Observing Slow Crustal Movement in Residential User Traffic

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

August 16 2008

motivation

many media coverage on explosive traffic growth by video content

- YouTube is just the beginning[Cisco2008b]

but technical sources report only modest traffic growth worldwide

- MINTS: 50-60% in U.S. and worldwide
- Cisco visual networking index: worldwide growth of 50% per year over last few years

why is traffic growth important?

- one of the key factors driving research, development and investiment in technologies and infrastructures
- with annual growth of 100%, it grows 1000-fold in 10 years
- with annual growth of 50%, it grows 58-fold in 10 years

key question: what is the macro level impact of video and other rich media content on traffic growth at the moment?

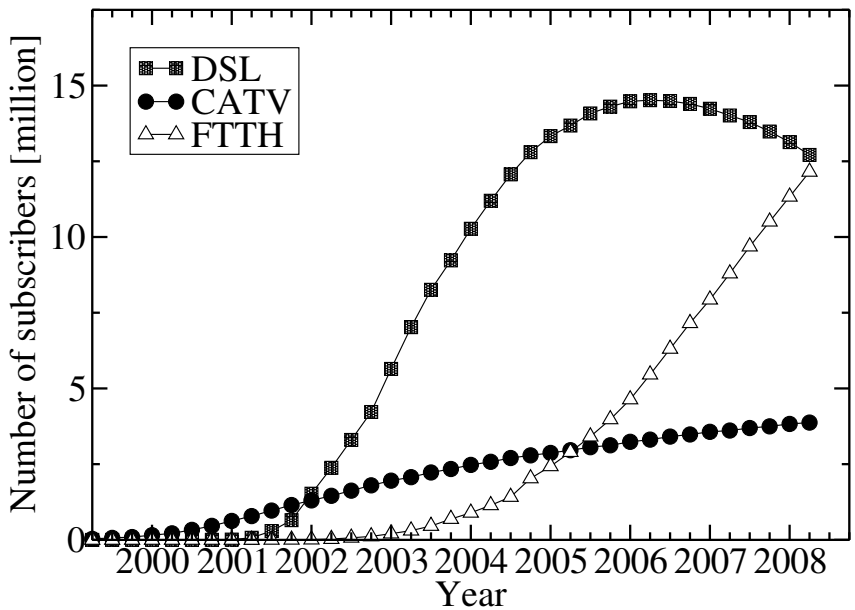
residential broadband subscribers in Japan

28.7 million broadband subscribers as of March 2008

- DSL:12.7 million, FTTH:12.2 million, CATV:3.3 million

shift from DSL to FTTH: about to exceed DSL

- 100Mbps bi-directional fiber access costs 40USD/month
- significant impact to backbones



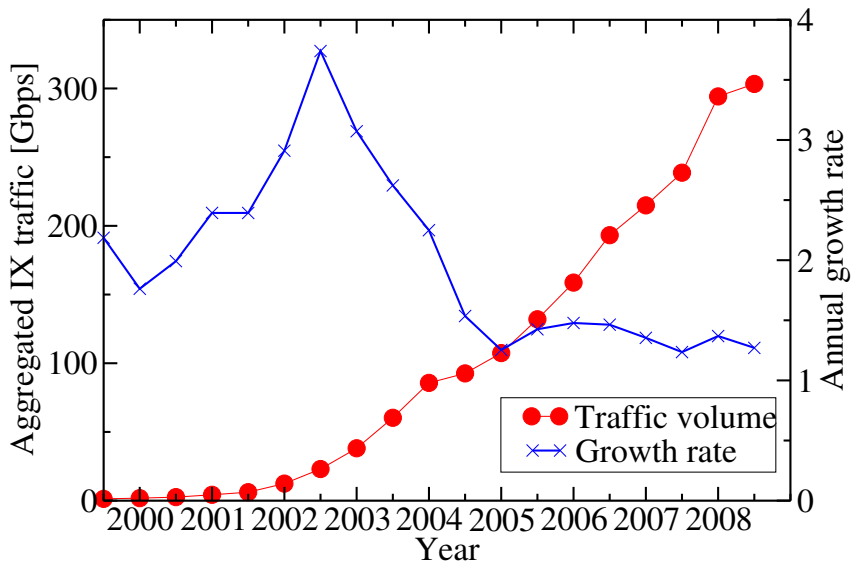
traffic growth in backbone

rapidly growing residential broadband access

- low-cost high-speed services, especially in Korea and Japan
- Japan is the highest in Fiber-To-The-Home (FTTH)

traffic growth of the peak rate at major Japanese IXes

- modest growth of about 40% per year since 2005



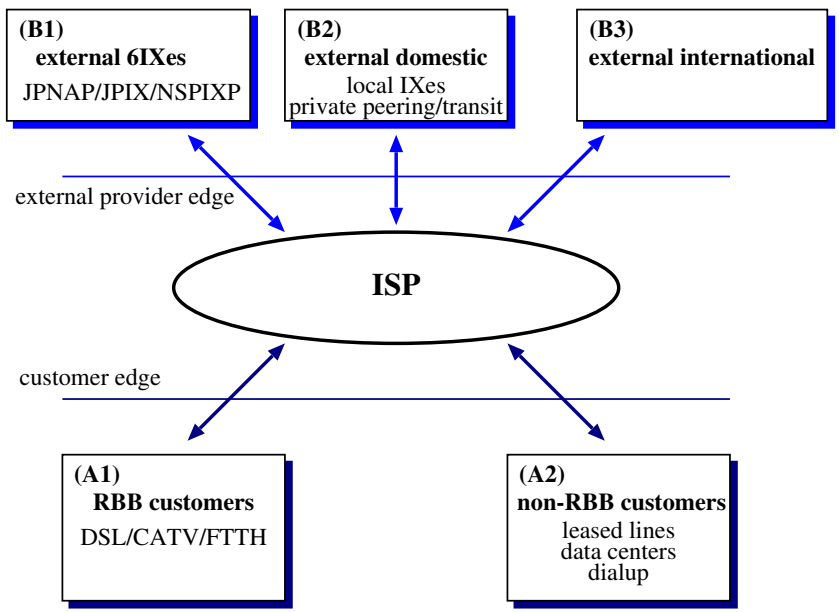
data collection across major ISPs

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

- 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

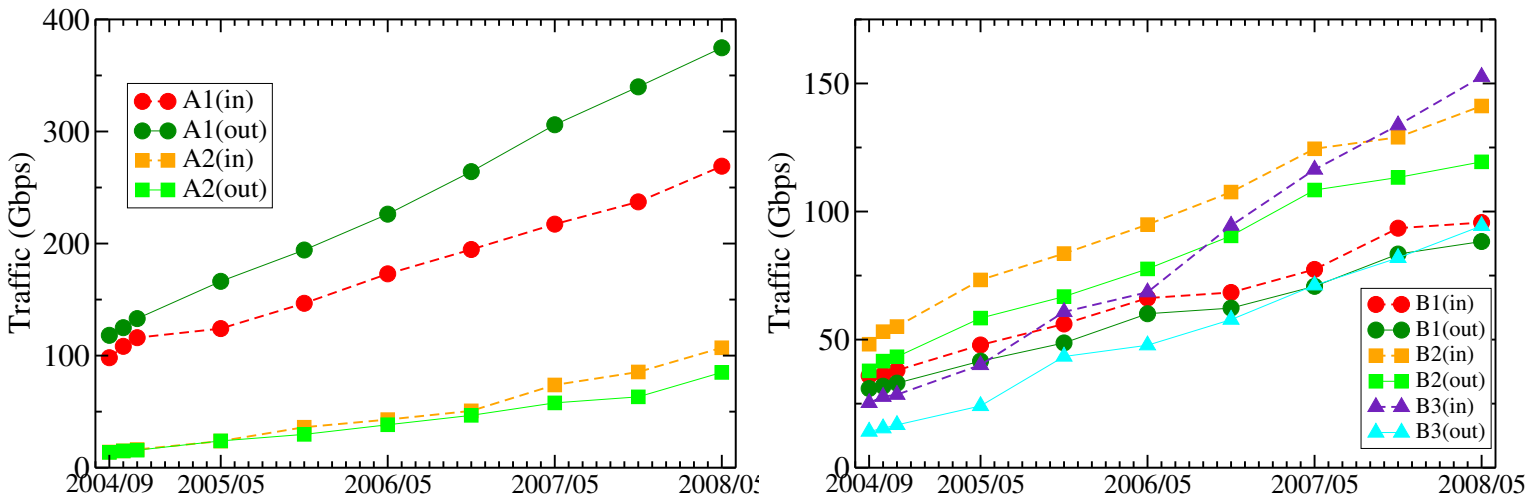
- 2-hour resolution interface counter logs
  - from Sep/Oct/Nov 2004, May/Nov 2005-2008
  - by re-aggregating logs provided by 7 ISPs

IN/OUT from ISPs' view

traffic growth

22-68% increase in 2007

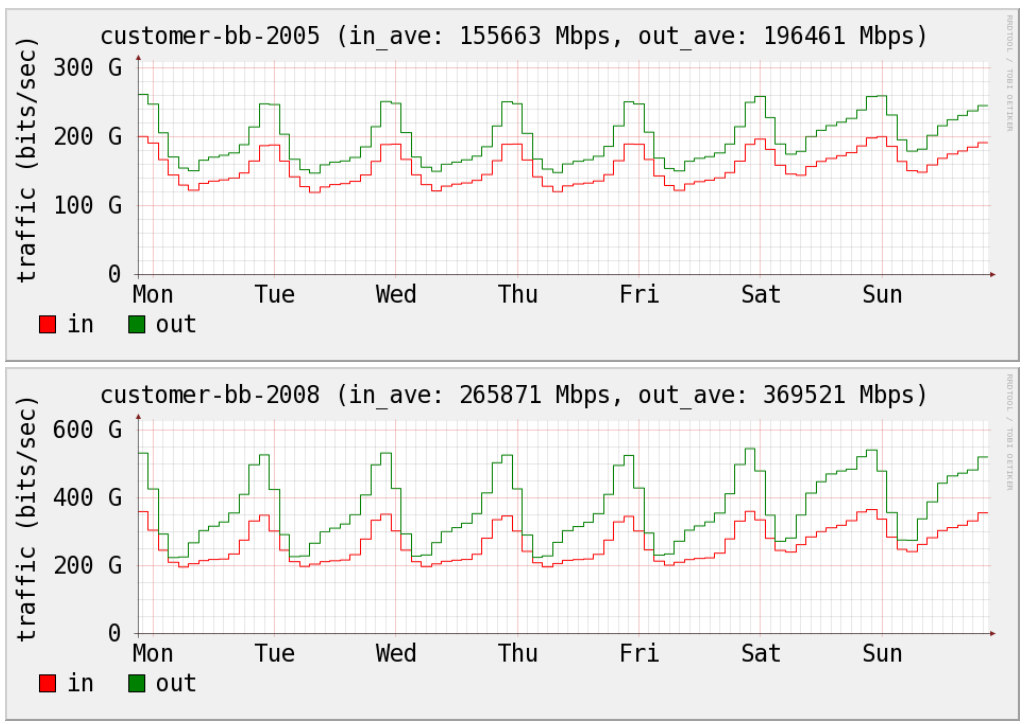
- RBB: 22% increase for inbound, 29% increase for outbound
- a sharp increase in international inbound due to popular video services



changes in RBB weekly traffic

in 2004, inbound and outbound was almost equal
in 2008, outbound (downloading to users) became larger
both constatnt portion and daily fluctuations grew in 2008

- implies a shift from p2p to video (e.g, YouTube)



analysis of per-customer traffic in one ISP

one ISP provided per-customer traffic data

- Sampled NetFlow data
  - from edge routers accommodating fiber/DSL RBB customers
- week-long data from Apr 2004, Feb 2005, Jul 2007, Jun 2008
  - Feb 2005 and Jun 2008, before and after the advent of YouTube

ratio of fiber/DSL active users and total traffic volumes

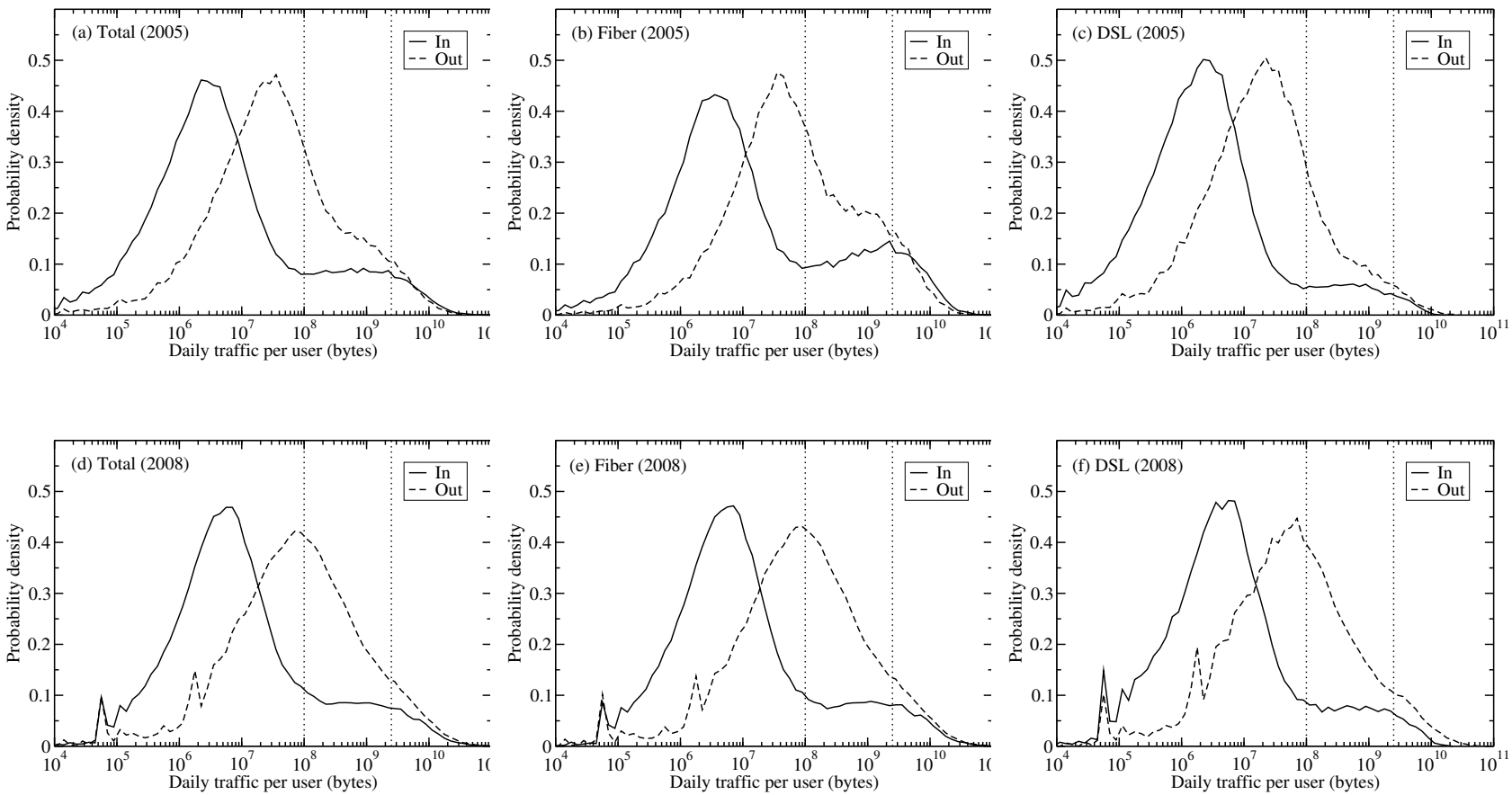
- in 2008, 80% of active users are fiber users, consuming 90% of traffic

		active users (%)	total volume (%)
2005	fiber	46	79
	DSL	54	21
2008	fiber	79	87
	DSL	21	13

PDF of daily traffic per user

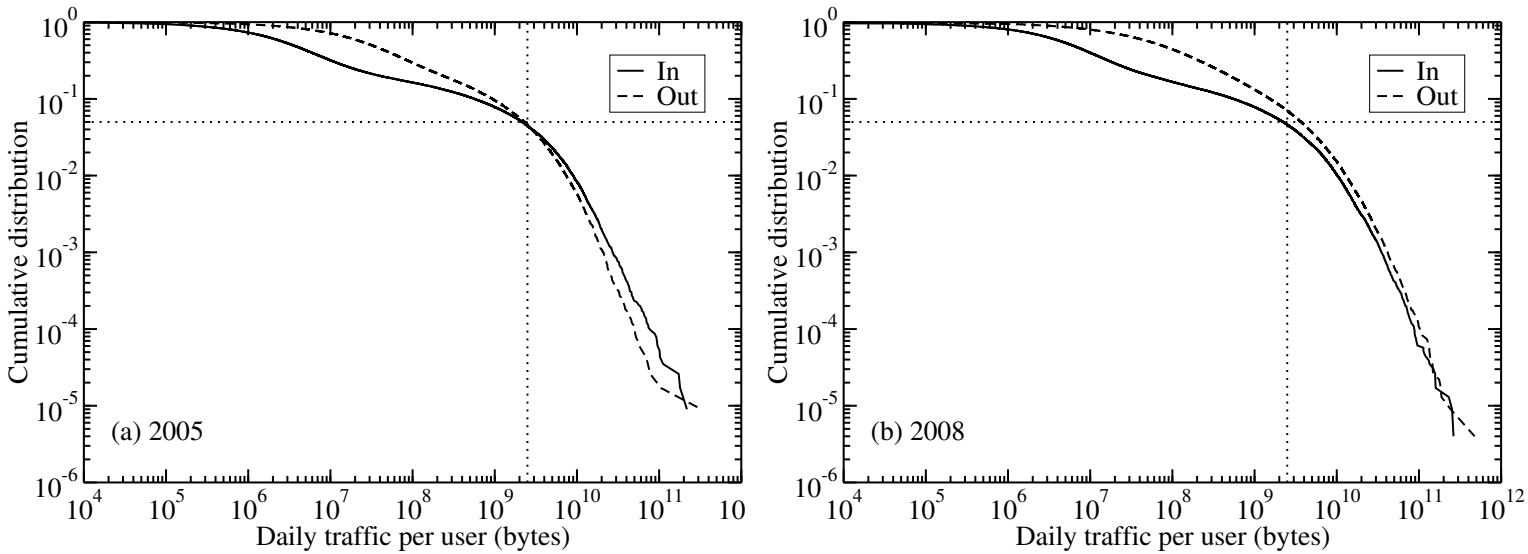
2 lognormal distributions: asymmetric, symmetric high-volume

- high-volume dist: not growing much
- total(left) fiber(middle) DSL(right) in 05(top),08(bottom)
- mode: 3.5MB,32MB/day(2005), 5MB,94MB/day(2008)



CCDF of daily traffic per user

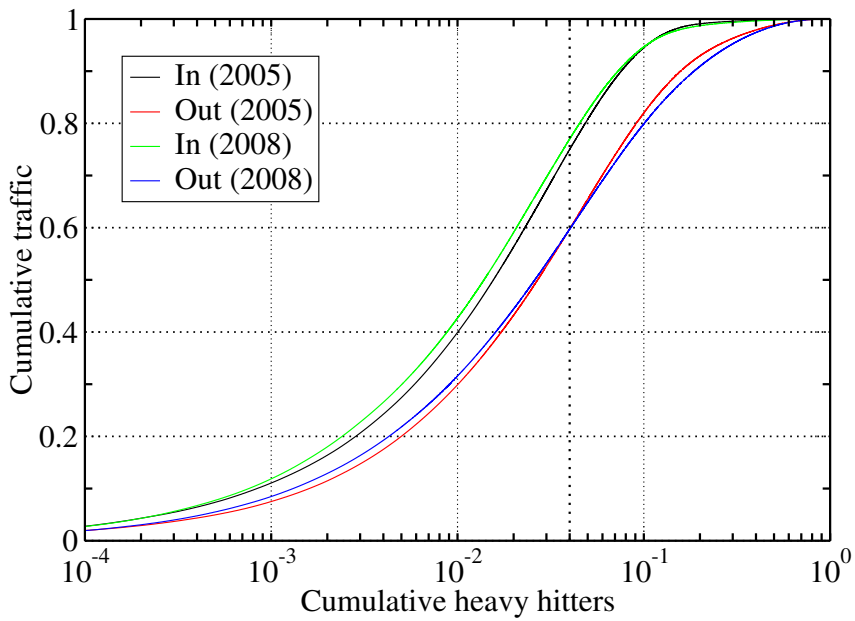
only outbound (download for users) increased



CDF of traffic volume consumed by top heavy-hitters

graph: the top N% of heavy-hitters use X% of the total traffic
highly skewed distribution in traffic usage
no noticeable change from 2005 to 2008

- probably because client-type users also have long-tailed distributions

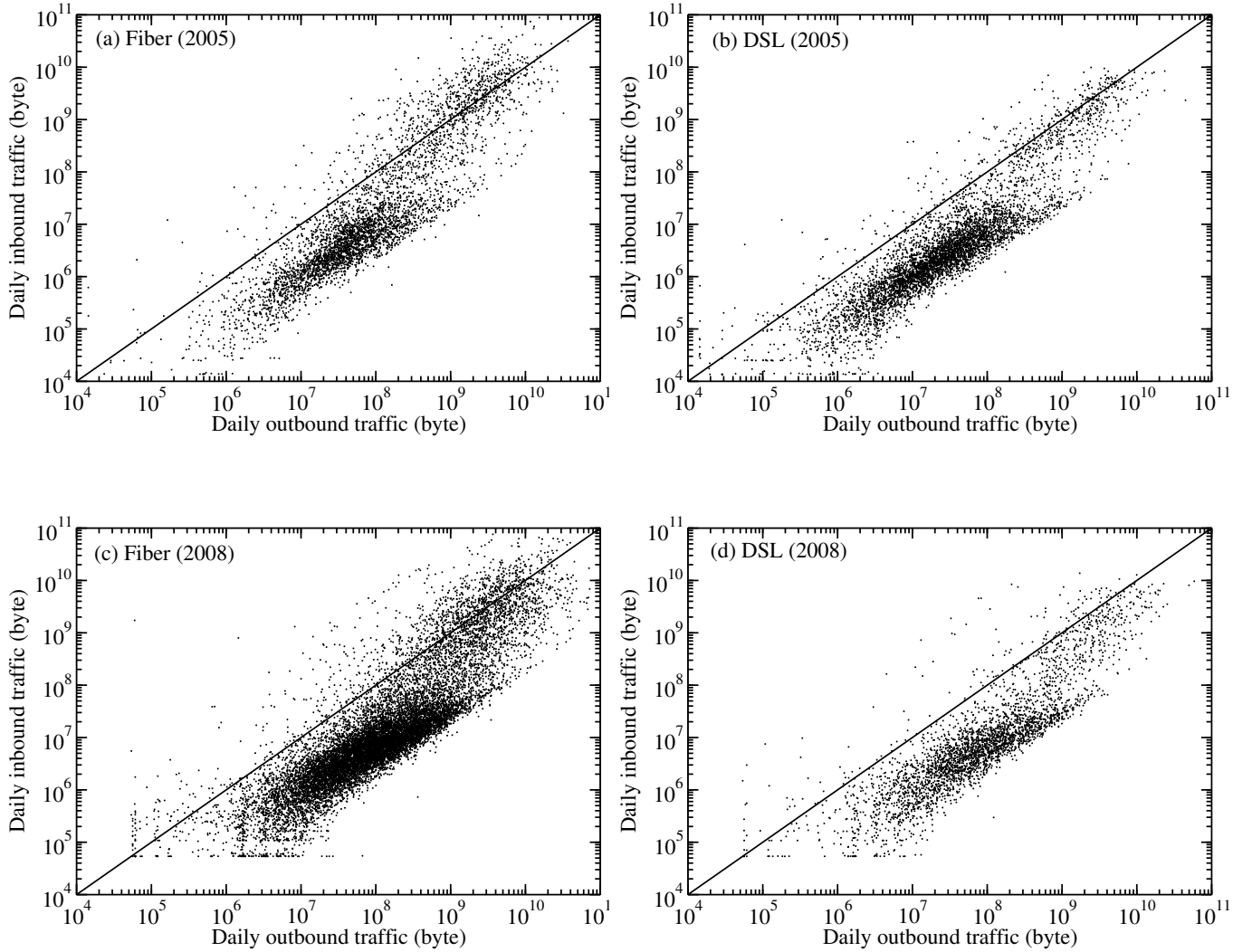


correlation of inbound/outbound volumes per user

fiber (left) and DSL (right) in 2005 (top) and 2008 (bottom)

2 clusters: one below the unity line, another in high volume region

no clear boundary: heavy-hitters/others, client-type/peer-type



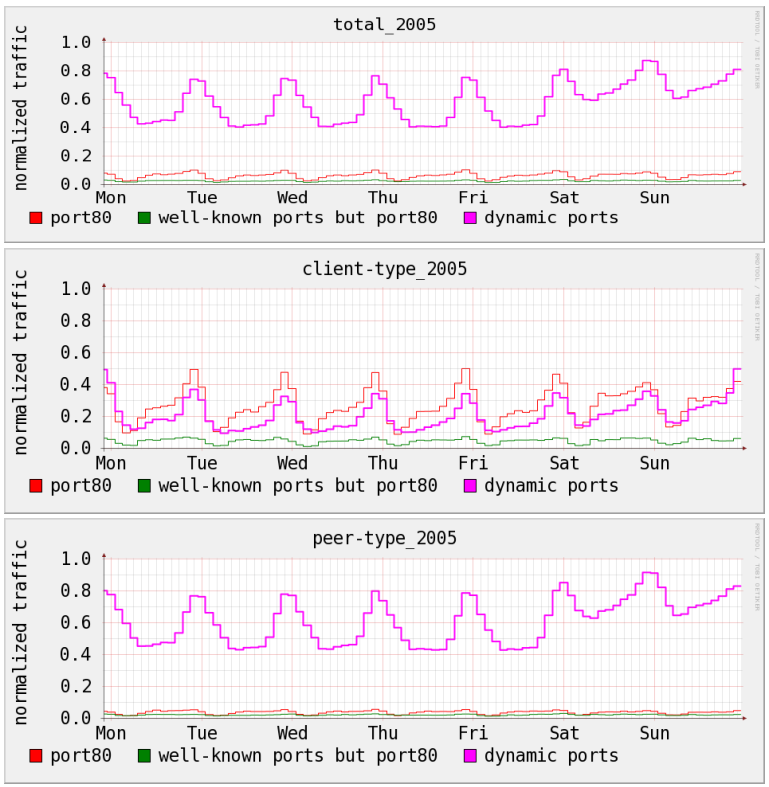
protocols/ports ranking

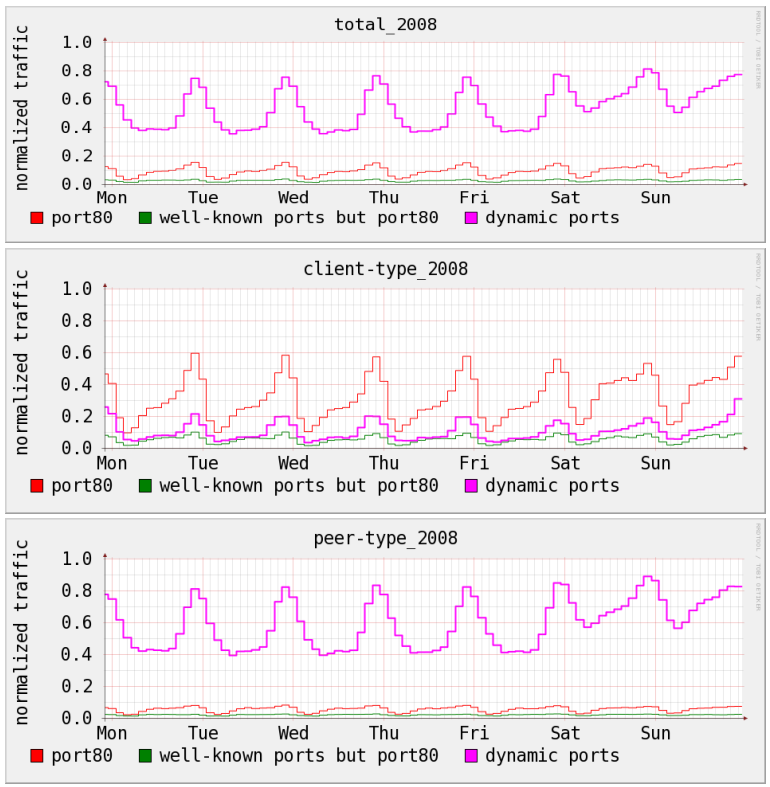
classify client-type/peer-type with threshold: 100MB/day upload

protocol port	2005 total (%)	2005 client type	2005 peer type	2008 **total** (%)	2008 client type	2008 peer type
TCP *	**97.43**	**94.93**	**97.66**	**96.00**	**95.51**	**96.06**
(< 1024)	*13.99*	*58.93*	*8.66*	*17.98*	*76.16*	*11.35*
80 (http)	9.32	50.78	5.54	14.06	64.96	8.26
554 (rtsp)	0.38	2.44	0.19	1.36	8.21	0.58
443 (https)	0.30	1.45	0.19	0.58	1.63	0.46
20 (ftp-data)	0.93	1.25	0.90	0.24	0.17	0.25
(>= 1024)	*83.44*	*36.00*	*89.00*	*78.02*	*19.35*	*84.71*
6346 (gnutella)	0.92	0.84	0.93	0.94	0.67	0.97
6699 (winmx)	1.40	1.14	1.43	0.68	0.24	0.73
7743 (winny)	0.48	0.15	0.51	0.30	0.04	0.33
1935 (rtmp)	0.20	0.81	0.14	0.22	0.73	0.16
6881 (bittorrent)	0.25	0.06	0.27	0.22	0.02	0.24
UDP *	**1.38**	**3.41**	**1.19**	**1.94**	**2.50**	**1.88**
53 (dns)	0.03	0.14	0.02	0.04	0.12	0.03
others	1.35	3.27	1.17	1.90	2.38	1.85
ESP	**1.09**	**1.35**	**1.06**	**1.93**	**1.85**	**1.94**
GRE	**0.07**	**0.12**	**0.06**	**0.09**	**0.08**	**0.09**
ICMP	**0.01**	**0.05**	**0.01**	**0.02**	**0.05**	**0.02**

temporal behavior of TCP port usage

3 types: port 80, well-kown port but 80, dynamic ports
total users (top), client-type (middle), peer-type (bottom) in 2005 (left) and 2008 (right)





summary of per-customer traffic analysis

overall traffic is still dominated by heavy-hitters, mainly using p2p

- but its traffic decreased in population share and volume share

current slow growth is due to stalled growth of dominant aggressive p2p traffic

client-type traffic slowly moving towards high-volume

- circumstantial evidence: driven by video and other rich media

growth model based on lognormal distributions

fitting client-type outbound volumes to lognormal dist.

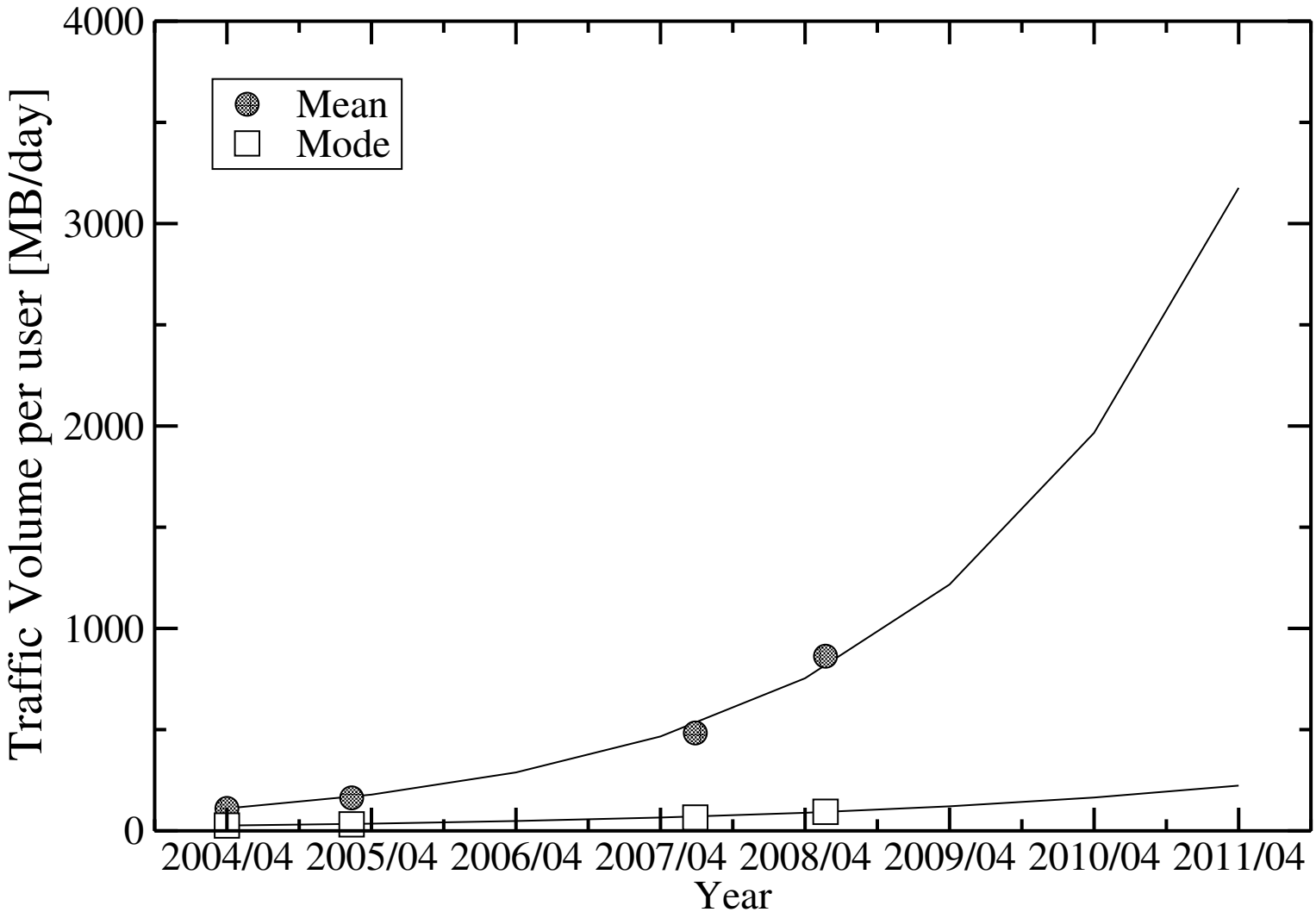
$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2}\right)$$

$$E(x) = \exp(\mu + \sigma^2/2)$$

- by definition, mean grows much faster than mode
- simplistic growth projections for outbound traffic per user (MB/day) for client-type users

	mode	mean
2004 Apr	26.2	110.6
2005 Feb	32.0	162.7
2007 Jul	65.7	483.2
2008 Jun	94.1	862.6
growth/yr	1.36	1.62
2009 Jun	121	1217
2010 Jun	164	1966
2011 Jun	223	3176

outbound traffic growth of client-type users



conclusion

apparent slow growth attributed to decline of p2p traffic

- but p2p willl not go away anytime soon
- p2p could evolve for large scale distribution

crustal is slowly swelling with video content

- similar to how web traffic was perceived in late 90es

network capacity also grows 50% per year (by various sources)
difficult to predict future traffic (lognormal!)
many challenges ahead

- technical factors: content caching, CDN, QoS
- economic factors: access cost, capacity/equipment costs
- political/social factors: net-neutrality, content management

acknowledgments

- IIJ, SoftBank Telecom, K-Opticom, KDDI, NTT Communications, SoftBank BB for data collection support
- ministry of internal affairs and communications for coordination

references

[CFEK2006] K. Cho, K. Fukuda, H. Esaki, and A. Kato.
The impact and implications of the growth in residential user-to-user traffic.
In *SIGCOMM2006*, Pisa, Italy, Aug. 2006.

[Cisco2008b] Cisco.
Approaching the zettabyte era.
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374_ns827_Networking_Solutions_White_Paper.html, June 2008.

[Cisco2008a] Cisco.
visual networking index – forecast and methodology, 2007-2012.
http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html, June 2008.

[Odlyzko2008] A. M. Odlyzko.
Minnesota Internet traffic studies.
http://www.dtc.umn.edu/mints/home.html.