Growth Trends in Japanese Broadband Traffic

Kenjiro Cho (IIJ)
Kensuke Fukuda (NII)
Hiroshi Esaki (Univ. of Tokyo)
Akira Kato (Univ. of Tokyo)

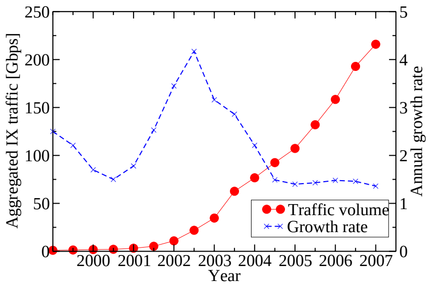
6th Asia Broadband Summit, Feb 26, 2007

about this talk

- extensive study on residentail broadband (RBB) traffic
  - aggregated traffic data from 7 Japanese ISPs
  - comparison of heavy-hitters/other-users, fiber/DSL users
- results show impact of RBB to Internet usage/backbone traffic
  - networking people should know
  - although each result may not be too surprising to experts

unprecedented traffic increase in backbone

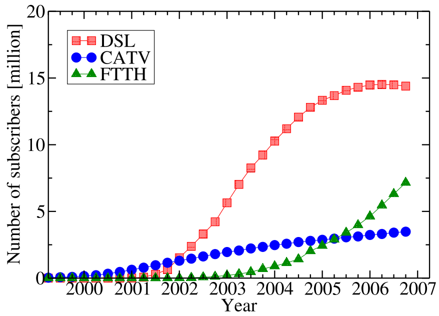
- rapidly growing residential broadband access
  - low-cost high-speed services, especially in Korea and Japan
  - Japan is by far the highest in Fiber-To-The-Home (FTTH)
- traffic growth of the peak rate at major Japanese IXes
  - still keeps growth of 50% per year
  - how much is contributed by residential broadband traffic?



Traffic growth at major Japanese IXes

residential broadband subscribers in Japan

- 25 million broadband subscribers as of September 2006
  - 14.4 million for DSL, 3.5 million for CATV, 7.2 million for FTTH
- exponential increase of FTTH, expected to exceed DSL in 2008
  - 100Mbps bi-directional fiber access costs 40USD/month
  - significant impact to backbones



RBB subscribers in Japan

motivation

- concerns about rapid growth of RBB traffic
  - backbone technologies will not keep up with RBB traffic
  - ISPs cannot invest in backbone simply for low-profit RBB
- ISPs and policy makers need to understand the effects of RBB
  - although most ISPs internally measure their traffic
    - data are seldom made available to others
    - measurement methods and policies differ from ISP to ISP

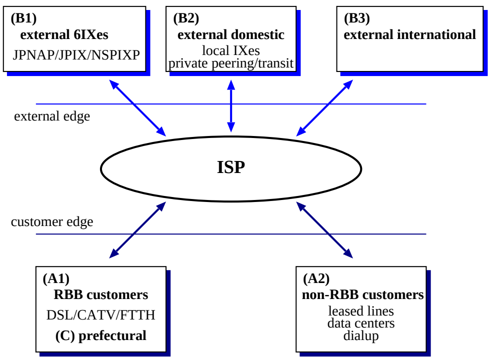
- to identify the macro-level impact of RBB traffic on ISP backbones
  - a study group with 7 major Japanese ISPs and government
- our approach consists of 2 analyses
  - aggregated traffic analysis
    - based on aggregated SNMP data from 7 major ISPs
  - per-customer traffic analysis
    - based on Sampled NetFlow data from one of the ISPs

major findings in aggregated traffic data

- our data is considered to cover 42% of total Japanese traffic
  - total RBB traffic in Japan is estimated to be 637Gbps (2006/11)
- 70% of RBB traffic is constant, peak in the evening hours
- p2p file-sharing was dominant in 2004
  - non-p2p video downloading has increased in 2006
- RBB traffic is much larger than office traffic, so backbone traffic is dominated by RBB traffic
- traffic volume exchanged via private peering is larger than volume exchanged via major IXes
- regional RBB traffic is roughly proportional to regional population

data collection across major ISPs

- focus on traffic crossing ISP boundaries (customer and external)
  - tools were developed to aggregate MRTG/RRDtool traffic logs
- only aggregated results published not to disclose individual ISP share
- challenges: mostly political or social, not technical



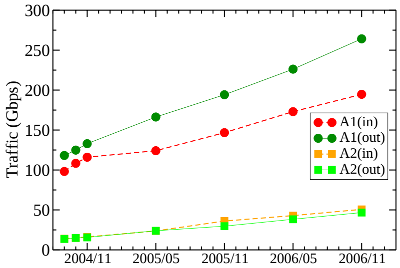
5 traffic groups at ISP cusomer and external boundaries

methodology for aggregated traffic analysis

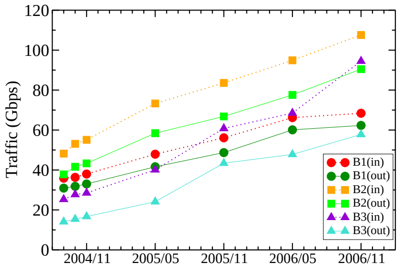
- month-long traffic logs for the 5 traffic groups with 2-hour resolution
  - MRTG's resolution for monthly log
- a script to read and aggregate a list of MRTG/RRDtool logs
  - each ISP creates log lists and makes aggreagated logs by themselves without disclosing details
- biggest workload for ISP
  - creating lists by classifying large number of per-interface logs
    - some ISPs have more than 100,000 logs!
  - maintaining the lists
    - frequent planned and unplanned configuration changes
- data sets
  - 2-hour resolution interface counter logs
    - from Sep/Oct/Nov 2004, May/Nov 2005, May/Nov 2006
    - by re-aggregating logs provided by 7 ISPs
- IN/OUT from ISPs' view

traffic growth

- 26-66% increase in 2006
  - RBB: 33% increase for inbound, 36% increase for outbound
- growth has slowed down from 100% in 2002 to 50% in 2005
  - observed worldwide



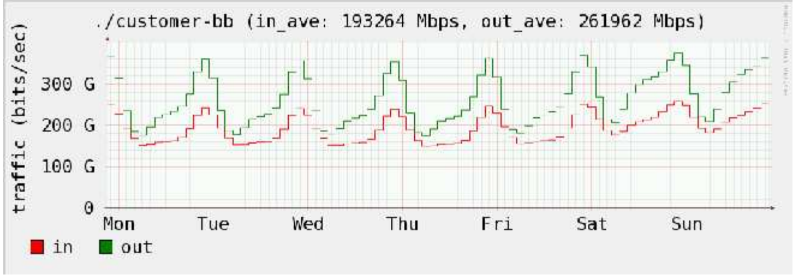
Growth of customer traffic



Growth of external traffic

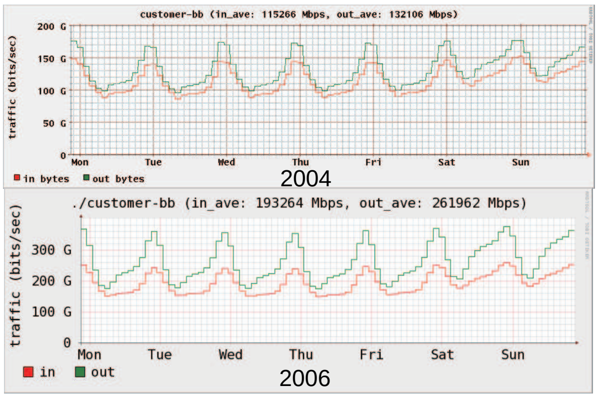
RBB customer weekly traffic
in November 2006

- DSL/CATV/FTTH customer traffic of the 7 ISPs
  - 200Gbps on average!
  - 150Gbps is constant, probably due to automated p2p applications
  - daily fluctuations: peak from 21:00 to 23:00



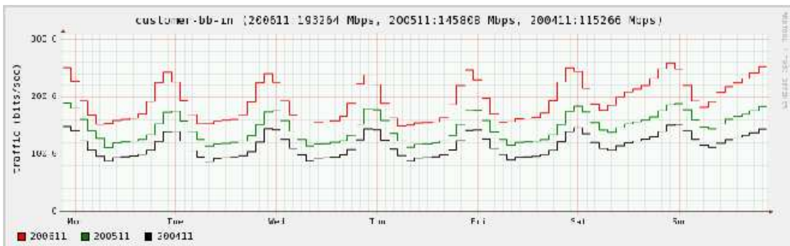
changes in RBB weekly traffic

- in 2004, inbound and outbound was almost equal
- in 2006, outbound (downloading to users) became larger



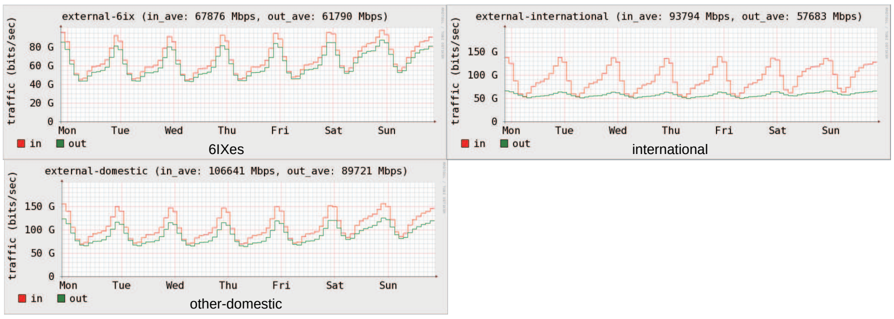
comparing RBB in-volumes among 2004, 2005 and 2006

- the growth came from the constant portion in 2005!
- both constatnt portion and daily fluctuations grew in 2006



weekly external traffic

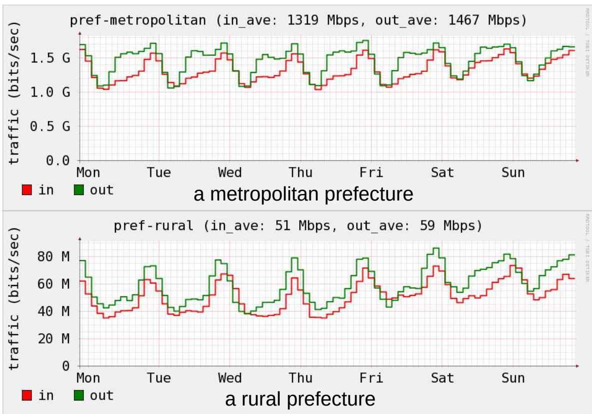
- external traffic is also strongly affected by RBB traffic
  - other-domestic: mainly private peering (also transit, regional IXes)
    - larger than traffic via majior IXes
  - international: inbound much larger than outbound
    - traditional content downloading seems still non-negligible



External weekly traffic in November 2006

prefectural differences in RBB traffic

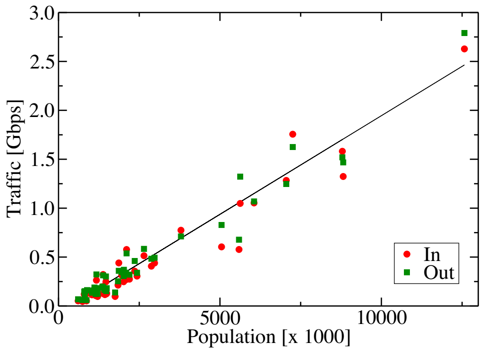
- similar temporal traffic pattern across different prefectures
  - e.g., peak in evening, 70% is constant, regardless the volume
  - metropolitan prefectures with larger office hour traffic



Example prefectural traffic

prefectural population and traffic

- traffic is roughly linear to population!
  - from a scatter plot of population and traffic volume
  - similar result with the number of Internet users
- no clear difference in usage or heavy-hitter ratio



Prefectural traffic volumes are roughly linear to populations

analysis of per-customer traffic in one ISP

- one ISP provided per-customer traffic data for Feb and Jul 2005

- data sets
  - Sampled NetFlow data
    - from edge routers accommodating fiber/DSL RBB customers
  - week-long logs from Feb and Jul 2005

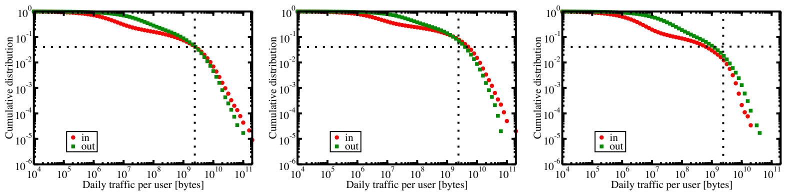
- heavy-hitters: denote users who upload more than 2.5GB/day
  - larger in fiber users

major findings in per-customer traffic data

- 4% of heavy-hitters account for 75% of the total inbound volume
- the fiber users account for 86% of the inbound volume
  - DSL is only 14%
  - even though the number of DSL active users is larger than fiber
- the distribution of heavy-hitters is heavy-tailed
  - no clear boundary between heavy-hitters and normal users
- dominant applications have poor locality and communicate with a wide range and number of peers

CCDF of daily traffic per user

- heavy-hitters are statistically distributed
  - over a wide range of traffic volume (heavy-tailed)
    - even up to 200GB/day (19Mbps)!
  - no clear boundary between heavy-hitters and normal users
- lines at 2.5GB/day (230kbps) and the top 4% heavy-hitters
  - knee of the total users's slope
- heavy-hitter population: 4% in total users, 10% in fiber, 2% in DSL

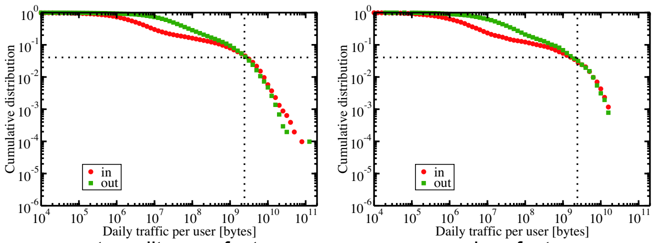


total fiber DSL

CCDF of daily traffic volume per user

prefectural comparison

- distribution similar in all prefectures
  - differences in tail length (population size)
    - probably due to universal broadband access in Japan

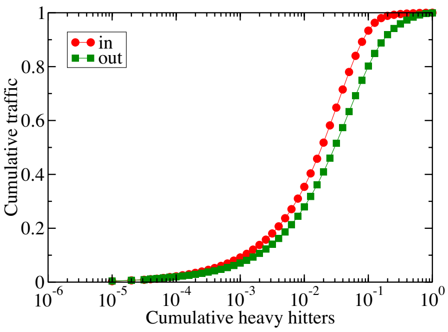


a metropolitan prefecture   a rural prefecture

CCDF of daily traffic volume per user

CDF of traffic volume of heavy-hitters

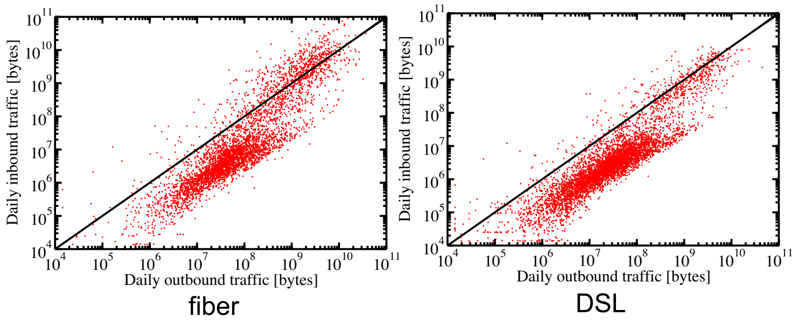
- graph: the top N% of heavy-hitters use X% of the total traffic
- highly skewed distribution in traffic usage
  - the top 4% use 75% of the total inbound traffic
  - the top 4% use 60% of the total outbound traffic



Cumulative Traffic of heavyhitters

correlation of inbound/outbound volumes per user

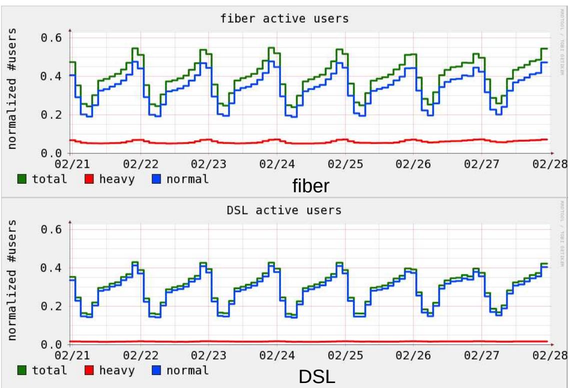
- 2 clusters: one below the unity line, another in high volume region
  - more heavy-hitters in fiber, more lightweight users in DSL
- no qualitative difference between fiber users and DSL users
  - except the percentage of heavy-hitters
- again, no clear boundary between heavy-hitters and normal users



daily per-user inbound/outbound volumes

number of active users

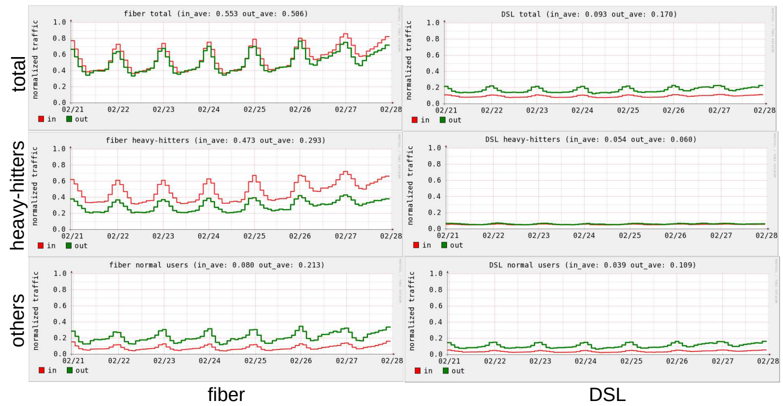
- numbers are normalized to the fiber/DSL combined peak
- total numbers are similar between fiber and DSL
- heavy-hitters are fairly constant, especially in DSL



Normalized number of active users

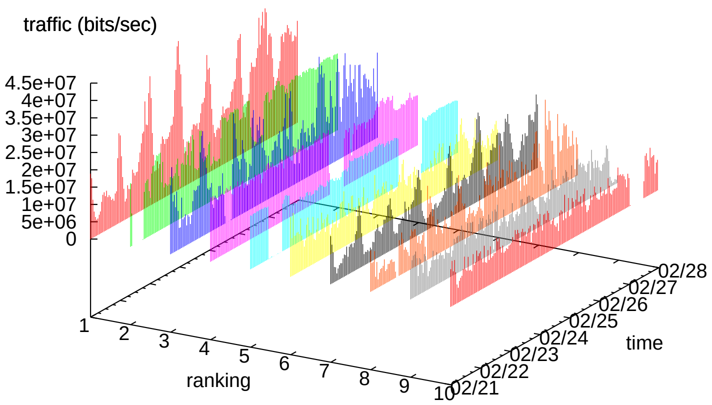
comparison of fiber/DSL traffic

- again, normalized to the combined peak
- inbound: 86% is from fiber users, DSL is only 14%
- total traffic is heavily influenced by fiber heavy-hitters



uploading behavior of top 10 heavy-hitters

- one hour average traffic over a week
  - considerable variations, suggesting differences in usage



protocols/ports ranking

- port 80 (http) is only 9%
- 83% is TCP dynamic ports!
  - each port usage is small except port 80

protocol	port	name	(%)	port	name	(%)
TCP	*		**97.43**			
	(< 1024		13.99)	81	-	0.15
	80	http	9.32	25	smtp	0.14
	20	ftp-data	0.93	119	nntp	0.13
	554	rtsp	0.38	21	ftp	0.11
	443	https	0.30	22	ssh	0.09
	110	pop3	0.17		others	2.27
	(>= 1024		83.44)	1935	macromedia-fsc	0.20
	6699	winmx	1.40	1755	ms-streaming	0.20
	6346	gnutella	0.92	2265	-	0.13
	7743	winny	0.48	1234	-	0.12
	6881	bittorrent	0.25	4662	edonkey	0.12
	6348	gnutella	0.21		others	79.41
UDP	*		**1.38**	6257	winmx-	0.06
	6346	gnutella	0.39		others	0.93
ESP			**1.09**			
GRE			**0.07**			
ICMP			**0.01**			
others			**0.02**			

geographic traffic matrix of RBB traffic

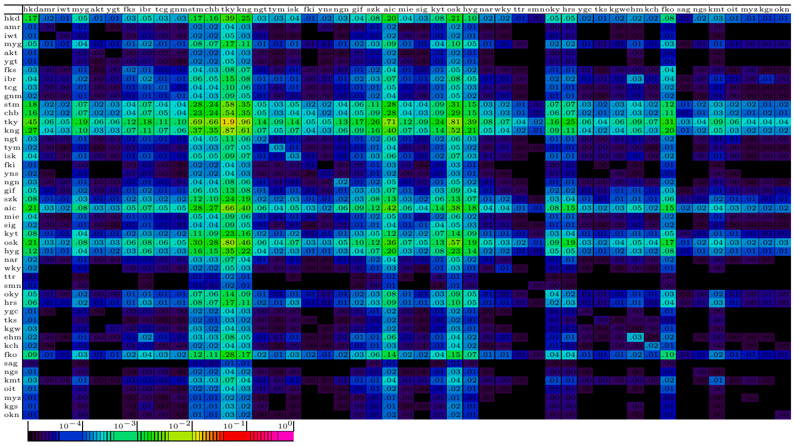
- RBB (home users), DOM (other domestic), INTL (international)
  - both ends are classified by commercial geo-IP databases
- 62% of residential traffic is user-to-user
- 90% is inside Japan (among RBB and DOM)
  - possible reasons are:
    - language and cultural barriers
    - p2p super-nodes among bandwidth-rich domestic fiber users

$src \backslash dst$	ALL	RBB	DOM	INTL
ALL	100.0	84.8	11.1	4.1
RBB	77.0	62.2	9.8	3.9
DOM	18.0	16.7	1.1	0.2
INTL	5.0	4.8	0.2	0.0

prefectural traffic matrix

(src on Y-axis, dst on X-axis)

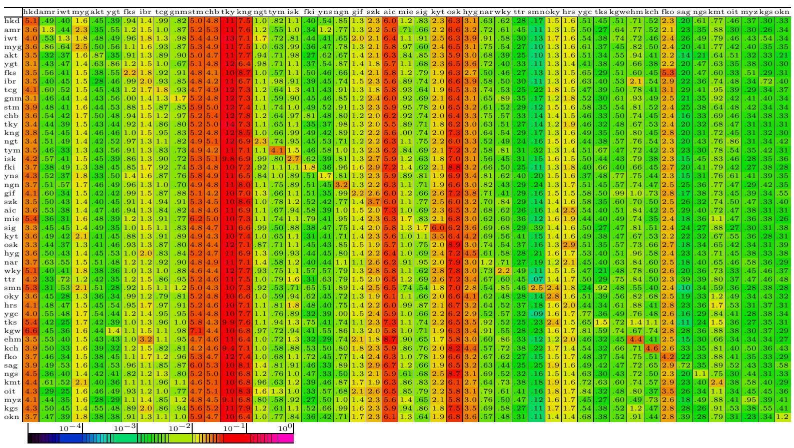
- looking into 47 prefectures
  - traffic volumes are roughly linear to prefectural populations



27

prefectural traffic matrix normalized to src

- the sum of columns is 100% for each row
- no clear difference among prefectures
  - similar distribution, only small locality (1-3%) is found
  - similar result when normalized to dst



28

implications

- we tend to attribute the skews in usage to the divide between a handful of heavy-hitters and the rest of the users
  - but there are diverse and widespread heavy-hitters
- heavy-hitters are no longer exceptional extremes
  - too many of them, statistically distributed over a wide range
    - casual users start playing with p2p applications, become heavy-hitters, and eventually shift from DSL to fiber
    - or, sometimes users subscribe to fiber first, and then, look for applications to use the abundant bandwidth
  - these users' behavior would be easily affected by social, economic or political factors (they don't care about underlying technologies)
    - in fact, a shift from p2p file-sharing to video downloading has been observed
  - but surely users as a whole are shifting towards high-volume usage
- is this specific to Japan?
  - a model of widespread symmetric residential broadband access
    - with language/cultural barriers, geographic concentration

conclusion

- we need to prepare for the future to accommodate innovations brought by empowered end-users
- our study to understand residential broadband traffic
  - cooperation with major ISPs and government
  - detailed analysis of traffic data from one ISP
- RBB traffic accounts for 2/3 of ISP backbone traffic
  - a significant impact on pricing and cost structures of ISP business
- future work
  - we will continue collecting aggregated traffic logs from ISPs
  - plans to compare results with other Japanese ISPs, other countries

- acknowledgments
  - IIJ, Japan Telecom, K-Opticom, KDDI, NTT Communications, POWEREDCOM, SOFTBANK BB for data collection support
  - Ministry of Internal Affairs and Communications for coordination