Traffic Classification on Mobile Core Network Considering Regularity of Background Traffic

KDDI R&D Laboratories,
Masaki Suzuki, Masafumi Watari, Shigehiro Ano

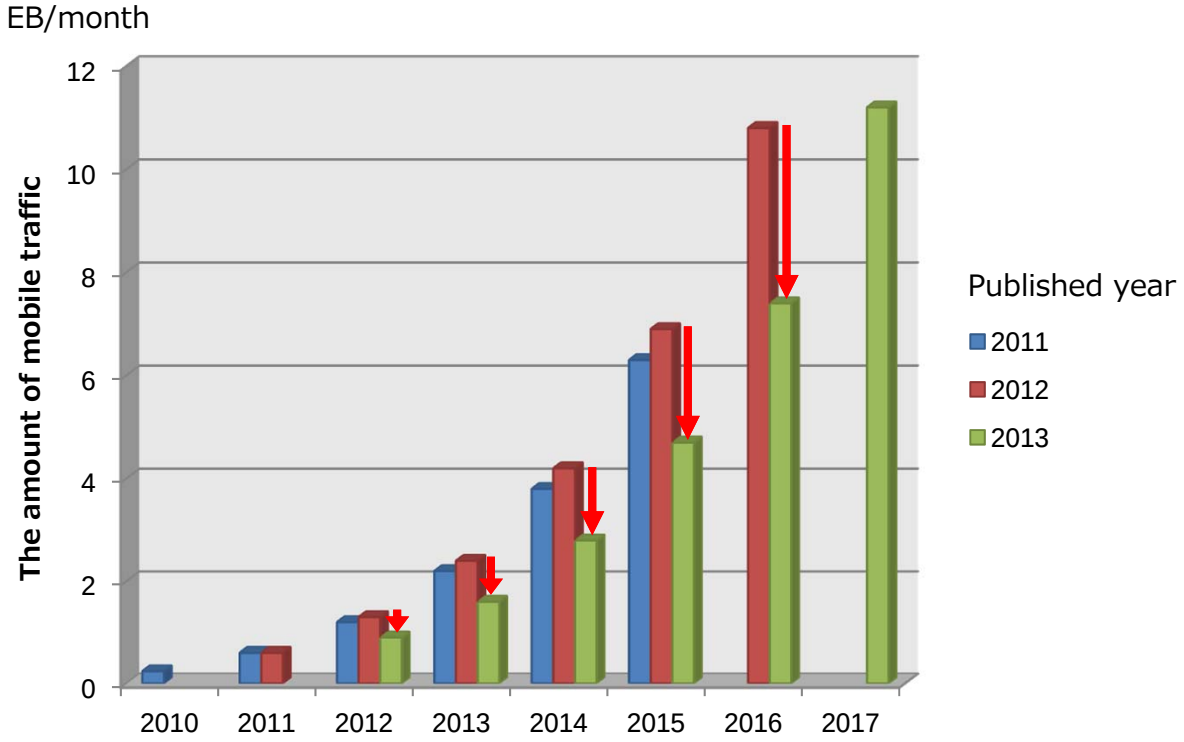
Kyushu Institute of Technology,
Masato Tsuru

May 12, 2015

KDDI R&D LABS

Considerable increase of mobile traffic

- Growth estimation for mobile traffic in 2013
  - Total amount was downwardly revised

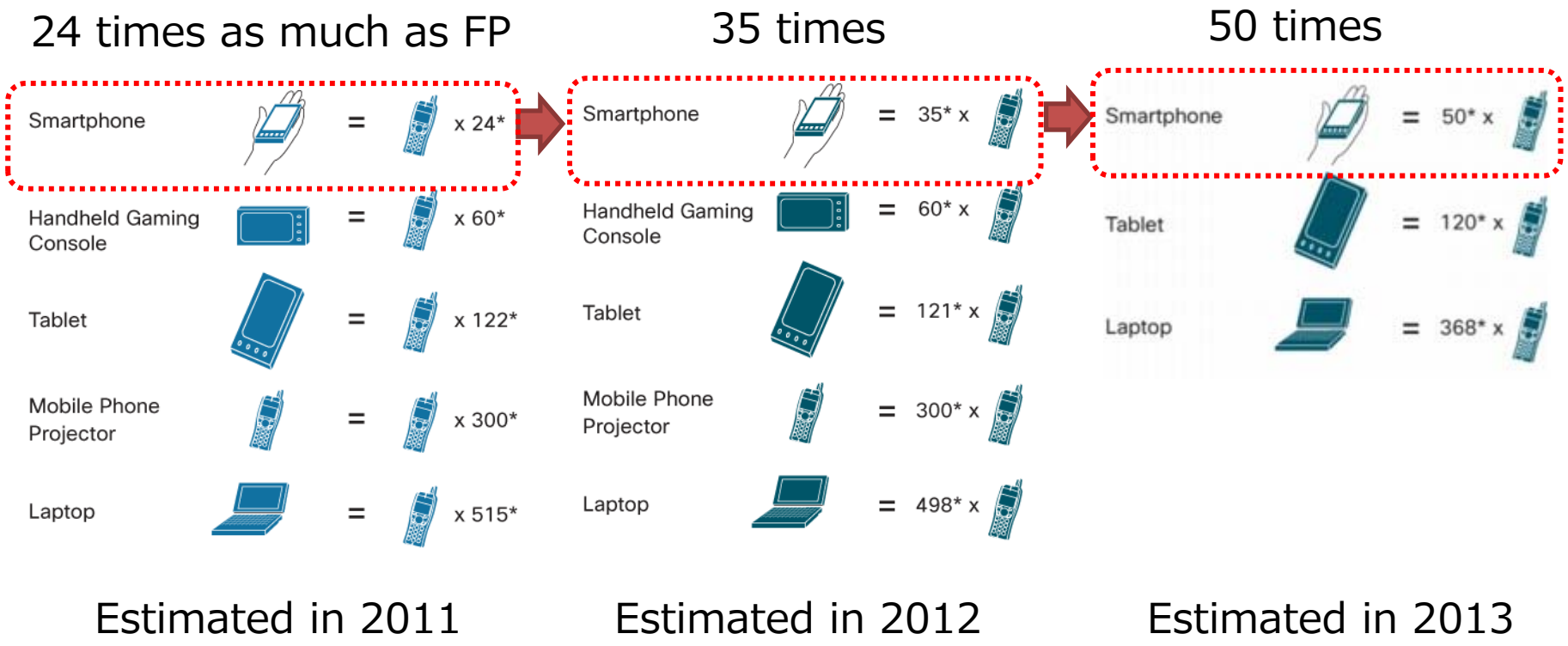


* Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015, 2011-2024, 2012-2017

KDDI R&D LABS

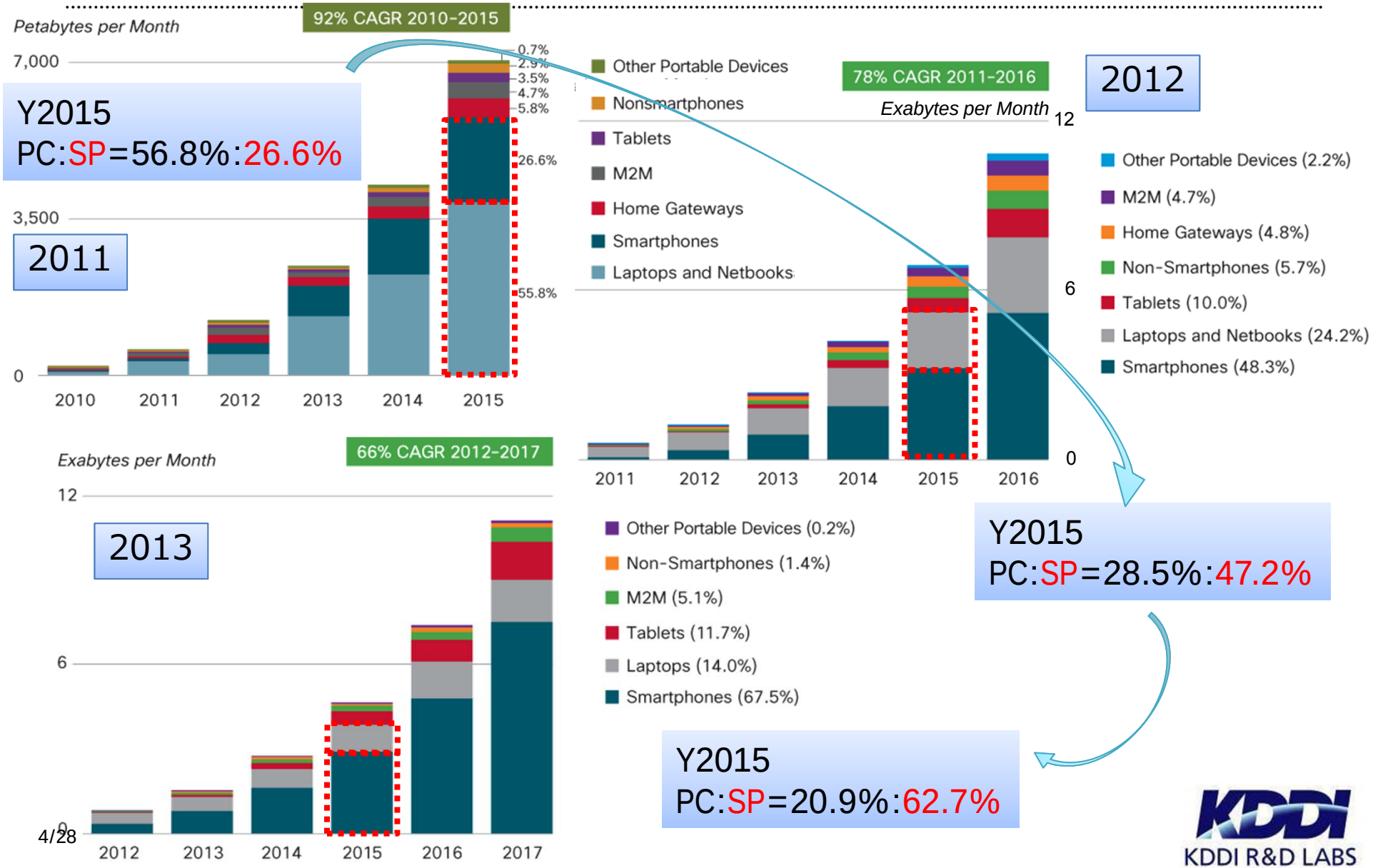
Smartphone traffic

- The amount of smartphone traffic in 2015 was upwardly revised (about twice)



KDDI R&D LABS

Growth of smartphone traffic



KDDI R&D LABS

Various apps on smartphone

Web browsing

Social networking

Streaming

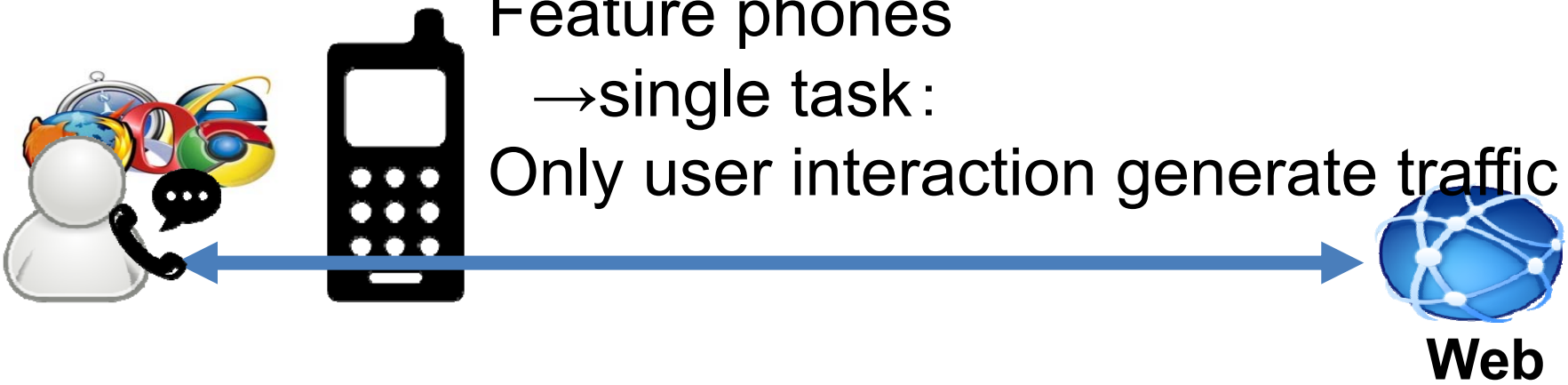
Information providers

Cloud storage

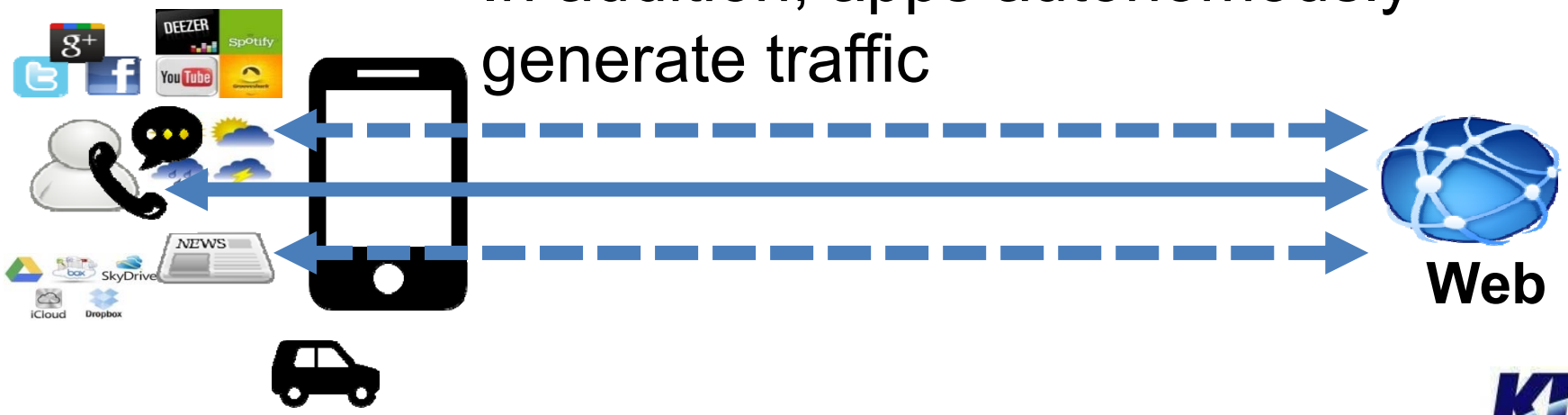
. . .

KDDI R&D LABS

Multi tasks on smartphone



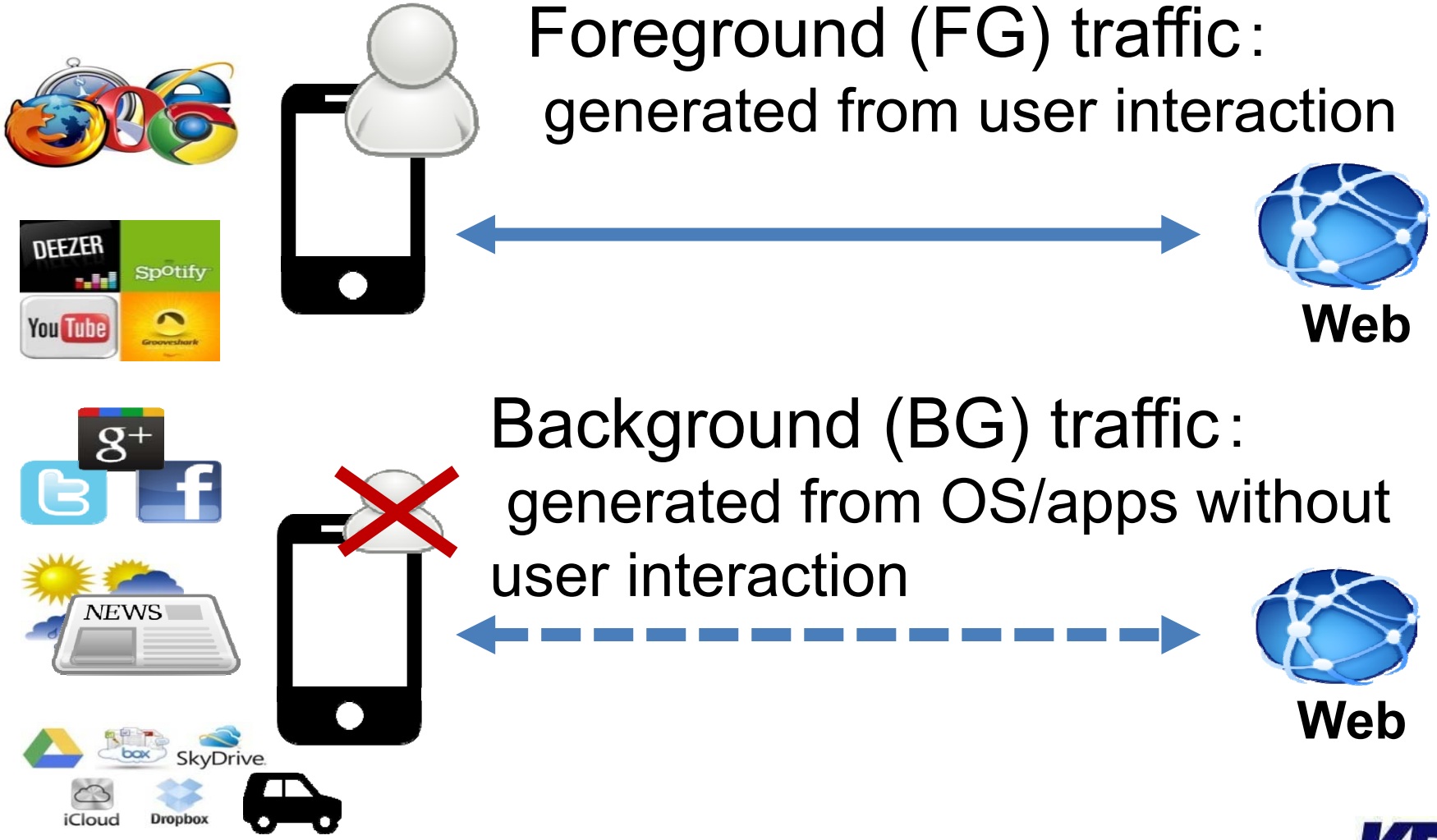
Feature phones
→single task：
Only user interaction generate traffic



Smartphone → multi tasks：
In addition, apps autonomously generate traffic

KDDI R&D LABS

FG/BG traffic



Foreground (FG) traffic:
generated from user interaction

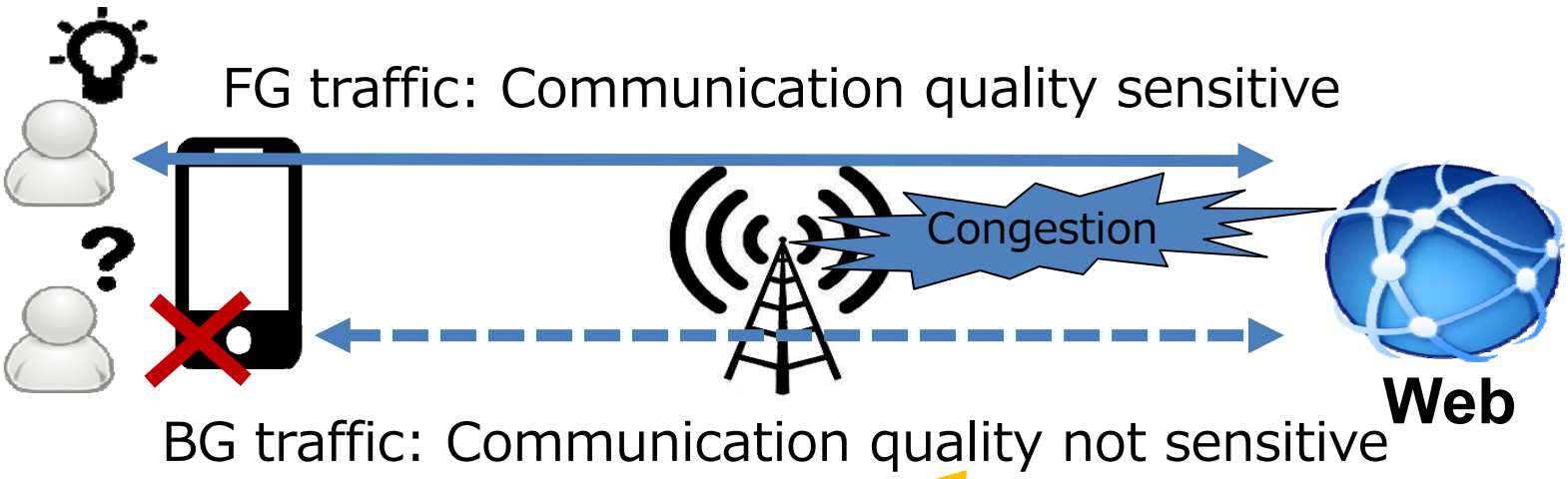
Background (BG) traffic:
generated from OS/apps without user interaction

KDDI R&D LABS

Features of FG/BG traffic

In the case of congestion

- FG traffic: users are likely to notice
- BG traffic: users are unlikely to notice

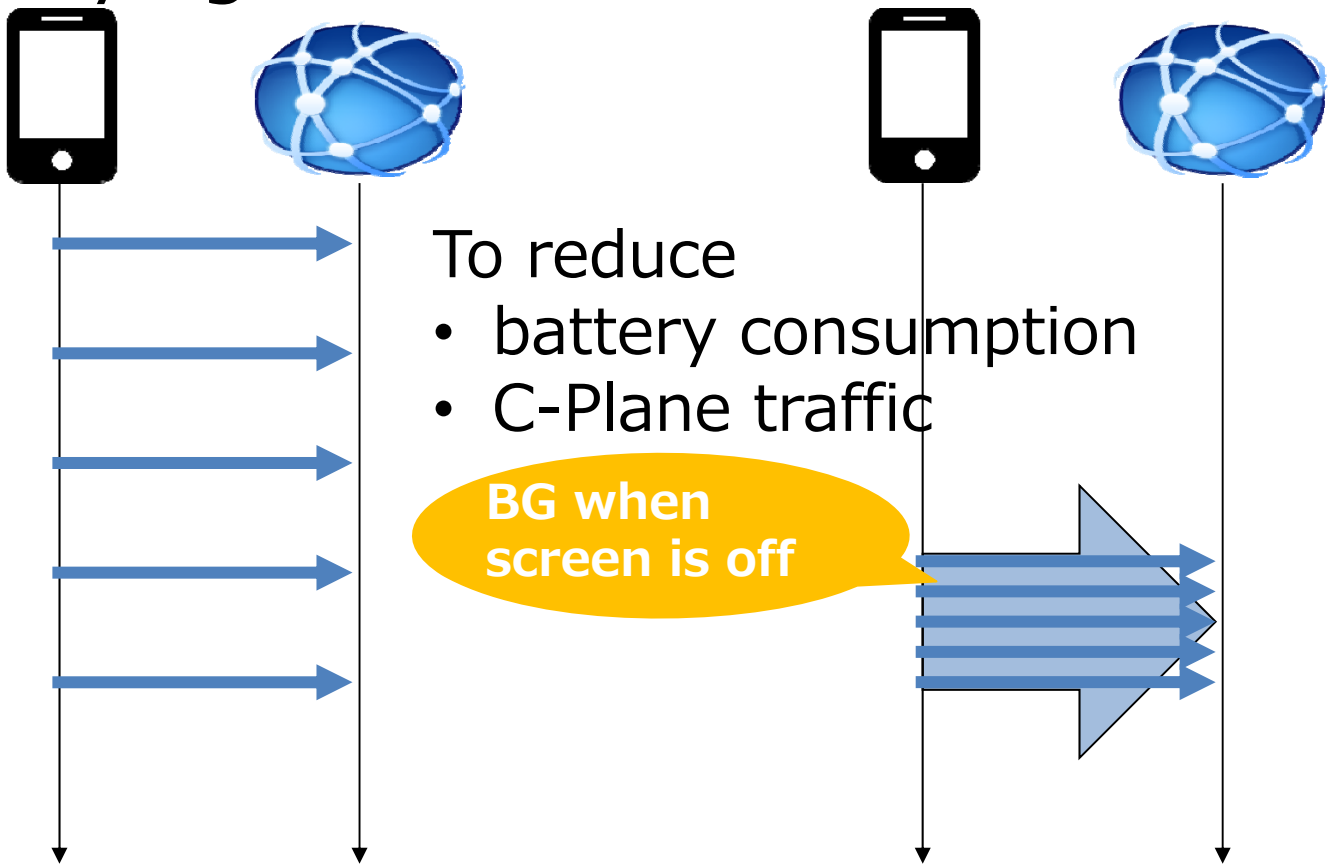


FG/BG classification is important

KDDI R&D LABS

Network Socket Request Manager

- Merging BG traffic into dedicated time slot
- Classifying FG/BG based on device condition

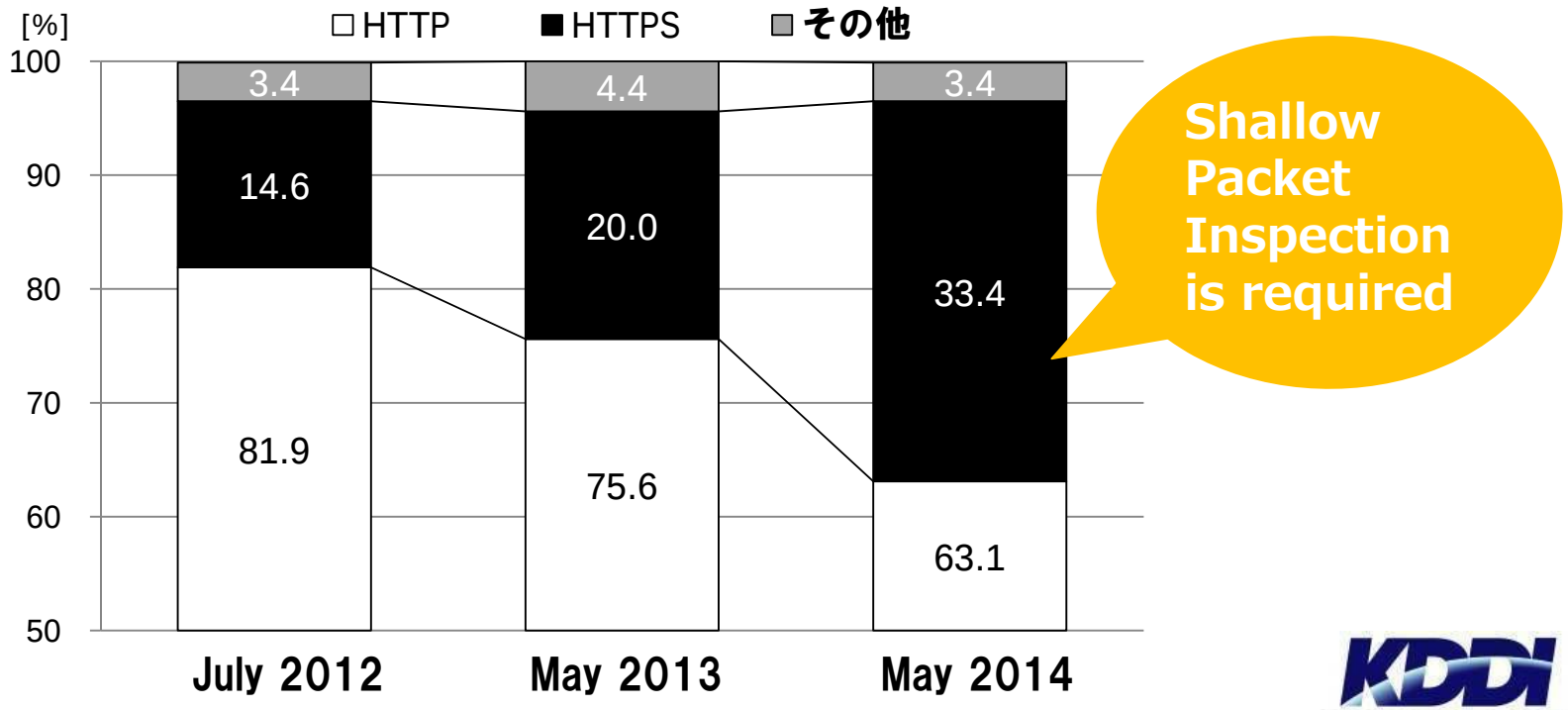


 * Qualcomm "Managing Background Data Traffic in Mobile Devices (2012)

KDDI R&D LABS

Increase of TLS traffic

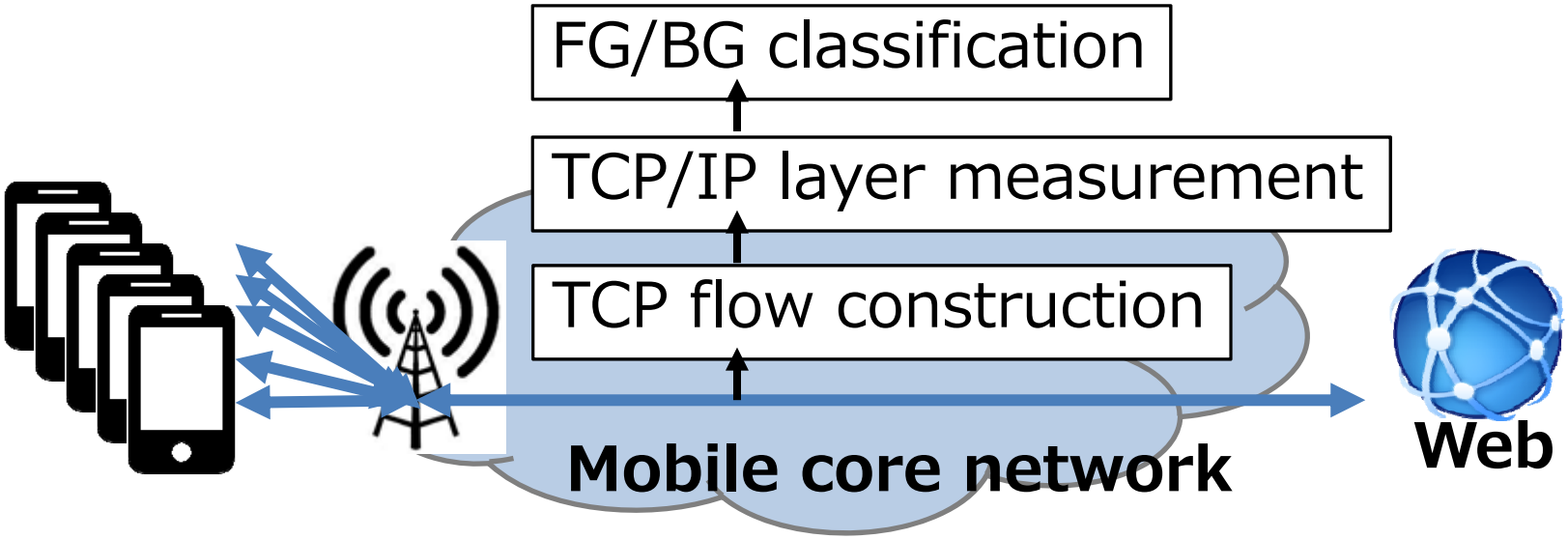
- TLS traffic in KDDI LTE network is gradually increasing
- Deep Packet Inspection will not be efficient any more



KDDI R&D LABS

Traffic classification on mobile core network

- Packet caption at b/w eNB and MME
- Measurement of TCP/IP layer metrics
- FG/BG classification by SVM classifier



KDDI R&D LABS

TCP flow construction

- Identify uplink and downlink
- 4-tuple is a key for TCP flows
- Skip packets if their payload size equal 0 (SYN, SIN, RST, ACK), for differentiation

Classifiers

- Single-index classifier:
  - Consists of single index
- Full-index classifier:
  - Combination of 40 indexes
- Shortcut-index classifier:
  - Combination of 37 indexes which can be measured in the initial part of flow

SVM classifier (base line)

	Index name
1	Entropy of pkt size for UL
2	Entropy of pkt size for DL
3	Avg pkt size for UL
4	Avg pkt size for DL
5	CV of pkt size for UL
6	CV of pkt size for DL
7	Flow duration
8	Frequency of pkt sending
9	Frequency of pkt receiving
10	Ratio of frequency of pkt
11	CV of sent pkt interval
12	CV of received pkt interval
13	Avg of throughput for UL
14	Avg of throughput for DL

label :

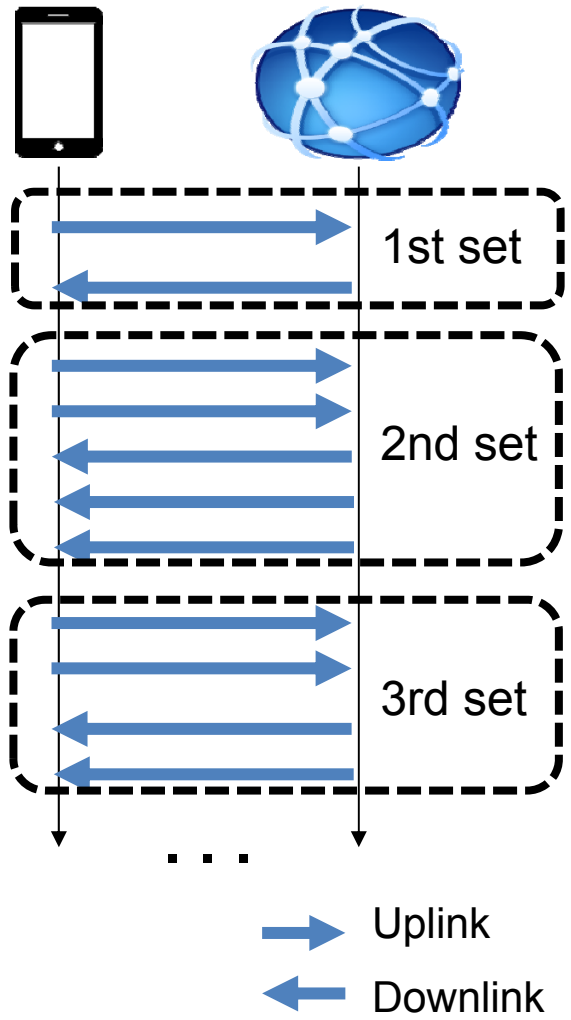
$$A = \begin{cases} \mathrm{FG} & (\text{if } \vec{a} \cdot \vec{x} + b \geq 0) \\ \mathrm{BG} & (\text{otherwise}) \end{cases}$$

- optimized weight parameters are defined by experiments

KDDI R&D LABS

SVM classifier (in addition)

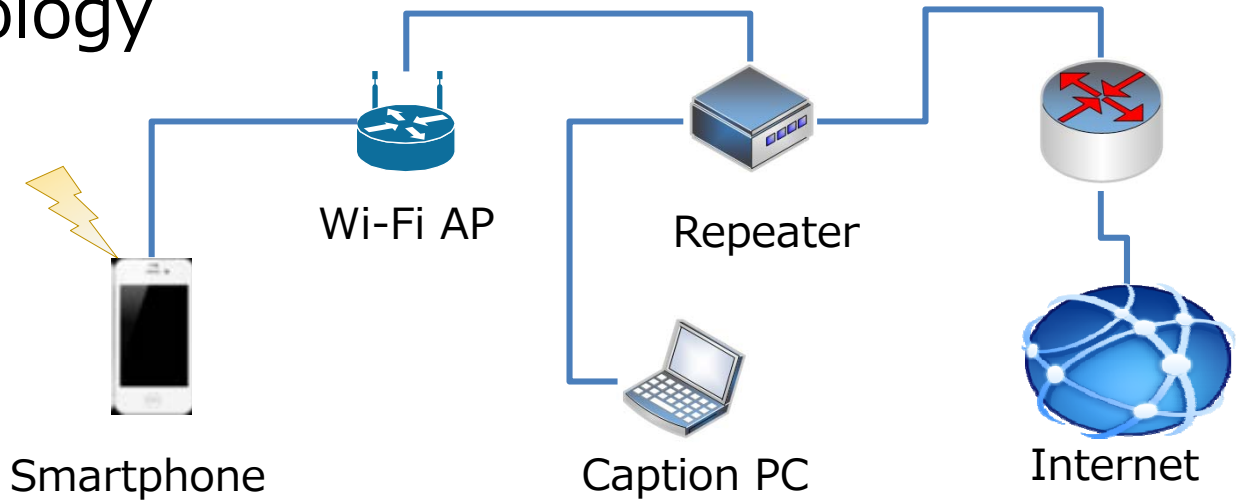
	Index name
15	Num. of same size of pkt sent for UL
16	Num. of same size of pkt sent for DL
17	Num. of MTU size of pkt sent for UL
18	Num. of MTU size of pkt sent for DL
19	Num. of not MTU size of pkt sent for UL
20	Num. of not MTU size of pkt sent for DL
21	Time duration from 1st pkt for UL to 2nd pkt for UL
22	Time duration from 1st pkt for UL to 3rd pkt for UL
23	Time duration from 1st pkt for DL to 2nd pkt for DL
24	Time duration from 1st pkt for DL to 3rd pkt for DL
25	Time duration from 1st pkt for UL to 1st pkt for DL
26	Time duration from 1st pkt for UL to 2nd pkt for DL
27	Time duration from 1st pkt for UL to 3rd pkt for DL
28	Time duration from 1st pkt for UL to 4th pkt for DL
29	Time duration from 1st pkt to 2nd pkt
30	Time duration from 1st pkt to 3rd pkt
31	Time duration from 1st pkt to 4th pkt
32	Time duration from 1st pkt to 5th pkt
33	Time duration from 1st set to 2nd set
34	Time duration from 1st set to 3rd set
35	Time duration from 1st set to 4th set
36	Time duration from 1st set to 5th set
37	Average time duration between sets from 1st to 5th
38	CV of time duration between sets from 1st to 5th
39	Average time duration between sets
40	CV of time duration between sets



KDDI R&D LABS

Experiments

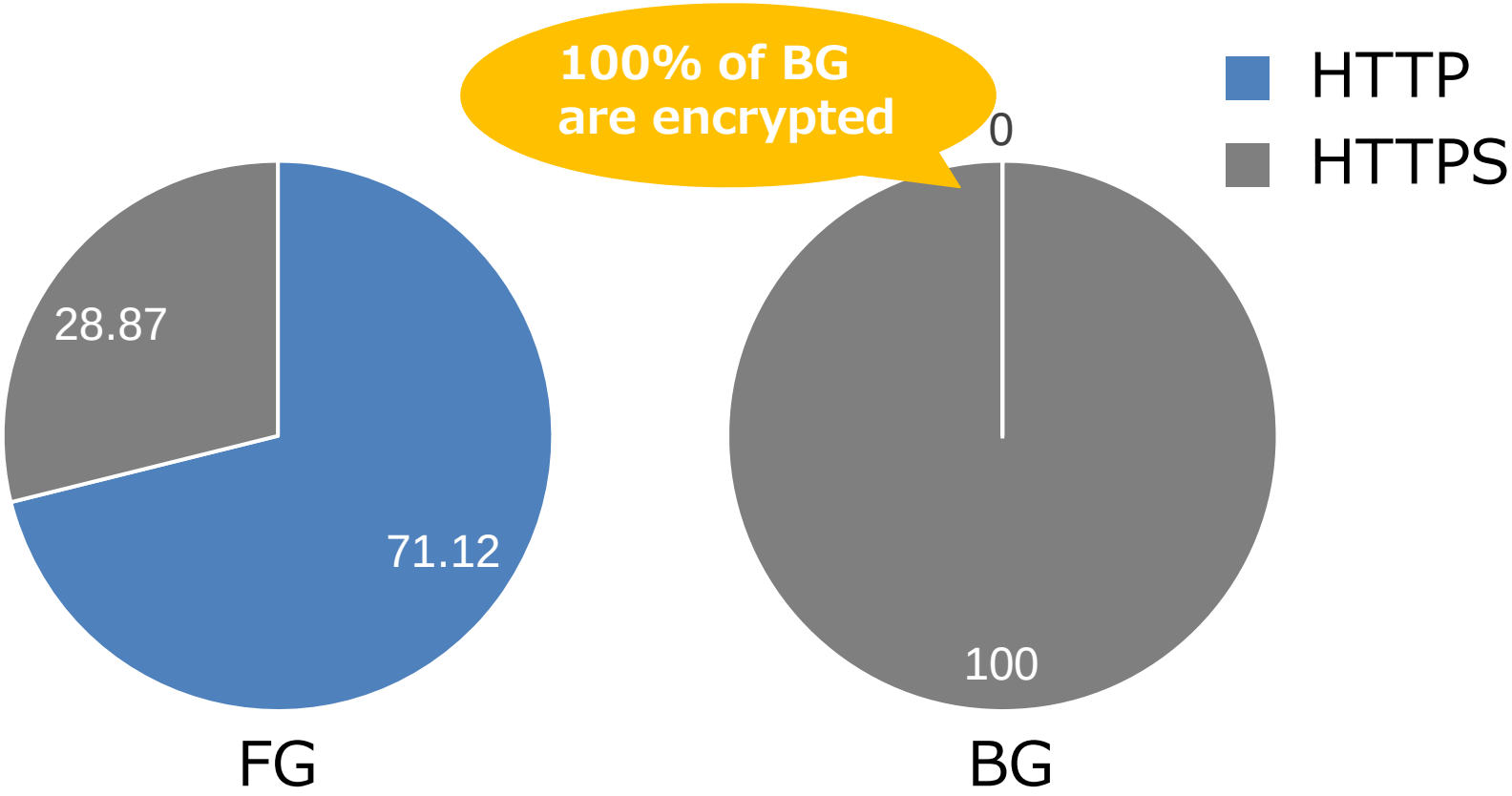
- Topology



- Smartphone
  - HTC ISW13HT（Android 4.0.3）
  - Applications which generate traffic without user interaction from Google play ranking in Aug., 2013
    → BG traffic
  - Testers use Safari, Facebook, Twitter and YouTube one by one
    → FG traffic

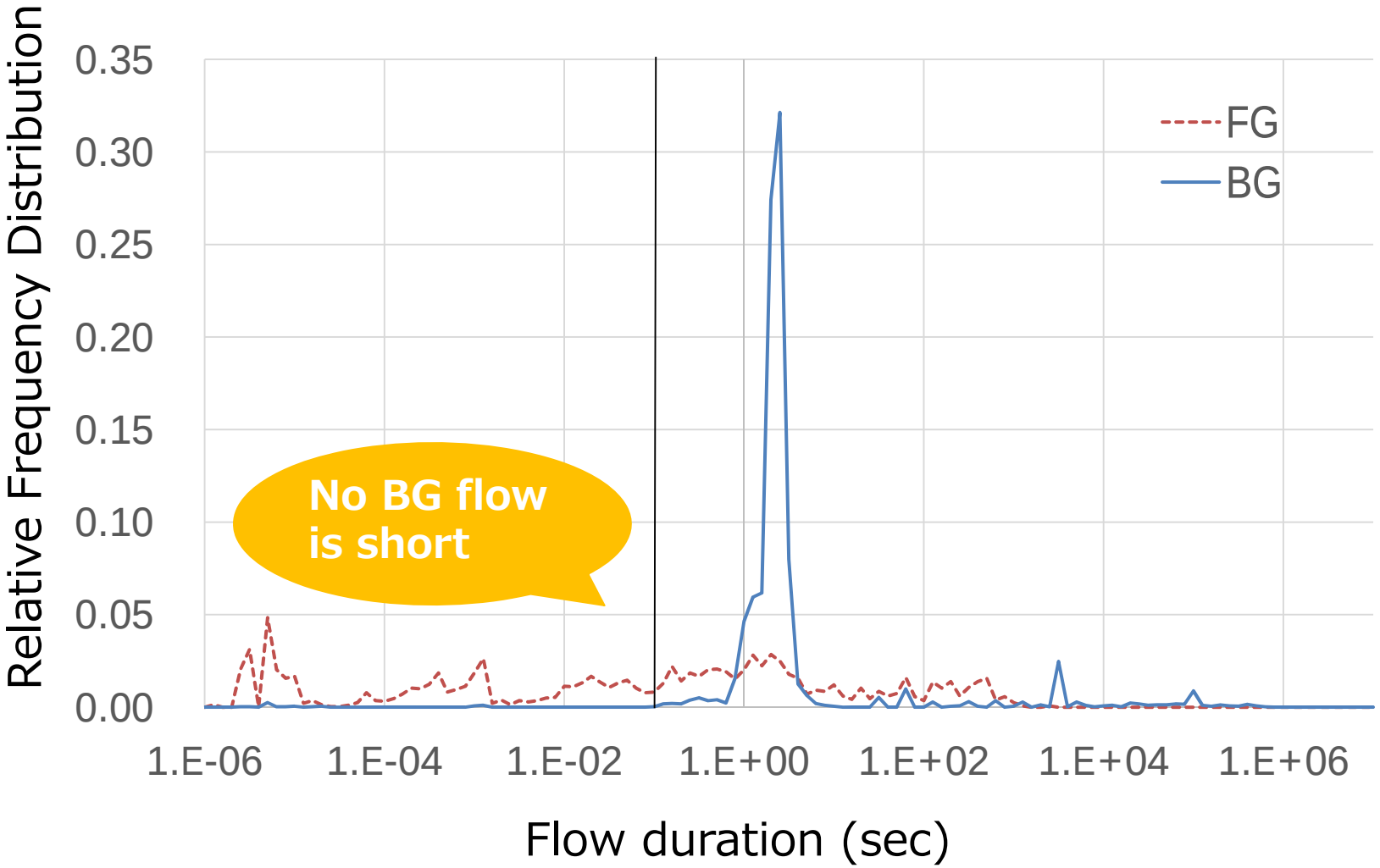
KDDI R&D LABS

Proportion of TLS traffic



KDDI R&D LABS

Flow duration



KDDI R&D LABS

Full-index classifier accuracy

- 10-fold cross validation using 2600 flows of data set

average accuracy is 96.1%

	Labeled as FG	Labeled as BG
FG	1270 flows precision: 0.978 recall: 0.947	29 flows
BG	71 flows	1221 flows Precision: 0.945 Recall: 0.977

KDDI R&D LABS

Accuracy/Weight of each metric

#	Accuracy of single classifier	F-value	Weight value
1	78.70	0.78	-1.36
2	80.43	0.80	-0.82
3	87.92	0.88	4.91
4	78.12	0.77	-0.66
5	77.50	0.78	-0.9
6	86.84	0.87	-6.03
7	53.30	0.50	1.59
8	58.20	0.55	-1.08
9	40.64	0.36	-0.75
10	59.44	0.56	3.24
11	66.38	0.62	1.57
12	67.19	0.63	0.49
13	46.28	0.39	4.38
14	42.18	0.37	0.51
…			

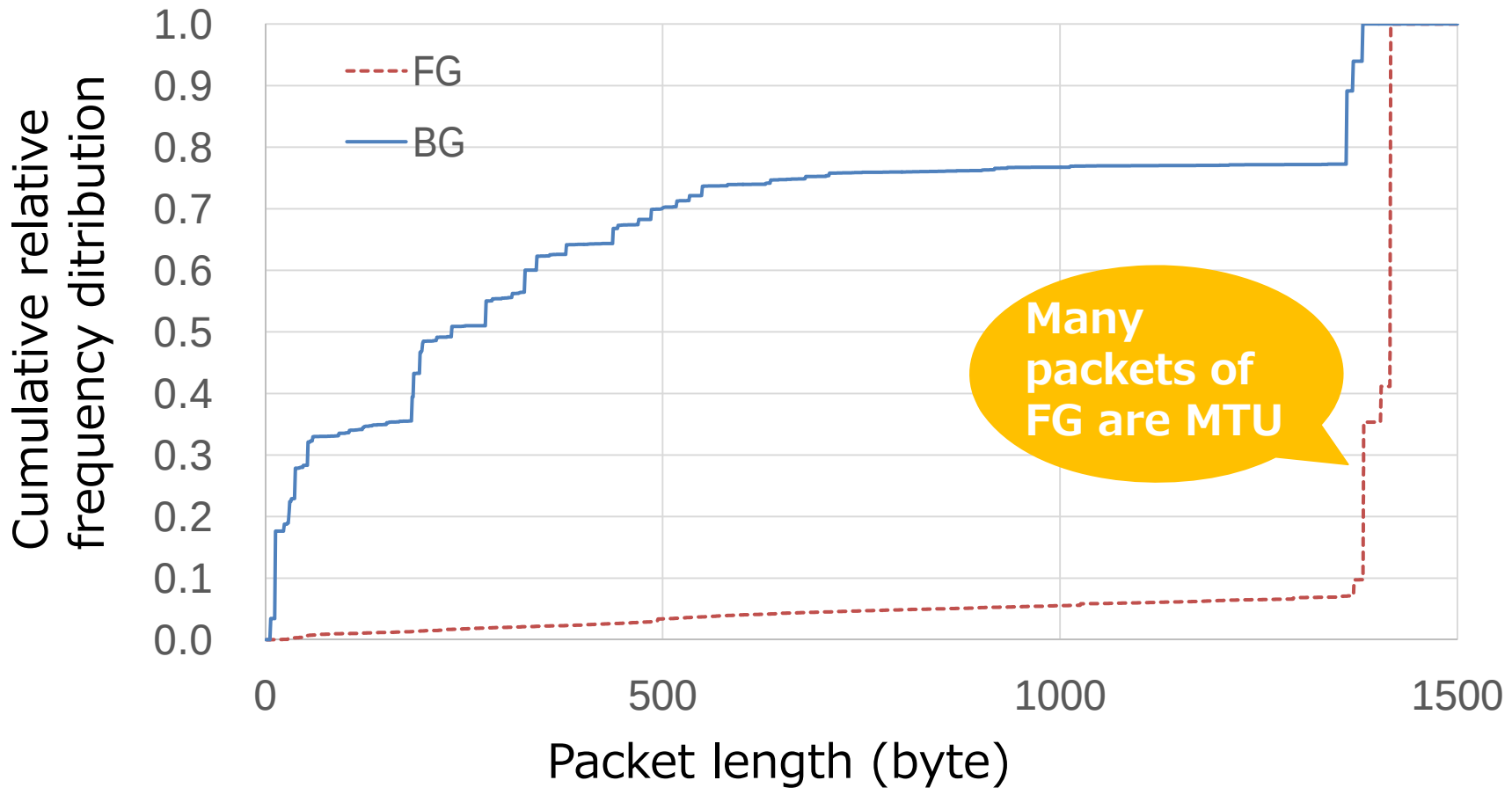
"Average packet size for uplink"

"CV of packet size for downlink"

These metrics are effective both in single-index and full-index classifier

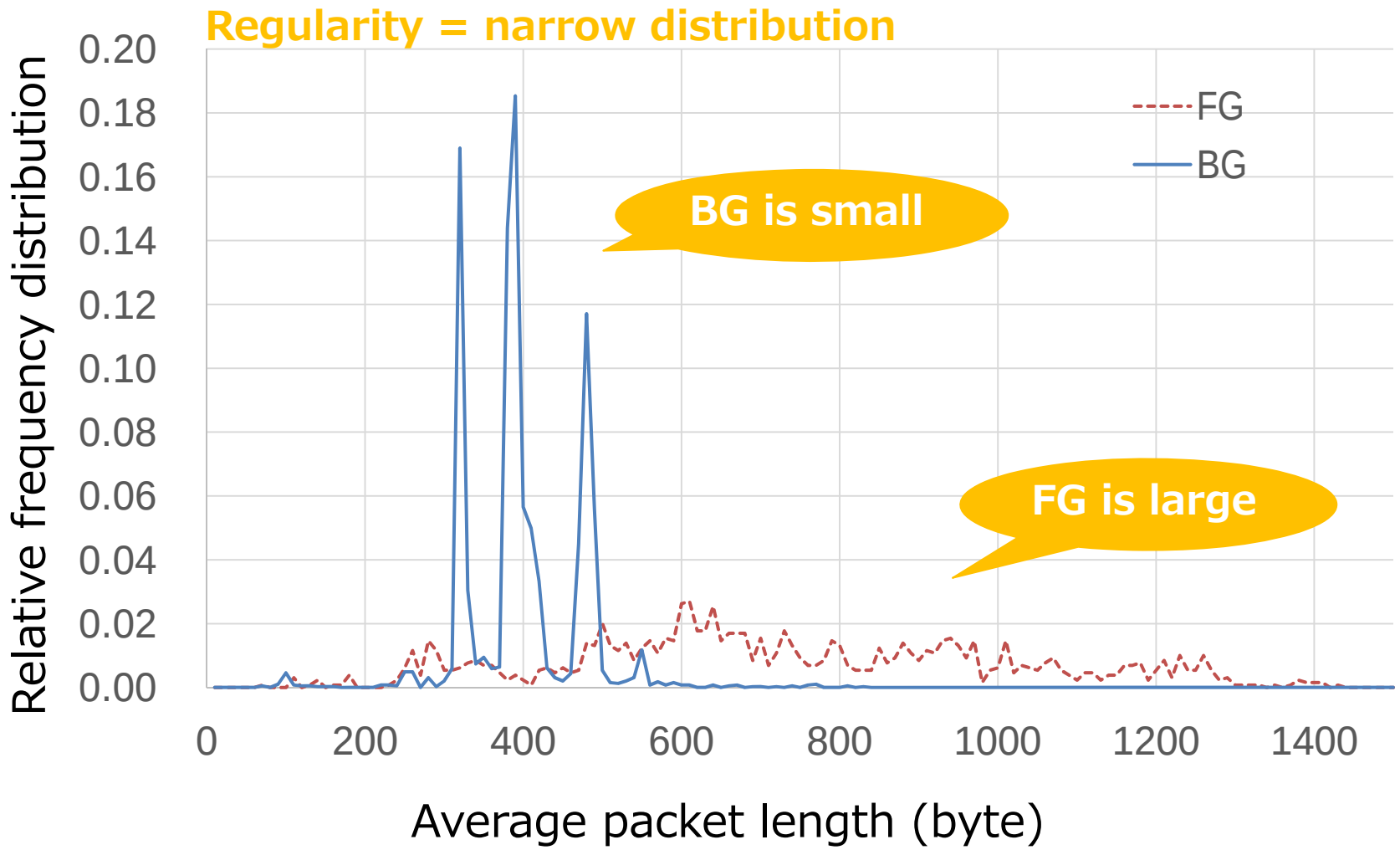
KDDI R&D LABS

Packet length: cumulative relative frequency distribution



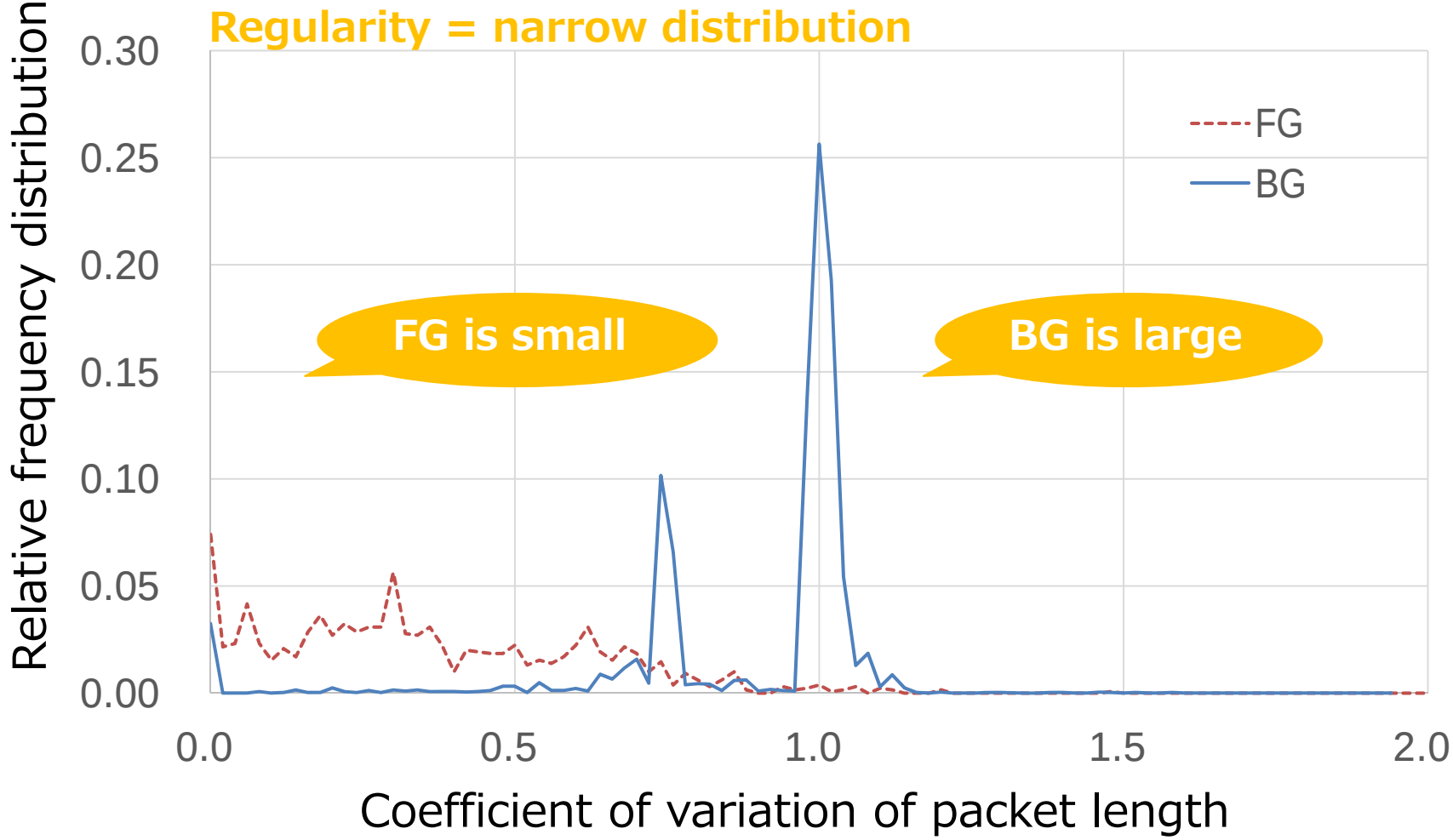
KDDI R&D LABS

Average packet length for uplink



KDDI R&D LABS

Coefficient of Variation of packet length



KDDI R&D LABS

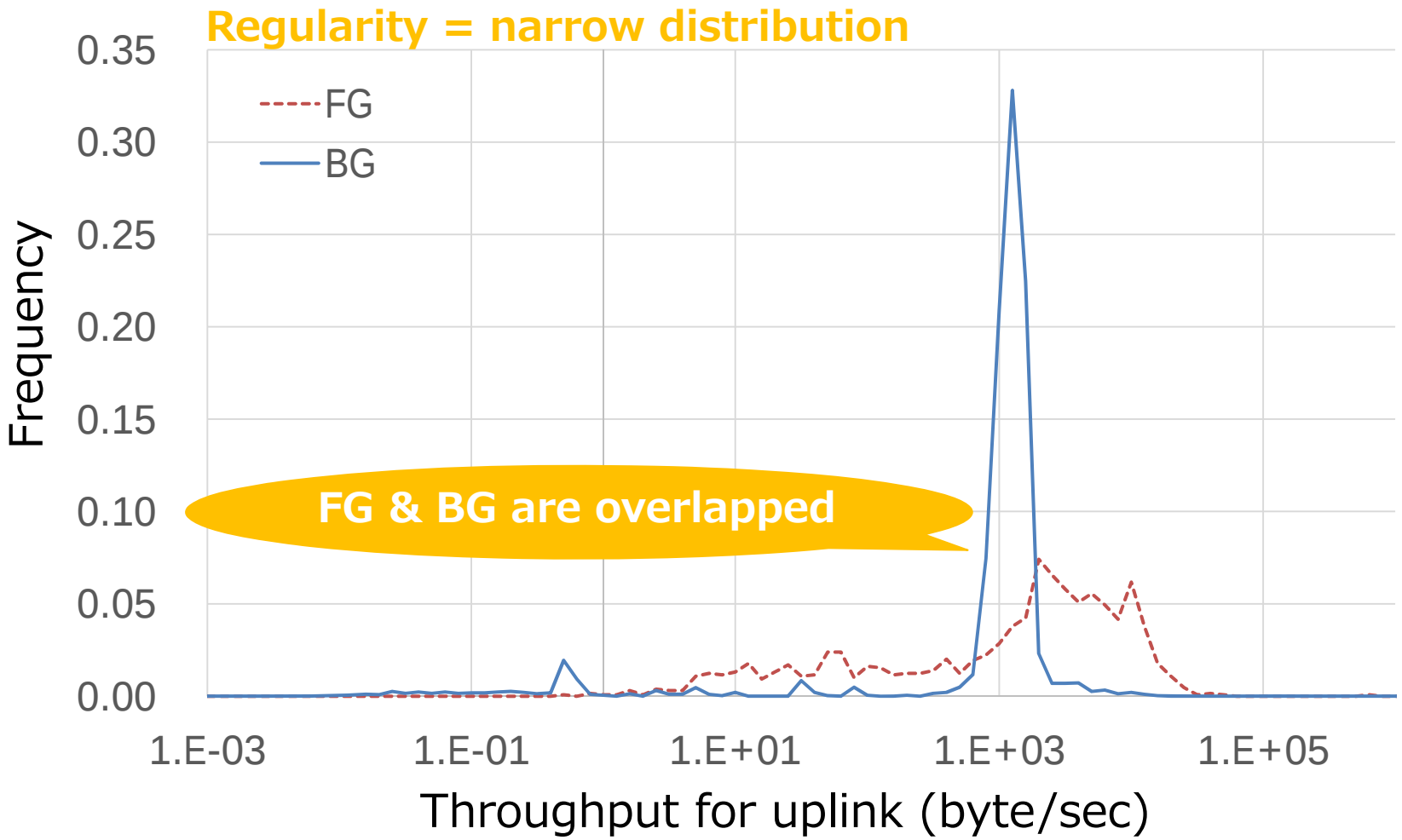
Accuracy/Weight of each metric

#	Accuracy of single classifier	F-value	Weight value
1	78.70	0.78	-1.36
2	80.43	0.80	-0.82
3	87.92	0.88	4.91
4	78.12	0.77	-0.66
5	77.50	0.78	-0.9
6	86.84	0.87	-6.03
7	53.30	0.50	1.59
8	58.20	0.55	-1.08
9	40.64	0.36	-0.75
10	59.44	0.56	3.24
11	66.38	0.62	1.57
12	67.19	0.63	0.49
13	46.28	0.39	4.38
14	42.18	0.37	0.51
…			

"Throughput for uplink" is not effective in single-index classifier. But effective in full-index classifier.

KDDI R&D LABS

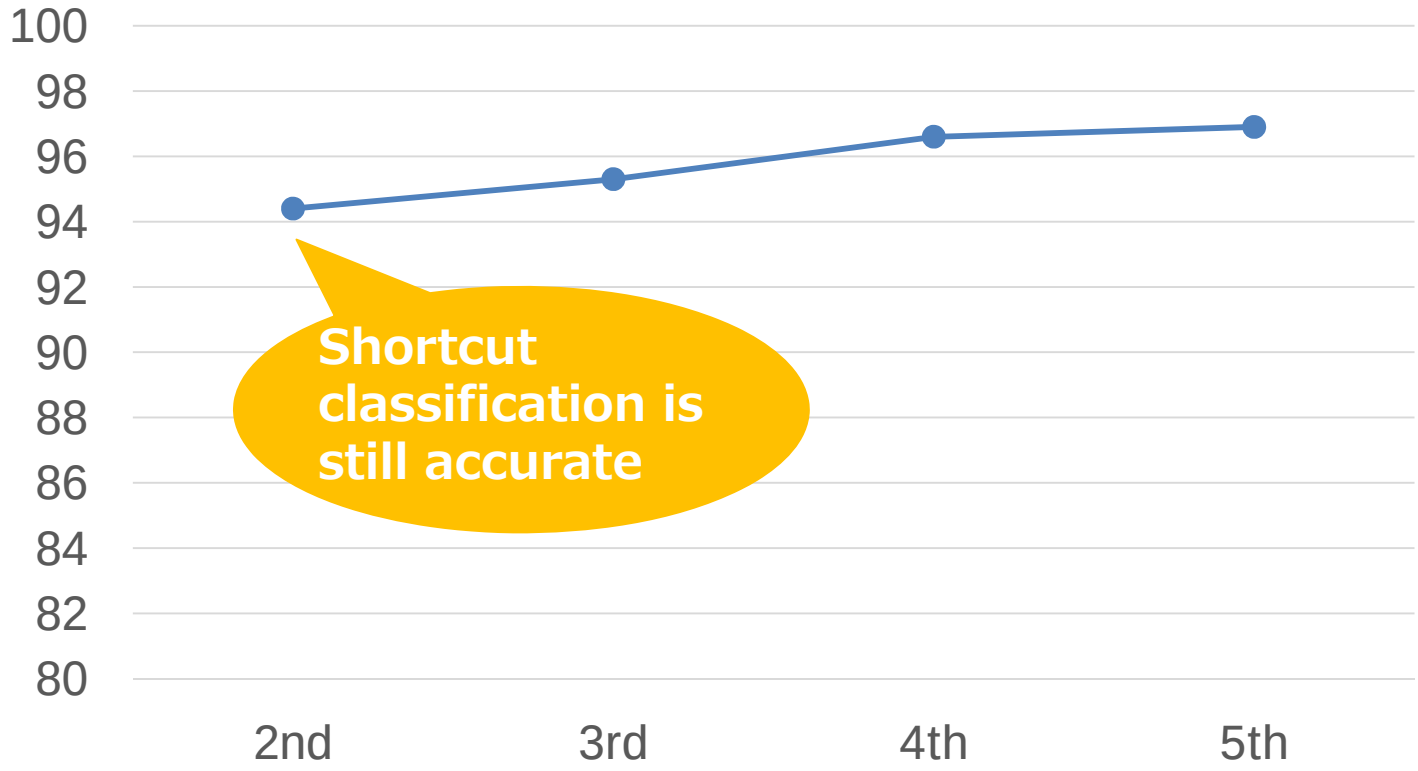
Throughput for uplink



KDDI R&D LABS

Shortcut classification

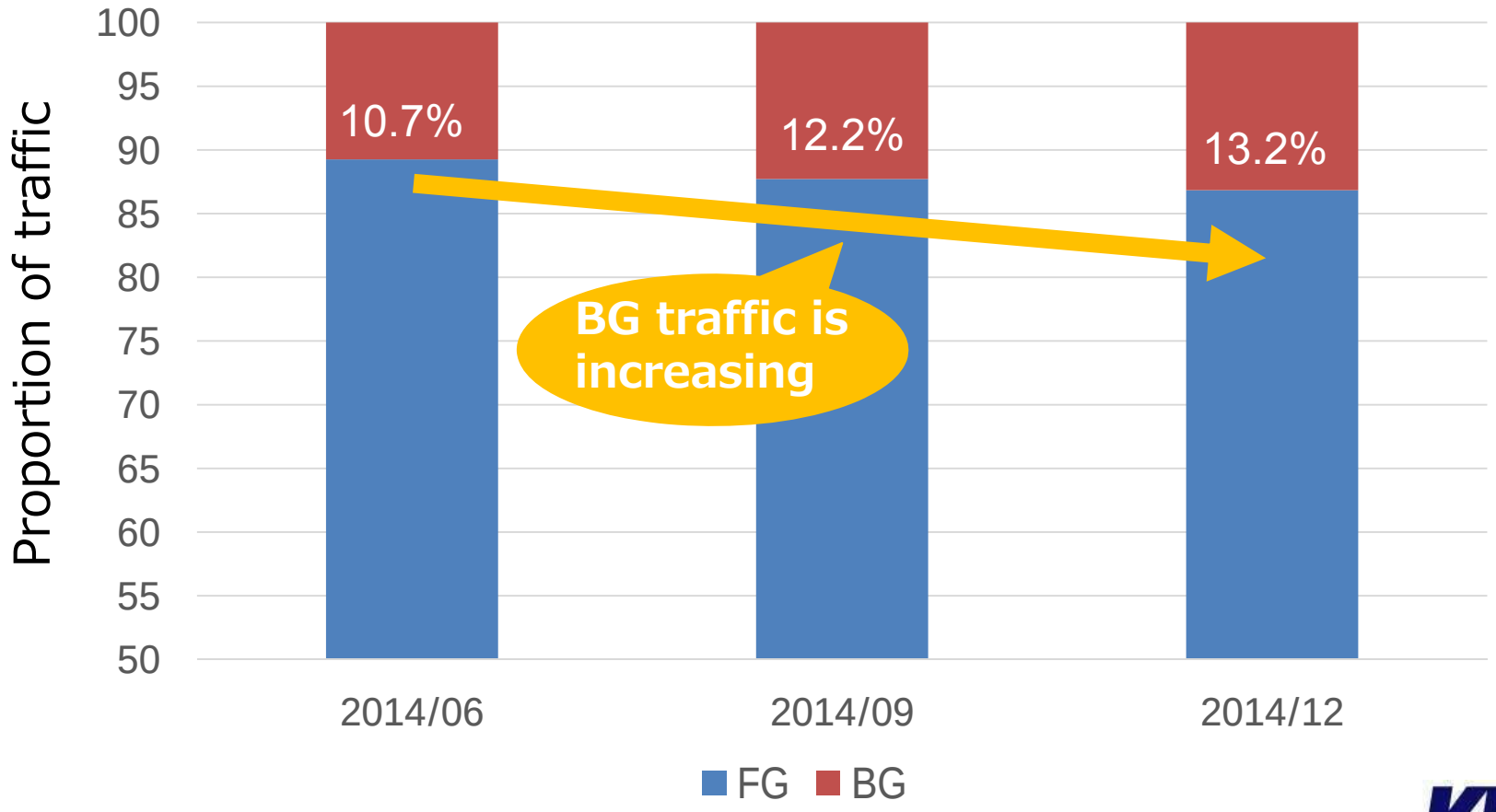
- Shortcut-index classifier measures until n-th set of packet



KDDI R&D LABS

Commercial data

- Applying shortcut classifier to commercial data



KDDI R&D LABS

Conclusion remarks

- Classification on mobile network
- High accuracy for experimental data set
  - Even using shortcut-index classifier
  - “Average packet size for uplink” and “CV of packet size for downlink” are especially effective for classification
- BG traffic is actually increasing in commercial LTE network

THANK YOU

masaki-suzuki@kddilabs.jp

KDDI R&D LABS