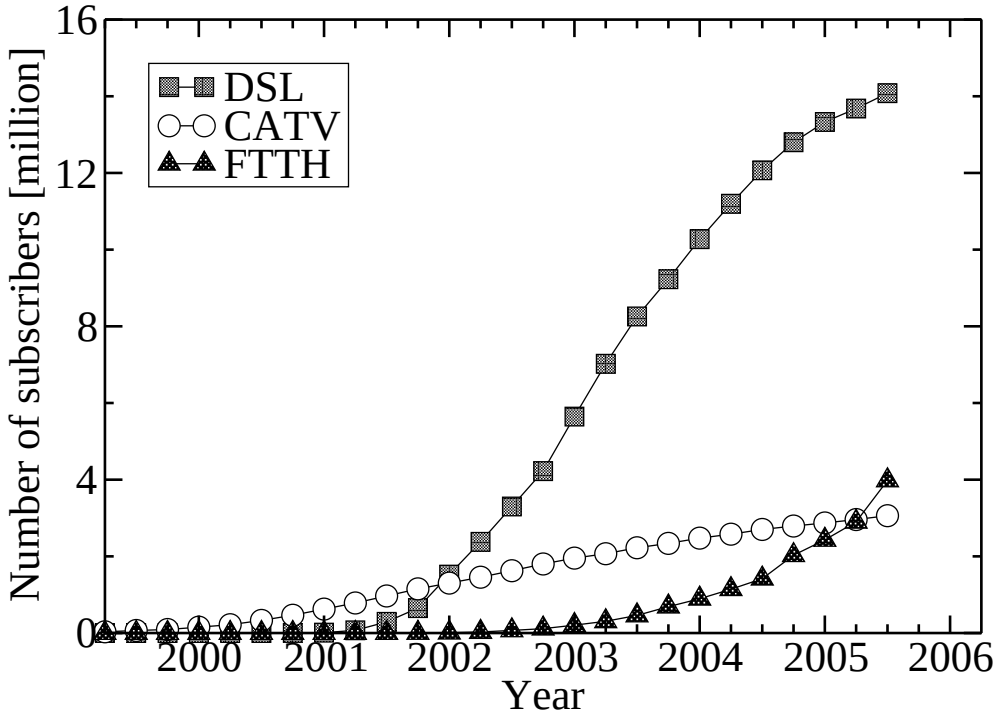
The impact of fiber access to ISP backbones in .jp

Kenjiro Cho (IIJ / WIDE)

residential broadband subscribers in Japan

- 21 million broadband subscribers as of September 2005
  - 14 million for DSL, 3 million for CATV, 4 million for FTTH
- exponential increase of FTTH
  - 100Mbps bi-directional fiber access costs 40USD/month
  - significant impact to backbones



findings

- 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 follows power law
  - no clear boundary between heavy-hitters and normal users
- dominant applications have poor locality and communicate with a wide range and number of peers

data sets

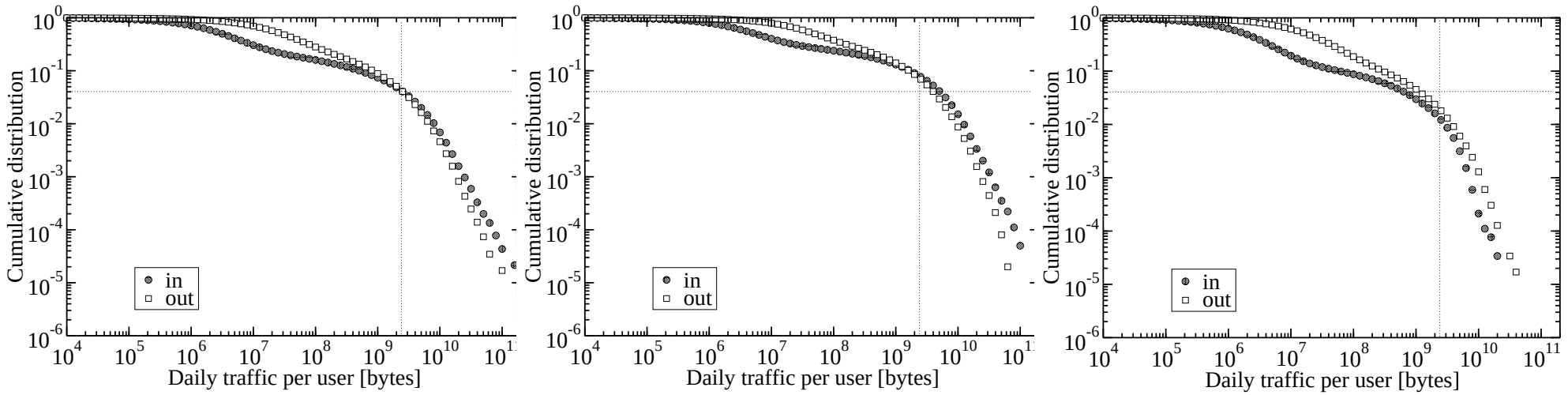
- Sampled NetFlow data from one Japanese ISP
  - edge routers accommodating fiber/DSL residential customers
  - week-long logs from February and July 2005

- ratio of fiber and DSL unique users in the data set
  - heavy-hitters: denote users who upload more than 2.5GB/day
    - larger in fiber users

	ratio (%)	$\geq 2.5GB/day$ (%)	$< 2.5GB/day$ (%)
total	100	4.46	95.54
fiber	46.4	3.66	42.79
DSL	53.6	0.80	52.75

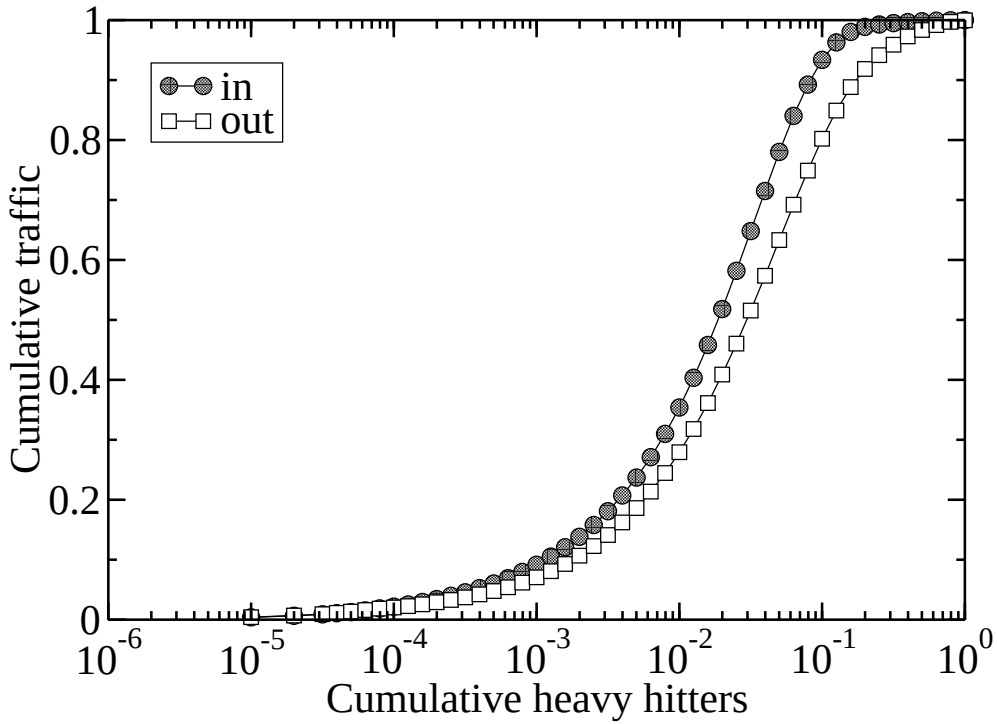
cumulative distribution of daily traffic per user
total users (left), fiber users (middle), DSL users (right)

- heavy-hitters are statistically distributed
  - over a wide range of traffic volume (follows power law)
    - 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



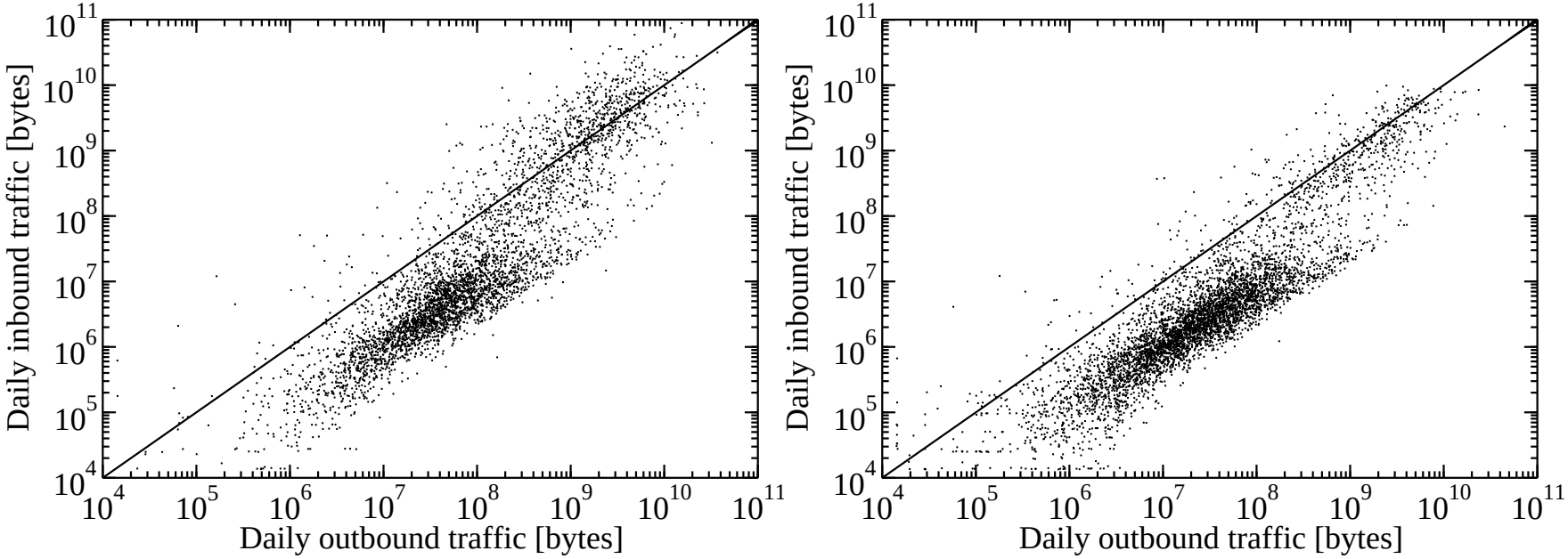
CDF of traffic volume of heavy-hitters
in decreasing order of volume

- 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



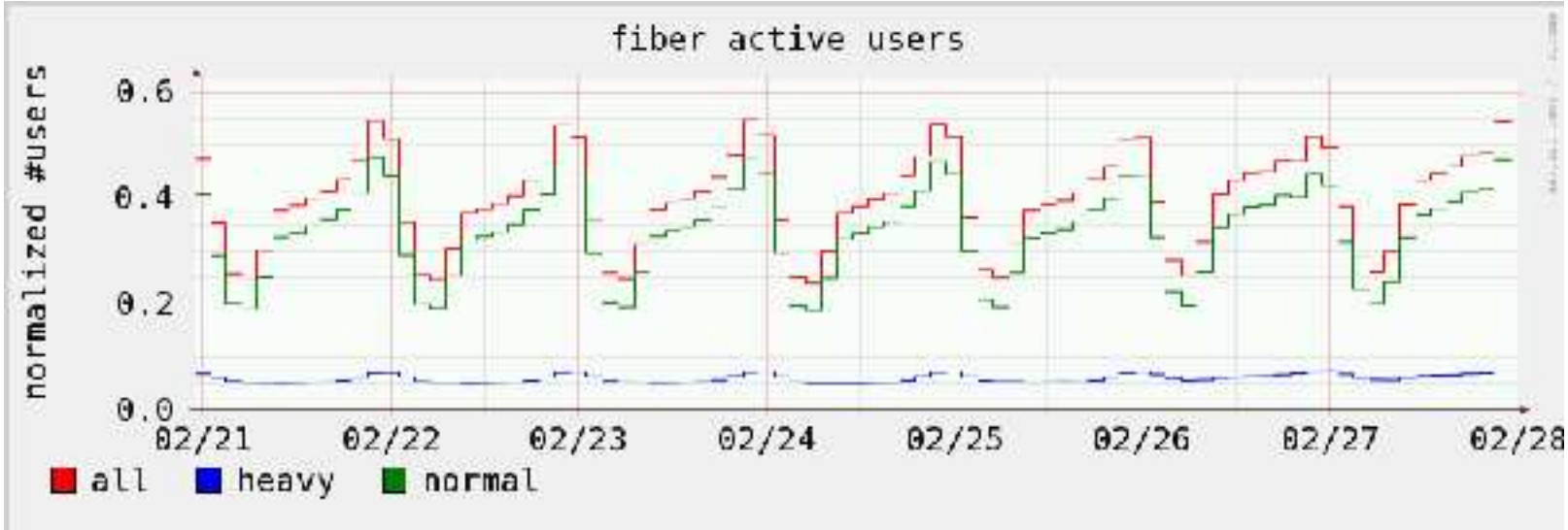
correlation of inbound and outbound volumes per user
fiber (left) and DSL (right)

- 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



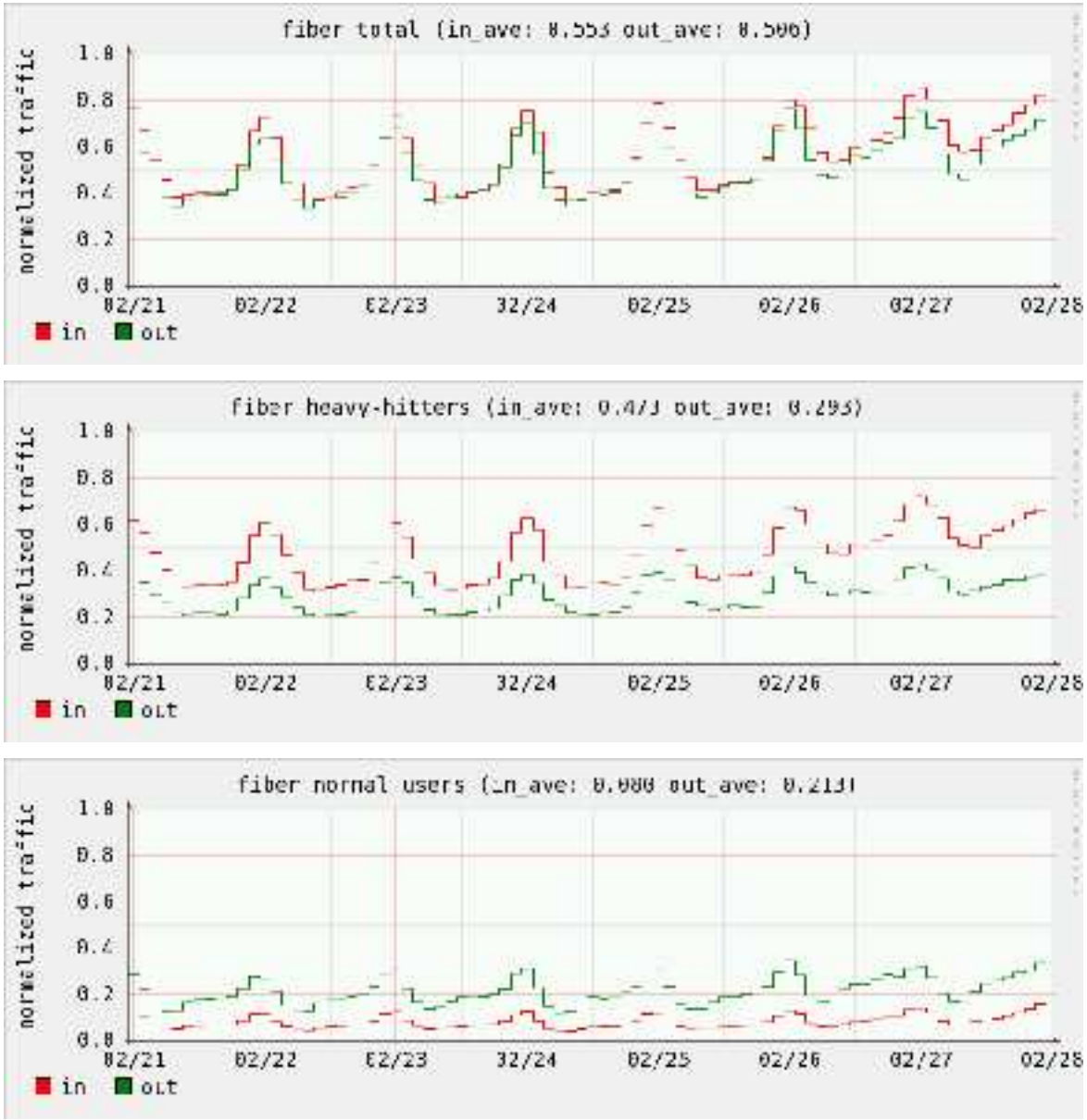
number of active users:
(normalized to the fiber/DSL combined peak)

- fairly constant for heavy-hitters, especially in DSL



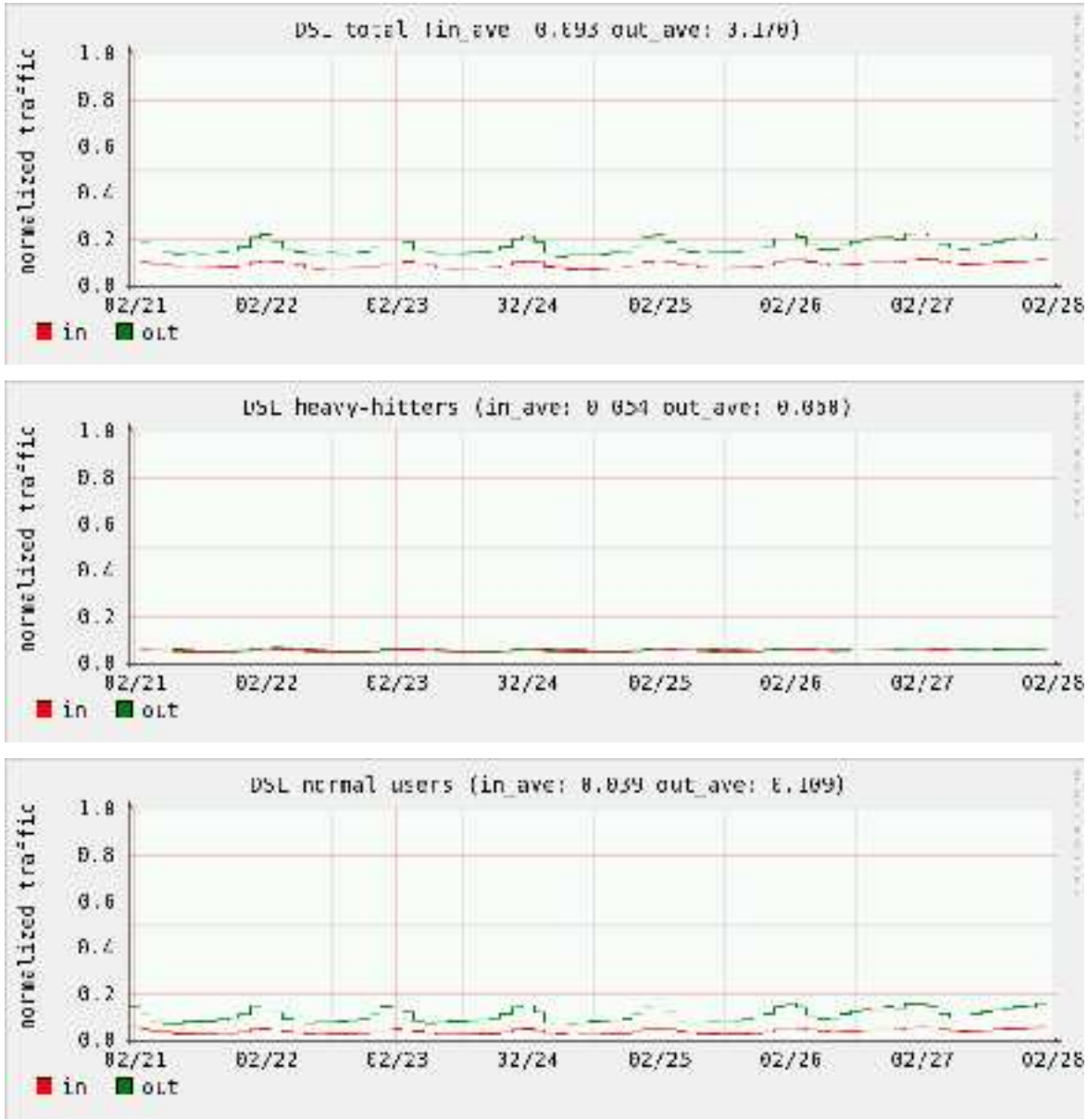
fiber users weekly traffic (normalized to the combined peak)

total users (top), heavy-hitters (middle), normal users (bottom)



DSL users weeky traffic (normalized to the combined peak)

total users (top), heavy-hitters (middle), normal users (bottom)



protocols/ports ranking

- 83% is TCP dynamic ports

protocol	ratio(%)	port #	name	ratio(%)
TCP	97.43			
($port < 1024$	13.99)	80	http	9.32
		20	ftp-data	0.93
		554	rtsp	0.38
		443	https	0.30
		110	pop3	0.17
		81	-	0.15
		25	smtp	0.14
		119	nntp	0.13
		21	ftp	0.11
		22	ssh	0.09
		-	other	2.27
($port >= 1024$	83.44)	6699	winmx	1.40
		6346	gnutella	0.92
		7743	winny	0.48
		6881	bittorrent	0.25
		6348	gnutella	0.21
		1935	macromedia-fsc	0.20
		1755	ms-streaming	0.20
		2265	-	0.13
		1234	-	0.12
		4662	edonkey	0.12
		8080	http-proxy	0.11
		-	other	79.30
UDP	1.38	6346	gnutella	0.39
		6257	winmx-	0.06
		-	other	0.93
ESP	1.09			
GRE	0.07			
ICMP	0.01			
OTHER	0.02			

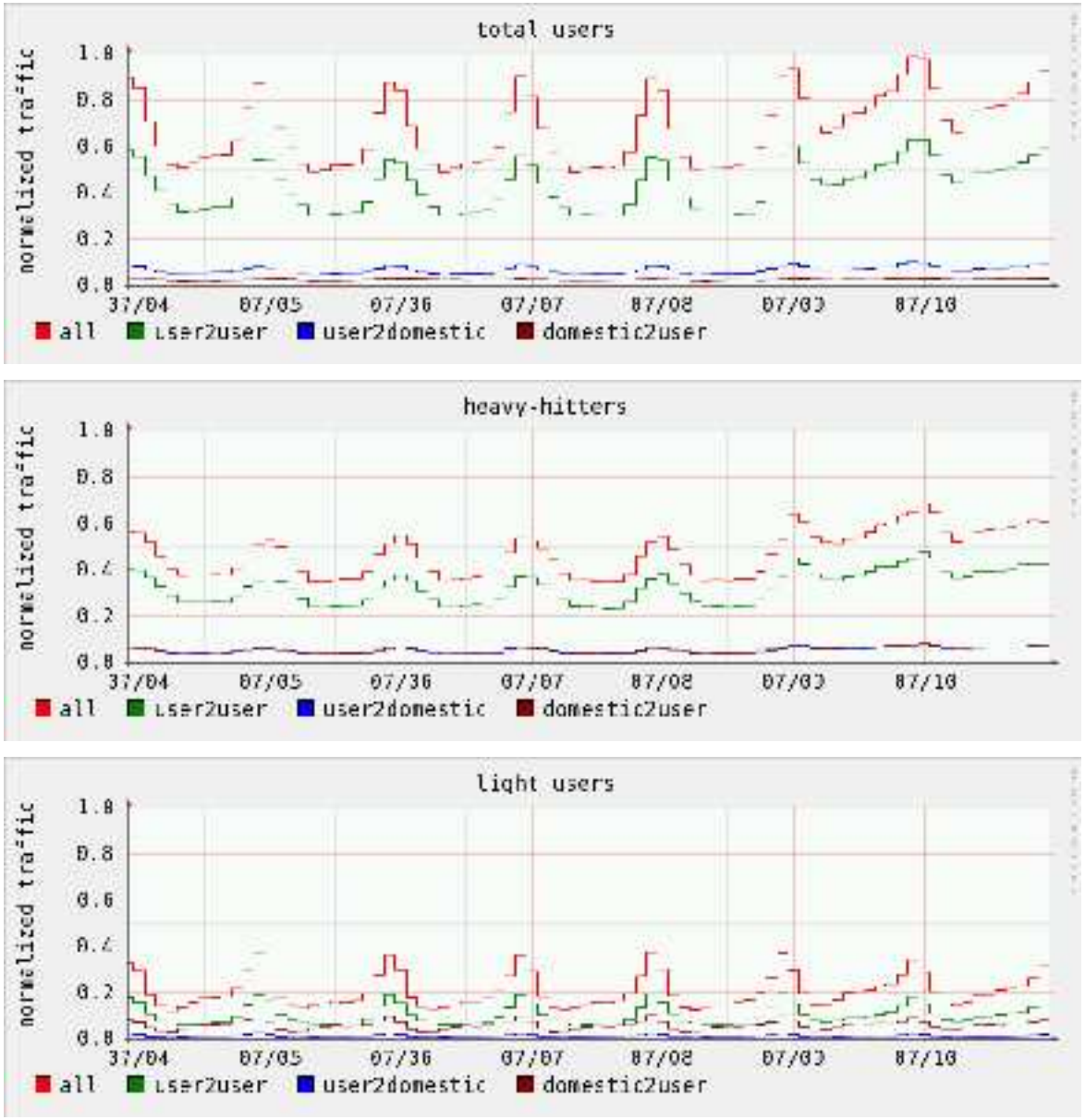
geographic traffic matrix

- 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

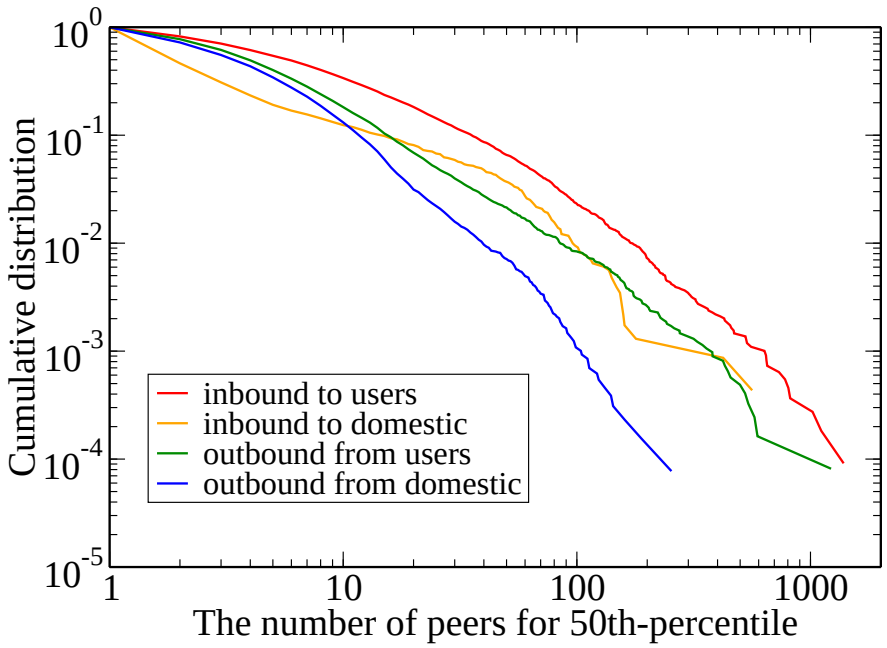
weekly traffic by traffic matrix

total users (top), heavy-hitters (middle), normal users (bottom)



CDF of number of peers for 50th-percentile traffic

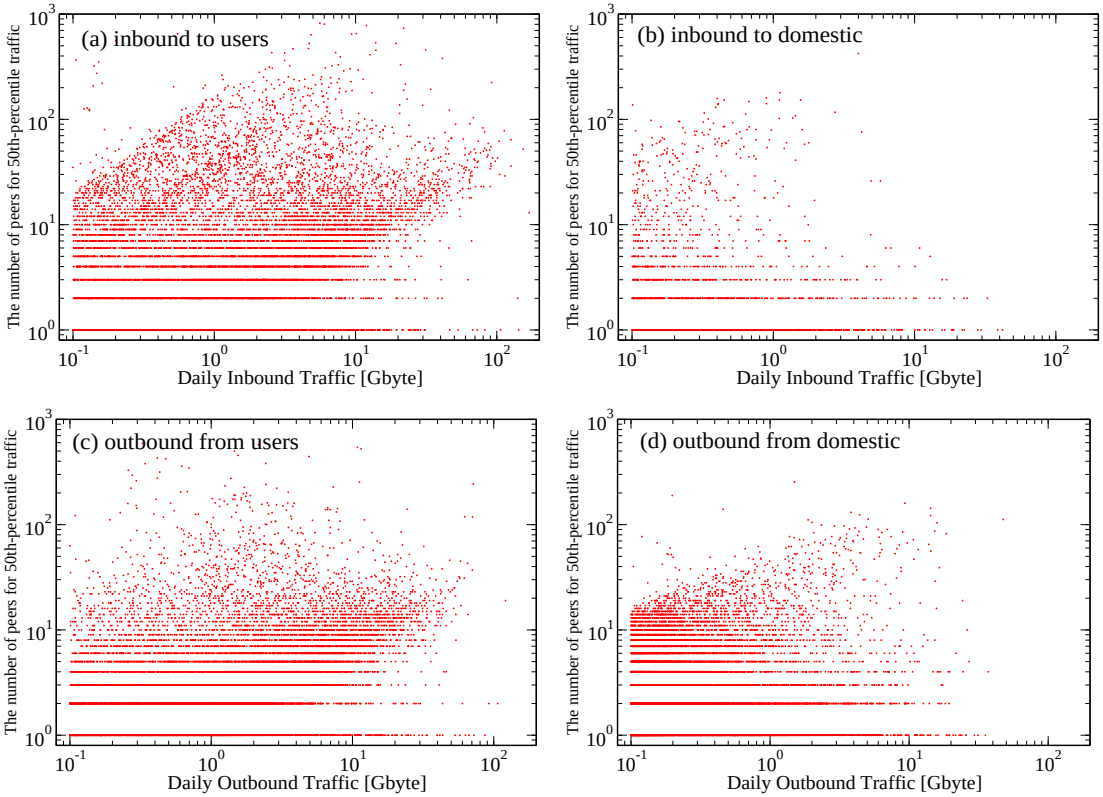
- count the number of peers for each user to distinguish app types
  - to exclude long tail, peers are sorted inversely and counted for exceeding 50th-percentile traffic
- we expected 2 application types:
  - a few peers for video-streaming/downloading from servers
  - a large number of peers for p2p file-sharing
- however, we can't distinguish them by peer numbers
  - users use both of them with different ratio?



correlation of traffic volume and number of peers

inbound to users (top-left), to domestic (top-right), outbound from users (bottom-left) and from domestic (bottom-right)

- positive correlation: peer numbers proportional to traffic volume
- high volume users with few peers: do not follow the correlation
  - other than file sharing?



implications

- we tend to attribute the skews 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
  - more natural to think
    - 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
- is this specific to japan?
  - other countries will take some time to deploy fiber access
  - a model of widespread symmetric residential broadband access
- need to reevaluate pricing and cost structures of ISP industry