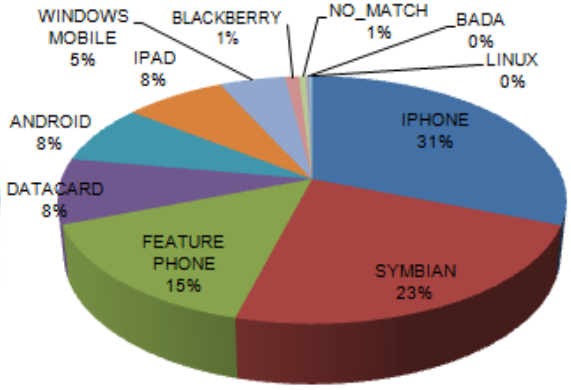
Avg. volume per second over devices

Use Case – Users' Preference & Behaviors

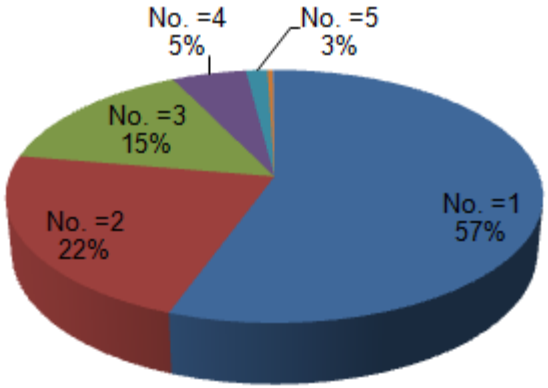
- **Why preference & behaviors**
  - ` **Better QoE Management**
    - ` **Resource Allocation;**
    - ` **Application-based Scheduling;**
    - ` **Network Planning;**
  - ` **Network Economics Considerations**
    - ` **Pricing;**
    - ` **Push Ads;**
    - ` **Recommendations;**
- **What we get …**
  - ` **Preferred Devices**
  - ` **Concurrent Applications**
  - **…**



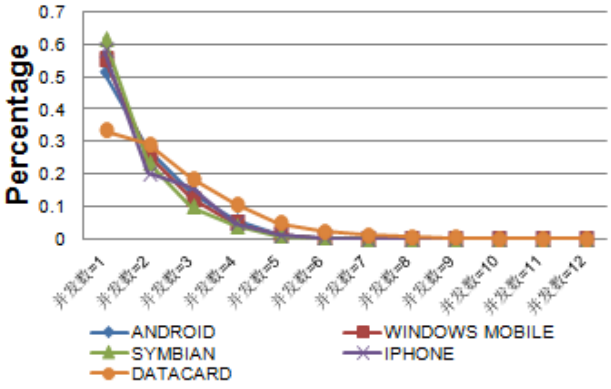
Distribution of the smart phones



Distribution of all devices



Concurrent Application

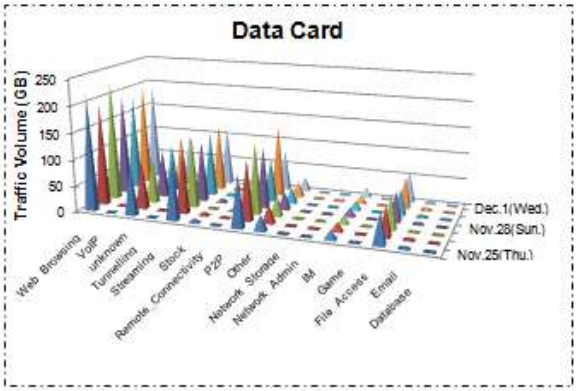


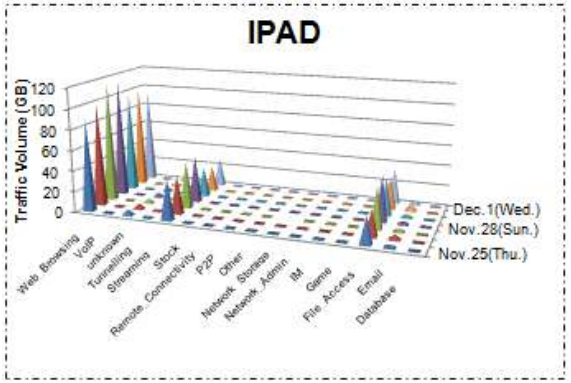
Concurrent Application over Devices

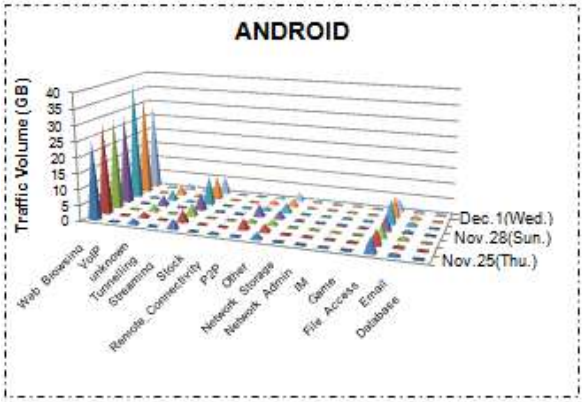
Use Case - Users 'preference & Behaviors

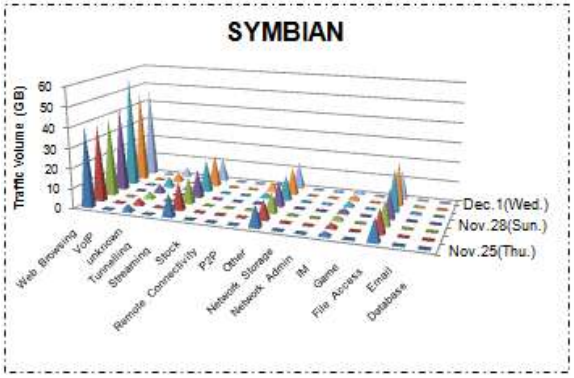
◆ **What's more …**

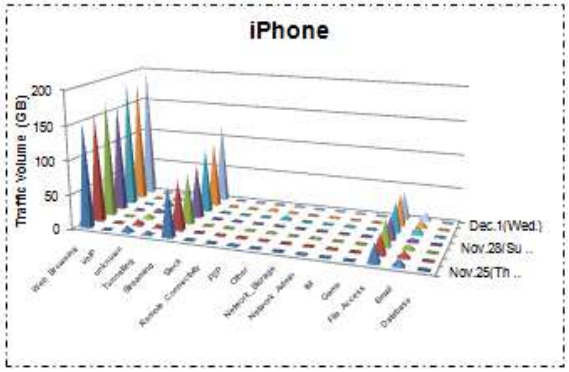
` Preferred Apps. classified by terminals

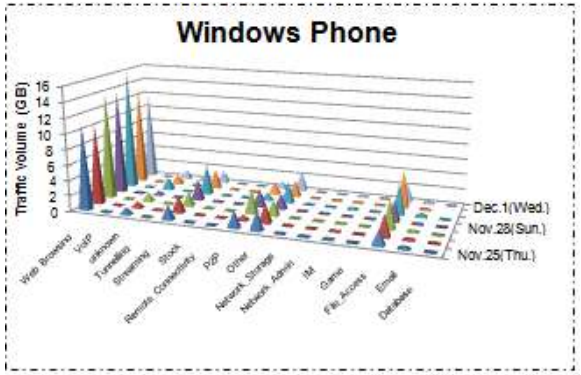












Use Case - Location-based analysis

◆ **Assumptions of Smart Transport**

` Enough GPS data;

` Sufficient sensor info.;

` Massive smart cameras;

◆ **What we can do:**

` Trajectory analysis of a single user;

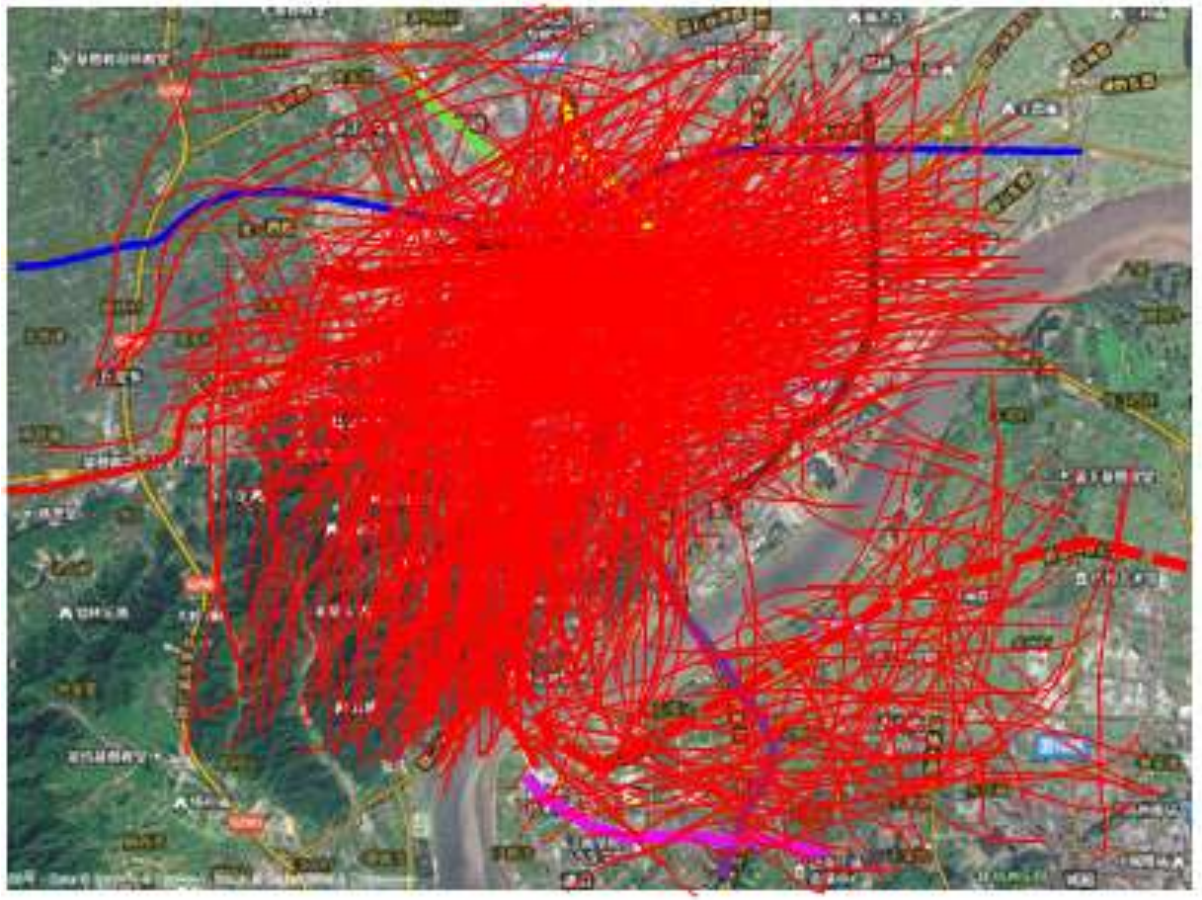
` Clustering of group users;

◆ **What the benefits we bring:**

` Traffic Prediction;

` Traffic Planning;

` Smarter Transport;



What else we can do…

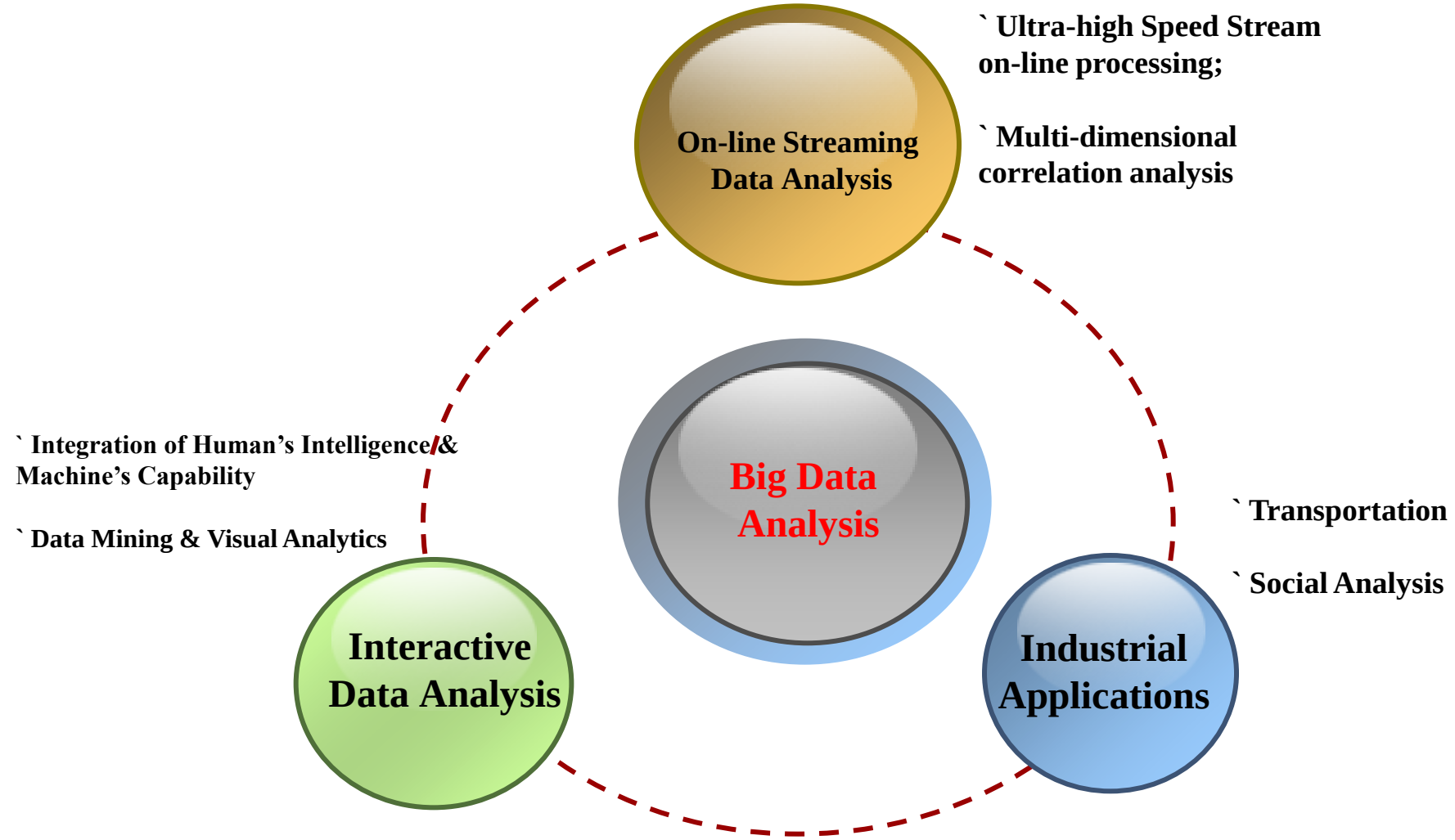
- **Further analysis based on the primary results**
  - ✓ Traffic modeling and synthetic traffic generation
  - ✓ Signal inference based only on data plane traffic
  - ✓ Traffic Prediction
  - ✓ Network Simulation
  - ✓ …

Contents

Summary & Future Plans

- **We almost can handle the challenges of the MBB Big Data:**
  - **Process nearly 9TB data per day;**
  - **The platform is scale free according to the new requirements;**
  - **Need App/Data specific algorithms enhancement;**
- **We do dig out knowledge from the Big Data, and used in:**
  - **Network Planning;**
  - **Consulting Service;**
  - **Simulation & Equipment Test;**
  - **Network Economics;**
  - **…**

Summary & Future Plans



诺亚方舟实验室
Noah's Ark Lab

Q&A

HUAWEI