

# 大規模映像共有サービスの ワークロード分析

NTTサービスインテグレーション基盤研究所  
森・川原・長谷川・下川

# 背景

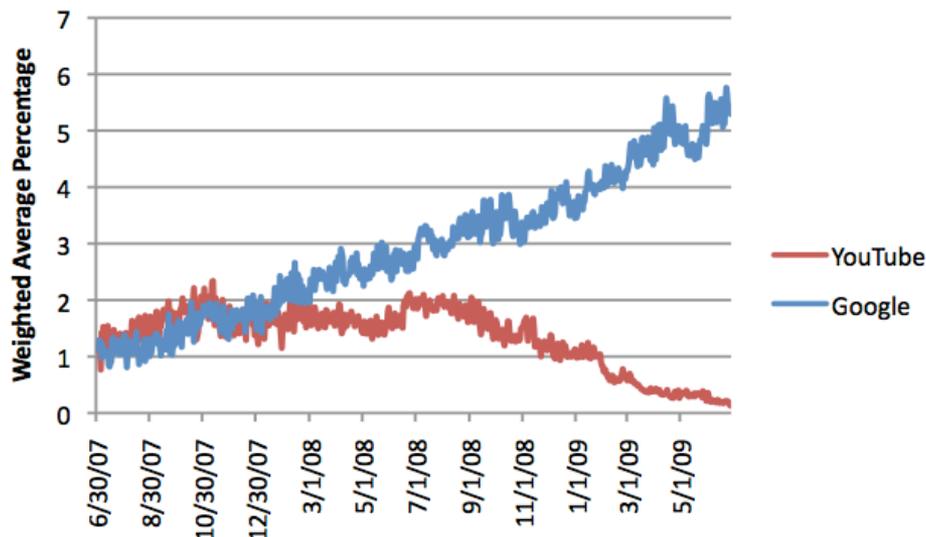
- インターネットトラフィックの変化～ Post YouTube

Rank	Provider	Percentage
1	Level(3)	5.77
2	Global Crossing	4.55
3	ATT	3.35
4	Sprint	3.2
5	NTT	2.6
6	Cogent	2.77
7	Verizon	2.24
8	TeliaSonera	1.82
9	Savvis	1.35
10	AboveNet	1.23

(a) Top Ten 2007

Rank	Provider	Percentage
1	Level(3)	9.41
2	Global Crossing	5.7
3	Google	5.2
4		
5		
6	Comcast	3.12
7		
8	<i>Intentionally omitted</i>	
9		
10		

(b) Top Ten 2009



Arbor Networks  
NANOG47 資料より抜粋

# 背景

Consumer Internet Traffic, 2008-2013							
	2008	2009	2010	2011	2012	2013	CAGR 2008-2013
By Sub-Segment (PB per month)							
Web/Email	1,239	1,595	2,040	2,610	3,377	3,965	26%
File Sharing	3,384	4,181	5,192	6,529	8,123	10,127	25%
Internet Gaming	47	87	135	166	217	239	39%
Internet Voice	103	129	152	174	183	190	13%
Internet Video Communications	36	57	94	160	239	354	58%
Internet Video to PC	1,087	2,421	4,232	6,794	9,415	12,087	62%
Internet Video to TV	29	150	358	1,000	1,679	2,507	144%
Ambient Video	110	224	633	1,329	2,079	2,685	90%

Cisco Visual Networking Index より抜粋

# 本研究のゴール

- 新しいトラヒックの理解・characterization
  - 特に「フロー」の視点で
  - ネットワーク設計・制御への影響把握

# 分析の対象とした映像共有サービス

サービス		拠点	容量(free)	容量(premium)	ファイル
YouTube		アメリカ	2GB or 10分	20GB	flv, 3gp, mp4
ニコニコ動画		日本	40MB	100MB	flv
MEGAVIDEO		中国(香港)	100MB	5GB	flv
DailyMotion		フランス	20分 or 150MB	--	flv, mp4

# フロー検出におけるチャレンジ

- 単純にIPアドレスの管理主体を調べるだけでは不十分
  - YouTube のトラフィック = AS 36561 (YouTube) 発のトラフィックというわけではない
  - YouTube サーバの一部はGoogle ASN に所属
    - YouTube (ASN): 18 GB
    - Google (ASN) + YouTube (ASN): 226 GB
    - YouTube (今回の分析結果): 182 GB
- ペイロード情報（パケットの中身）を使わない分析
  - そもそも100Gbps 超の世界では DPI (Deep Packet Inspection) の限界が来る可能性大
  - 一方ネットワークに何が流れているかを把握しておきたいという要求は普遍的に存在

# YouTube ドメイン発トラヒック

- http/content-type が様々
  - application/atom+xml, application/javascript, application/octet-stream
  - application/opensearchdescription+xml, application/rss+xml
  - application/vnd.google-earth.kmz, application/x-javascript
  - application/x-shockwave-flash, application/x-www-form-urlencoded
  - application/x-xpinstall, application/xml, application/xslt+xml
  - image/gif, image/jpeg, image/png, image/x-icon
  - text/css, text/html, text/javascript, text/plain, text/x-cross-domain-policy, text/xml
  - video/mp4, video/x-flv ← 興味があるのはこれ
  - Observation: ビデオ配送サーバ群と html (UI) 配送サーバ群は異なる

# YouTubeトラヒックの抽出方法(1)

## 1. 大規模URL 収集と分析

- web proxy サーバのログ(1ヶ月間で1億レコード)より content-type: が video かつURLのドメインが /youtube\.com\$/ にマッチするドメイン を収集
- 規則性の発見(リバースエンジニアリング)  
v◇◇.cache◇◇.c.youtube.com

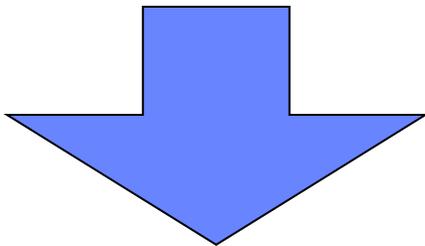
## 2. ドメインリスト作成

- ドメインを補完し、実際にDNS正引きが存在するものをリストに追加

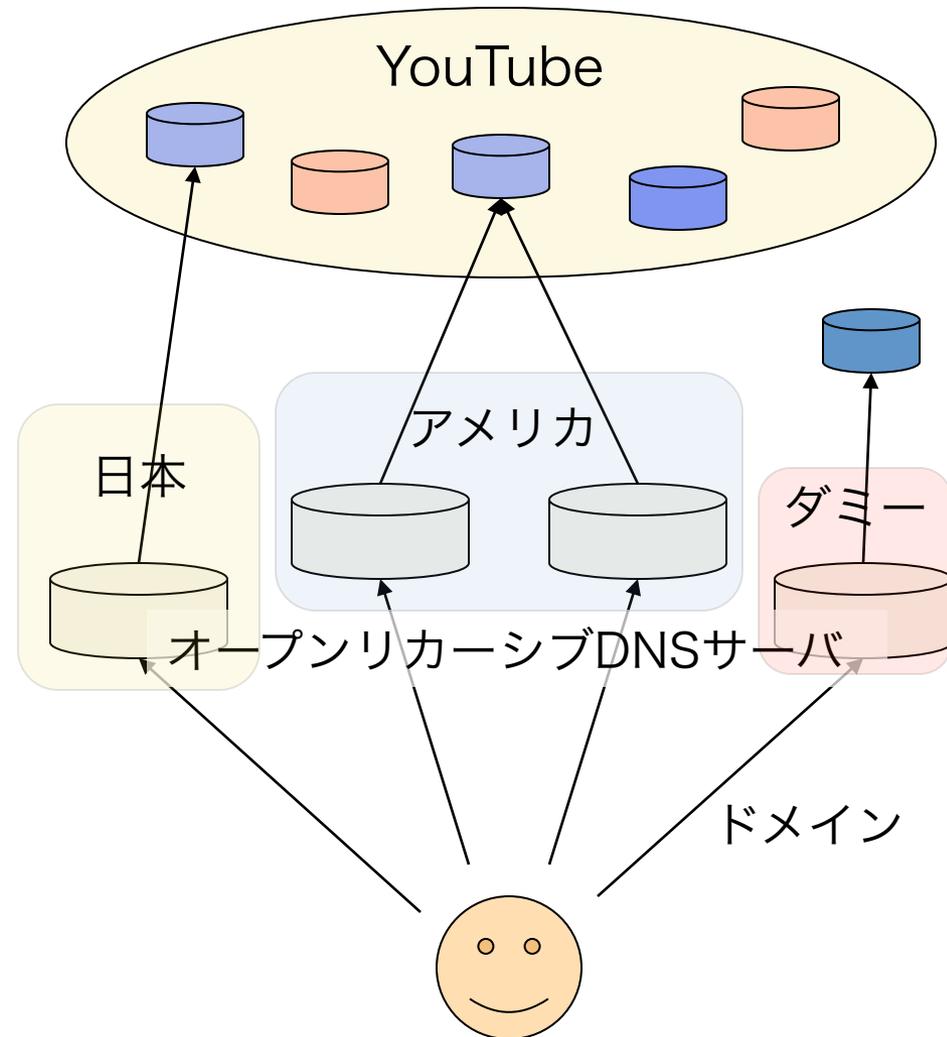
# YouTubeトラヒックの抽出方法(2)

## 3. IPアドレスリスト作成

- 世界中に分散したオープンリカーシブDNSサーバ経由でDNSリストのドメインをlook up (Aレコード)



動画共有サービスにおける  
動画配信サーバのIPアドレスを網羅的に抽出可能



# 収集したドメインとIPアドレス

ドメイン

hostnames	complemented range	#observed transactions
v◎.lscache◎.c	$1 \leq \odot \leq 24, 1 \leq \otimes \leq 8$	130,286
v◎.cache◎.c	$1 \leq \odot \leq 8, 1 \leq \otimes \leq 8$	27,485
tc.v◎.cache◎.c	$1 \leq \odot \leq 24, 1 \leq \otimes \leq 8$	1626
v◎.nonxt◎.c	$1 \leq \odot \leq 24, 1 \leq \otimes \leq 8$	25
lax-v◎.lax	$1 \leq \odot \leq 308$	19
sjl-v◎.sjl	$1 \leq \odot \leq 50$ (with exceptions)	19

service	hostnames	complemented range
Smiley videos	smile-com◎◎.nicovideo.jp	$0 \leq \odot \leq 6, 0 \leq \otimes \leq 3$
Smiley videos	smile-clb◎◎.nicovideo.jp	$0 \leq \odot \leq 6, 0 \leq \otimes \leq 3$
Megavideo	www◎.megavideo.com	◎ can be any positive integer.
Dailymotion	proxy-◎◎.dailymotoin.com	$0 \leq \odot \leq 9, 0 \leq \otimes \leq 9$

IPアドレス

Service	#addresses
YouTube	2,138
Smiley videos	74
Megavideo	670
Dailymotion	100

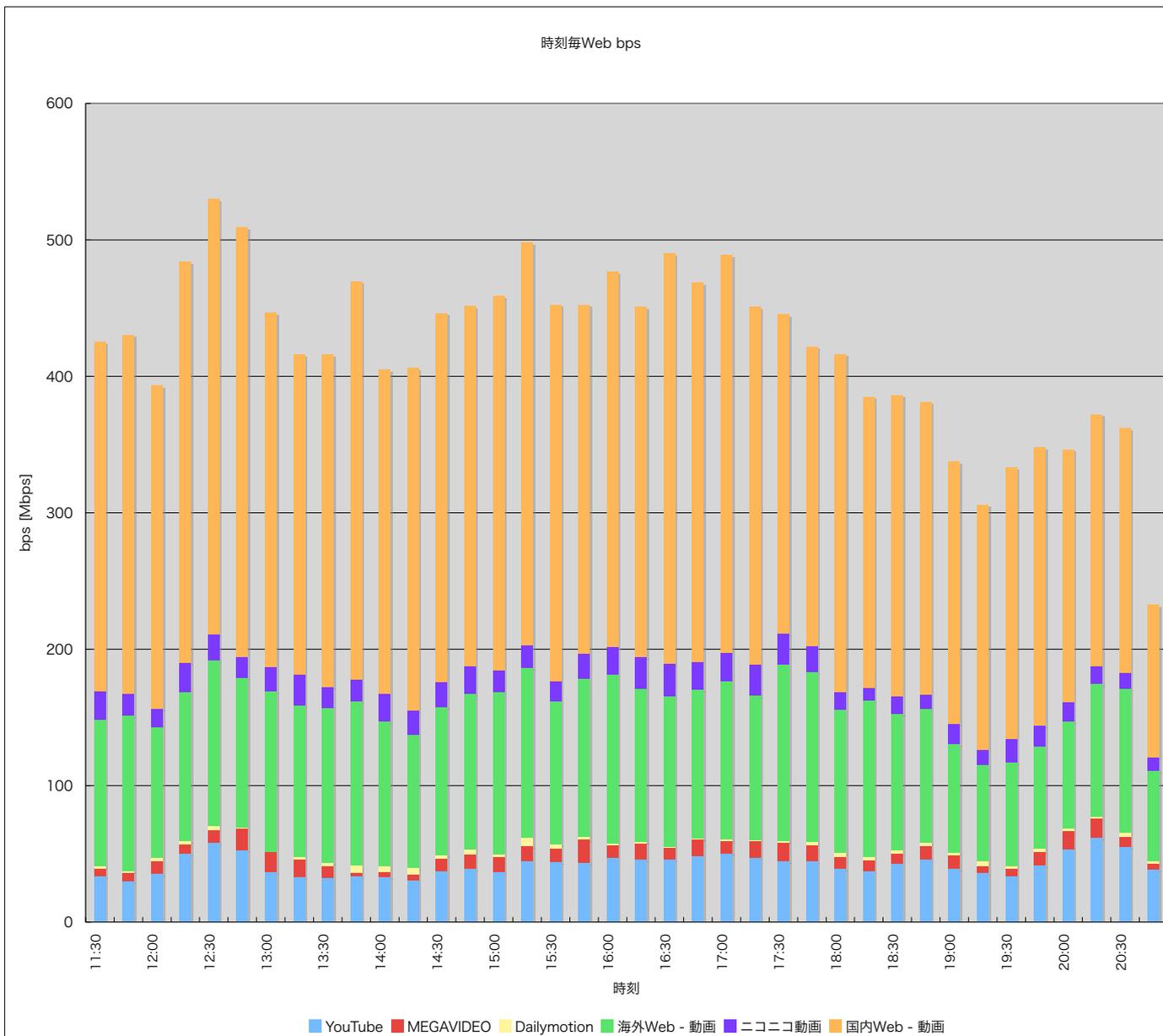
# トラフィックデータ

- あるインターネットエッジを計測
- 下りの通信路における9時間分のトラフィックデータ
- 下記のようなフローレコードを取得

通信開始時刻	通信終了時刻	サーバ側IP	クライアントID	ポート番号	フローサイズ	TCPヘッダ
1236911063	1236911070	133.x.x.x	d93jfoe0afm3	80	34530	APSF--
1236911075	1236911185	193.y.y.y	hfe983h1r091	25190	2901948	APS---

- 全フロー数 ~ 約1億
- 取得したIPアドレスリストを用い、映像共有サイトのトラフィックフローを分析する

# Webトラフィックの内訳



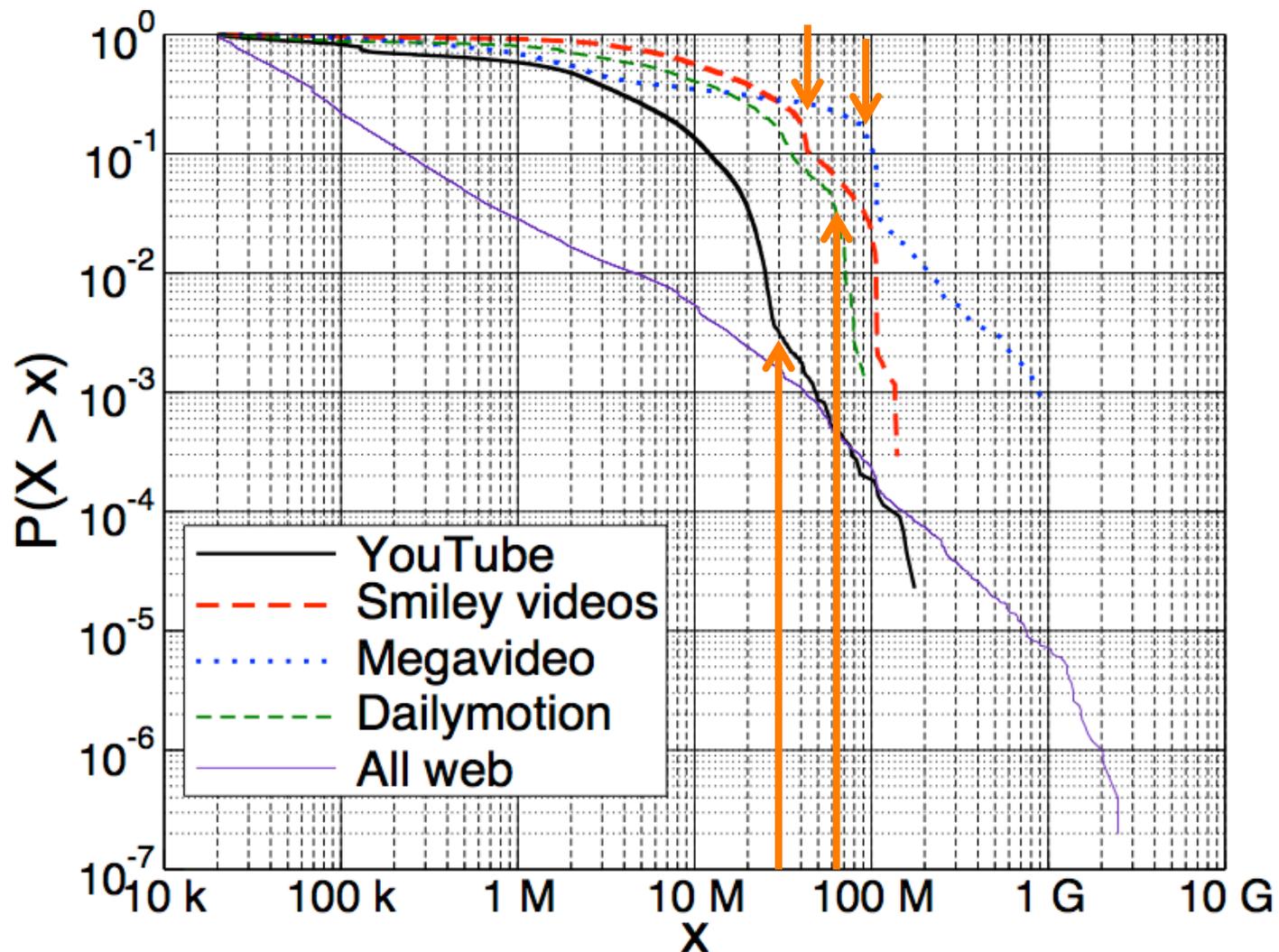
- YouTubeなどの動画トラフィックは海外Webトラフィックの3分の1程度を占める

# フロー統計

サイズが20KB 以上のフローのみ集計

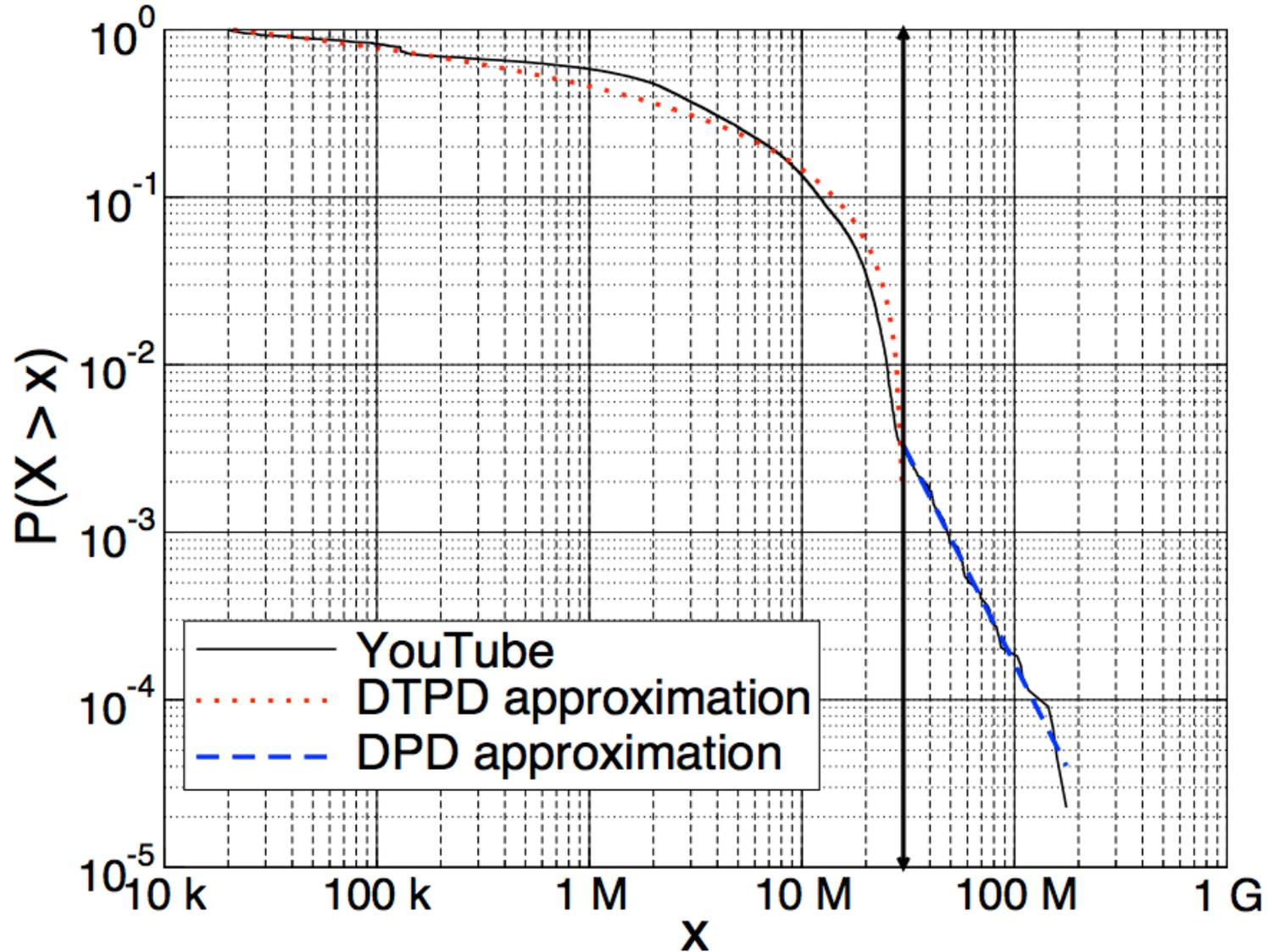
Service	#flows	mean size	mean rate	mean duration
YouTube	43,960	4.1 MB	1.3 Mbps	41.8 sec
Smiley videos	3,438	21.3 MB	2.6 Mbps	139.8 sec
Megavideo	1,354	30.0 MB	1.3 Mbps	232.6 sec
Dailymotion	730	13.7 MB	1.5 Mbps	96.0 sec
All web	5,043,927	0.33 MB	0.9 Mbps	16.5 sec

# フローサイズ分布

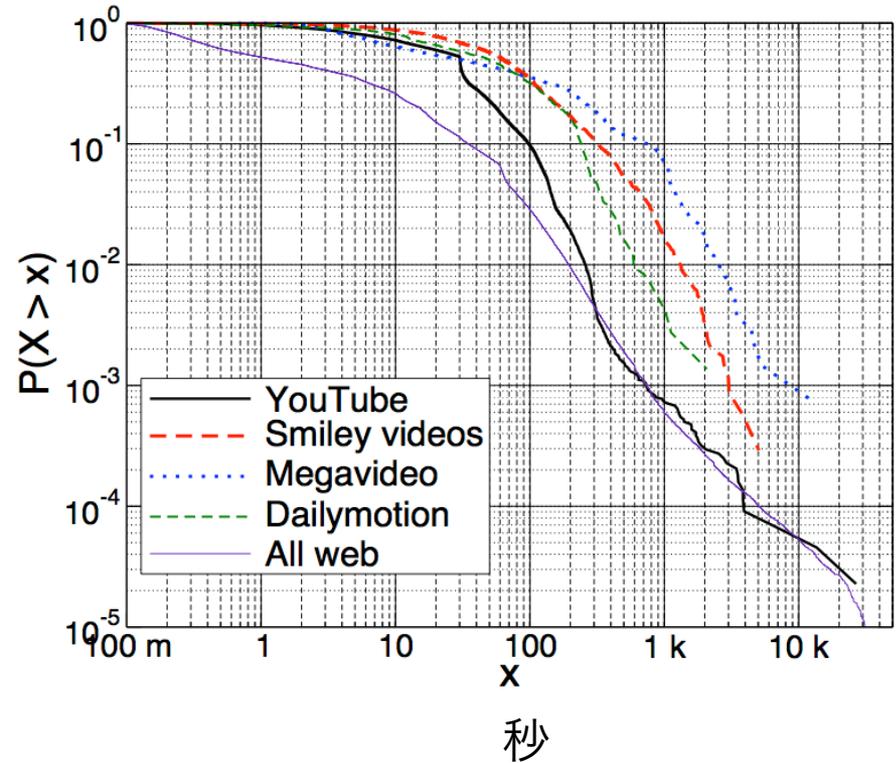
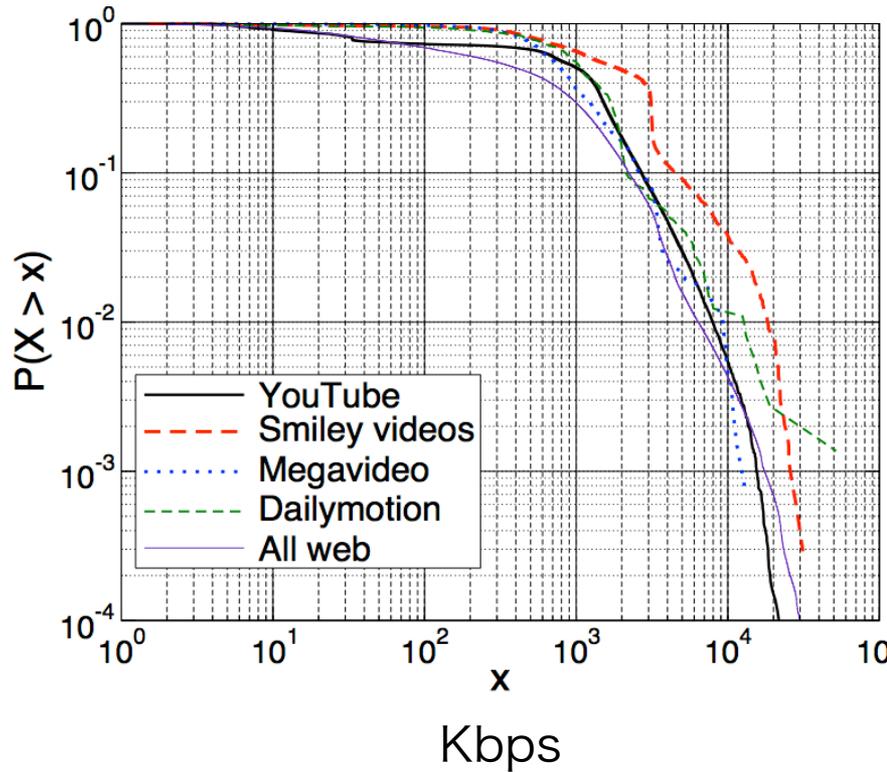


容量制限によるサービス毎の特徴が明確

# フローサイズ分布のフィッティング

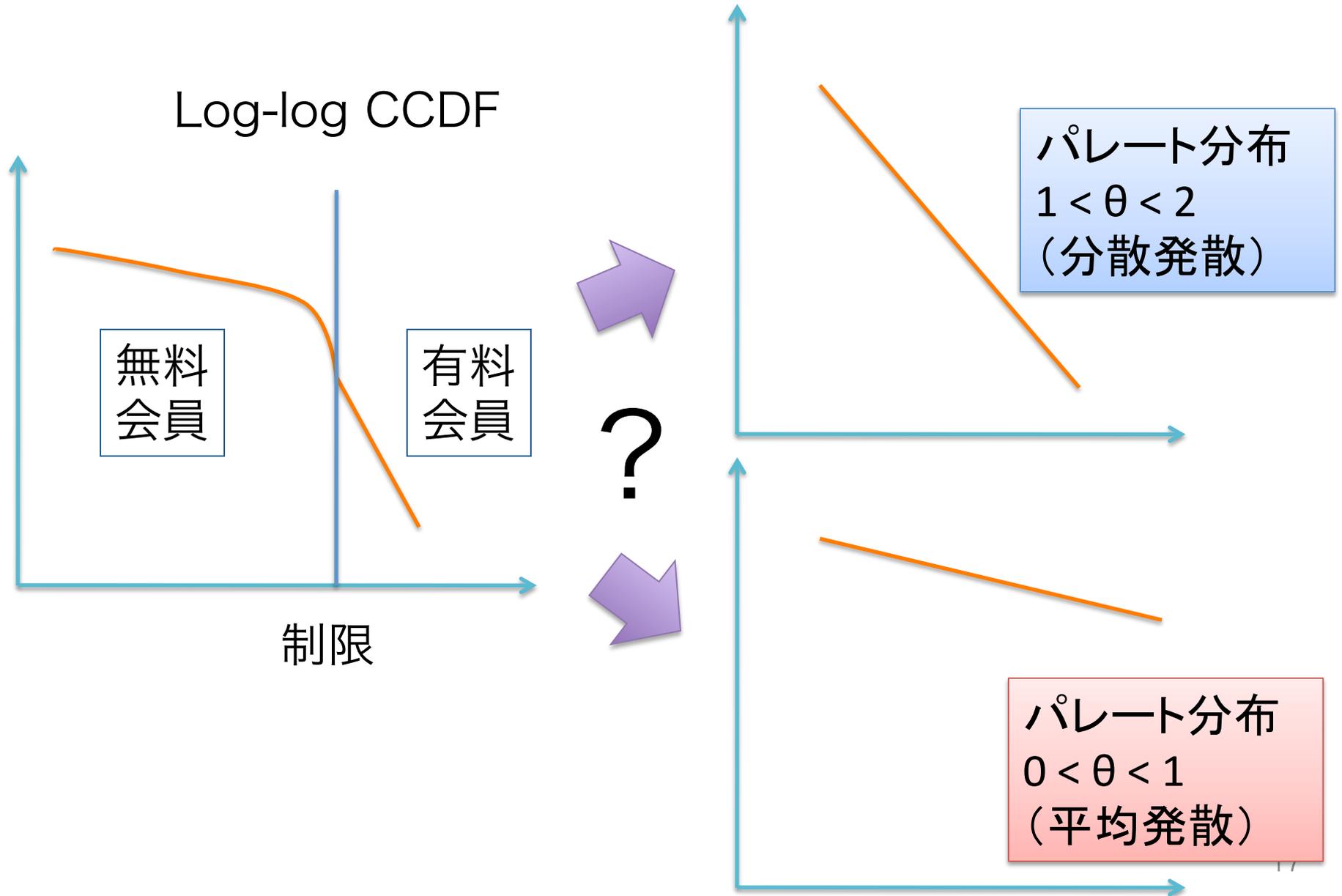


# フローレート・フロー持続時間



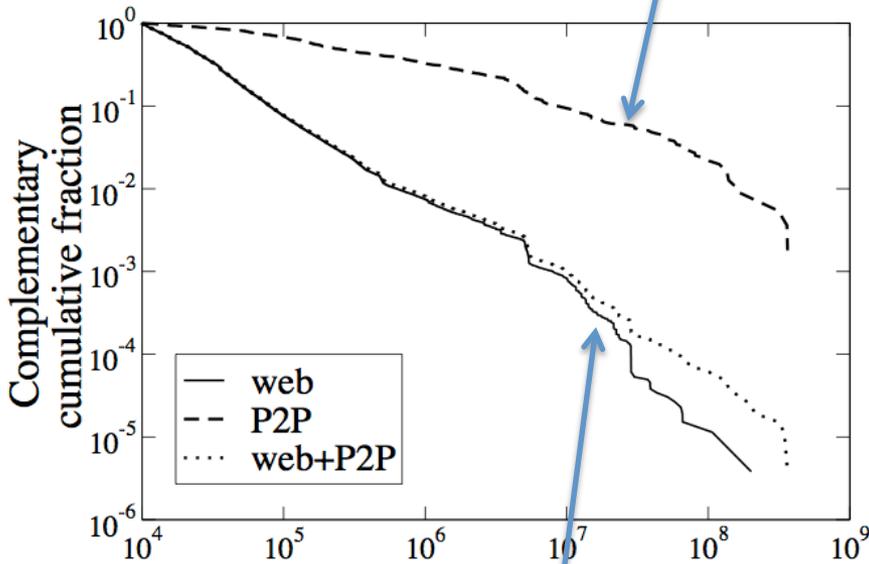
サービス間で大きな分布形状の差は認められない  
持続時間は巾乗分布的

# 容量制限がなくなるとどうなる？



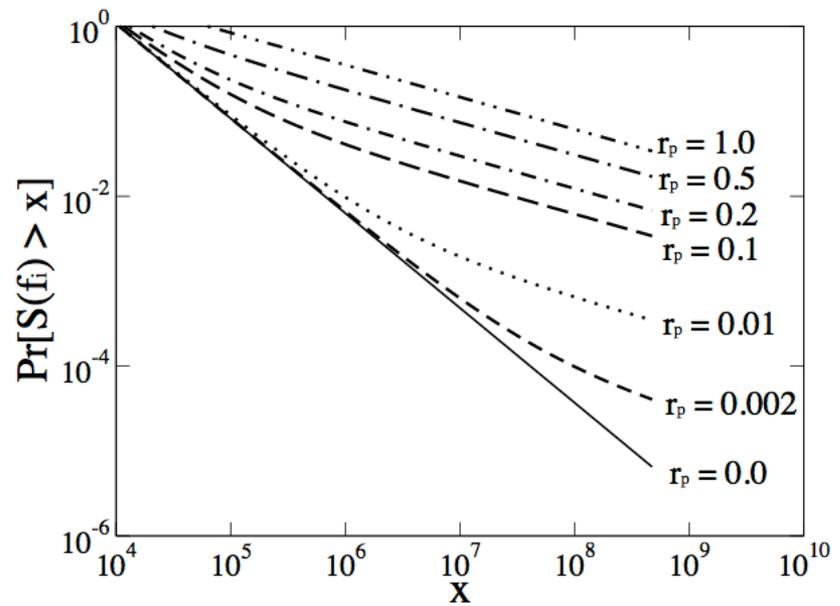
# P2Pフローとの比較

P2P:  $\theta=0.38$



Web:  $\theta=1.12$

## P2P増加時のインパクト



Tatsuya Mori, Masato Uchida, Shigeki Goto,  
*Flow analysis of internet traffic: World Wide Web versus peer-to-peer.*  
Systems and Computers in Japan 36(11): 70-81 (2005)

# まとめ

- 現在の映像共有サービスのトラフィックの特徴を分析
- DPIや統計的推定によらないフロー検出方法の提案
- サービス容量制限がフローサイズ分布特性に影響
- 制限付きのP2Pフローと近い特性
  - ただしトラフィック源がピアからデータセンターにシフト
  - 容量制限の解除による変化は未知数（今後の課題）
- 今後の変化・成長とネットワークに対する影響把握は引き続き課題である