ビデオ共有サービスを対象とした重複トラフィック削減手法の提案とシミュレーションによる評価

◯安藤翔伍, 中尾彰宏
東京大学大学院 中尾研究室

1

1.研究背景と目的

2. 関連研究
3. 重複ビデオトラフィックの解析
4. 重複ビデオトラフィック削減手法の検討
5. 結論

2

1. ビデオトラフィックが多量に存在している問題

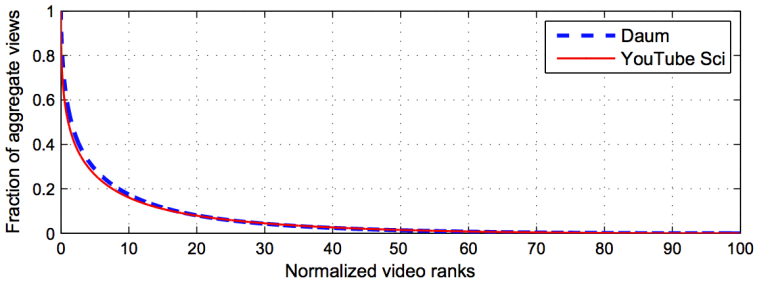
- インターネットにおいてビデオトラフィックが増加傾向にある
  - ビデオトラフィックがP2Pファイル共有トラフィックを上回る
  - ビデオトラフィックはインターネット全体の36-41%*
  - YouTubeは1日あたり165万ドルの赤字を出している**
- インターネット全体のトラフィックも増加***
  - 日本の2010年のインターネットトラフィックは，前年比で25.4%増加
  - 時間別のトラフィック変化でも2005年と比較して，低値とピーク値の差が1.5倍に増加
    - P2P中心のトラフィックからオンラインビデオ等のユーザ行動に付随したトラフィックへと変化

* Cisco Visual Networking Index：Usage http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/Cisco VNI Usage WP.pdf
** Google Losing up to $1.65M a Day on YouTube http://www.internetevolution.com/author.asp?section_id=715&doc_id=175123
*** 我が国のインターネットにおけるトラヒック総量の把握. http://www.soumu.go.jp/ main content/000109282.pdf

3

2. インターネットに重複ビデオトラフィックが含まれる問題

- YouTubeの人気ランキング上位10%が全体のリクエストの80%を占めるパレートの法則に則っている*
- リクエストの偏り＝ビデオトラフィックには同一ビデオの再生が存在
- 同一ビデオの再生は，削減可能な重複したトラフィックを発生させ，ネットワーク負荷となる

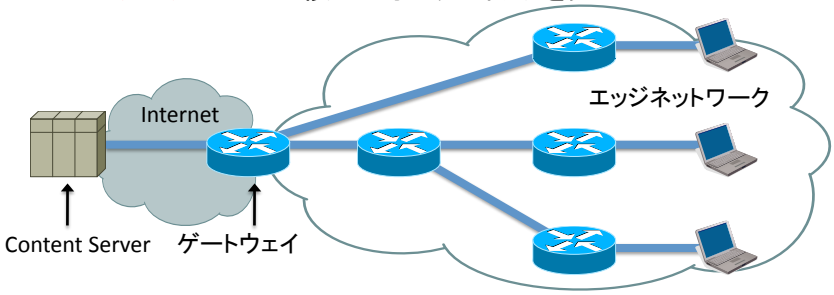


* MeeyoungCha,HaewoonKwak,PabloRodriguez,Yong-YeolAhn,andSueMoon.I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, IMC '07, pages 1–14, New York, NY, USA, 2007. ACM.

4

3.エッジネットワークに 重複ビデオトラフィックが含まれる問題

- 前述のリクエストの偏りは，インターネットのみではなく，エッジネットワークにおいても重複ビデオトラフィックを発生



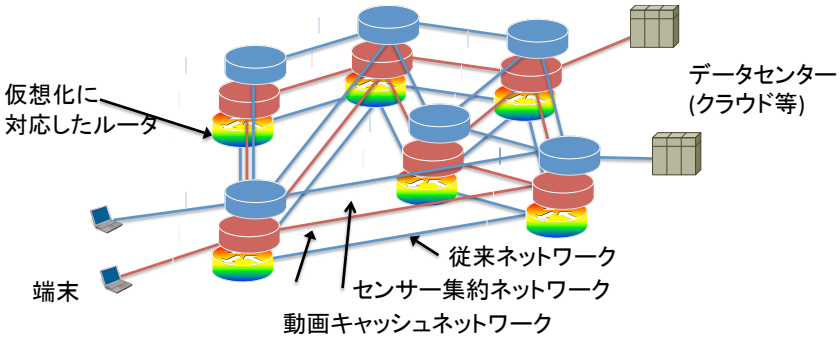
– 336時間のキャンパスネットワークからのYouTubeリクエスのうち，トラフィックの34.1%が重複*

* Michael Zink, Kyoungwon Suh, Yu Gu, and Jim Kurose. Characteristics of YouTube network traffic at a campus network - measurements, models, and implications. vol- ume 53, pages 501–514, March 2009.

5

研究背景:ネットワーク仮想化

- ネットワーク仮想化によりネットワーク内部のパケット処理が可能
  – 独立に確保された資源を用いて,任意のプログラムが実行可能*

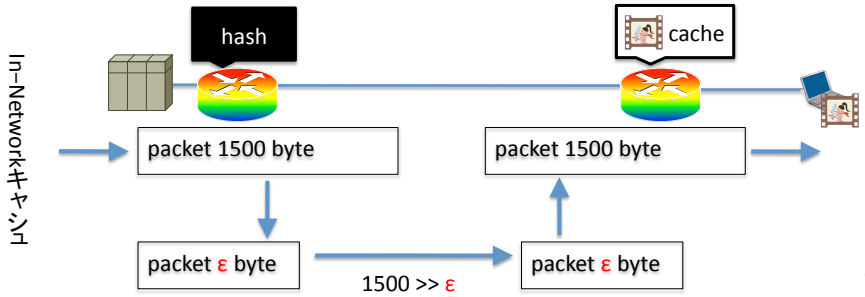


* Akihiro Nakao Network Virtualization as Foundation for Enabling New Network Architectures and Applications. IEICE TRANSACTIONS on Communications 1 Volume E93-B. No.3. 454-457

6

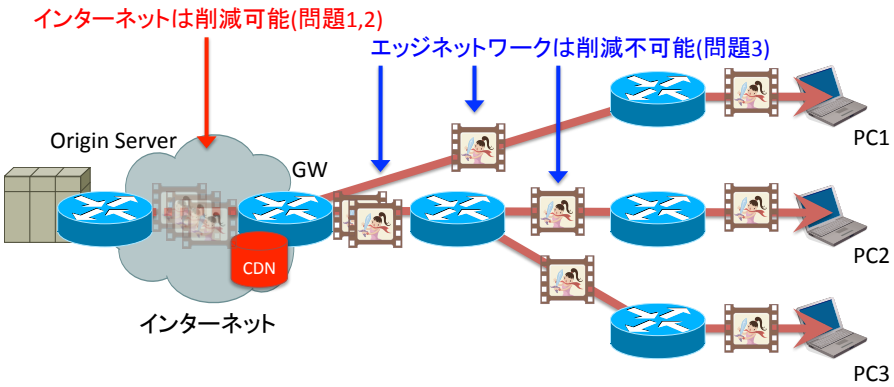
研究目的

- ネットワークの仮想化を利用し，インターネットの重複トラフィック(問題1,2)だけではなく，これまで削減対象とされていなかったエッジネットワークの重複トラフィック(問題3)も削減可能な，In-Network キャッシュ手法を提案し，削減効果を検討する
- In-Network キャッシュ手法
  - エッジネットワークの全てのルータ･スイッチがストレージを有し，パケットキャッシュ機能を実装し，冗長トラフィックを削減する手法



7

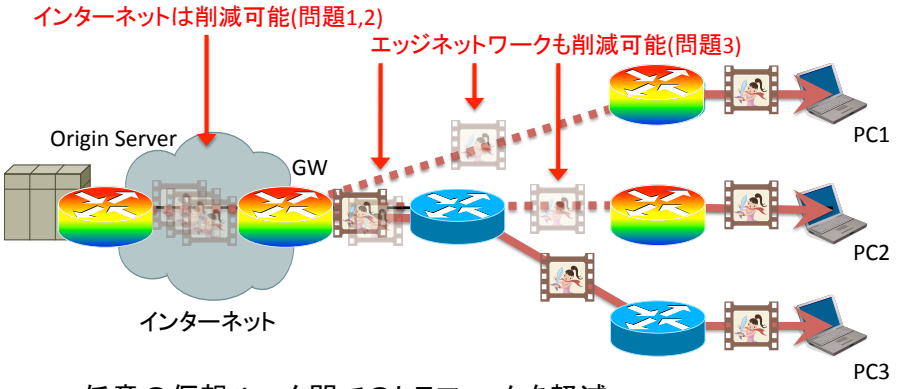
CDNを用いた配信方式



- インターネットの重複トラフィックは削減可能
- エッジネットワークの重複トラフィックは削減不可能

8

提案手法:In-Networkキャッシュ



- 任意の仮想ルータ間でのトラフィックを軽減
- インターネットの重複トラフィックは削減可能
- エッジネットワークの重複トラフィックも削減可能

9

2.関連研究

10

本研究と関連研究との差異

	Chaらの研究 *	Zinkらの研究 **	本研究
使用データ	YouTubeの クローリング	実トラフィック/ 合成トラフィック	実トラフィック
シミュレーション単位	リクエスト	分布に基づく転送量	実転送量
代替配布手段	仮想的なグローバル キャッシュ/P2P	ローカルキャッシュ/ GWプロキシ/P2P	In-Networkキャッシュ
削減対象	YouTube	Edge (campus)	Edge (campus)
インターネットの トラフィック削減	○	○	○
エッジネットワークの トラフィックの削減	×	×	○

– YouTubeを対象としたキャッシュのシミュレーション
  – インターネットのトラフィック削減のみを議論

*Meeyoung Cha,Haewoon Kwak,Pablo Rodriguez,Yong-Yeol Ahn,and Sue Moon.I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, IMC '07, pages 1–14, New York, NY, USA, 2007. ACM.
** Michael Zink, Kyoungwon Suh, Yu Gu, and Jim Kurose. Characteristics of YouTube network traffic at a campus network - measurements, models, and implications. vol- ume 53, pages 501–514, March 2009.

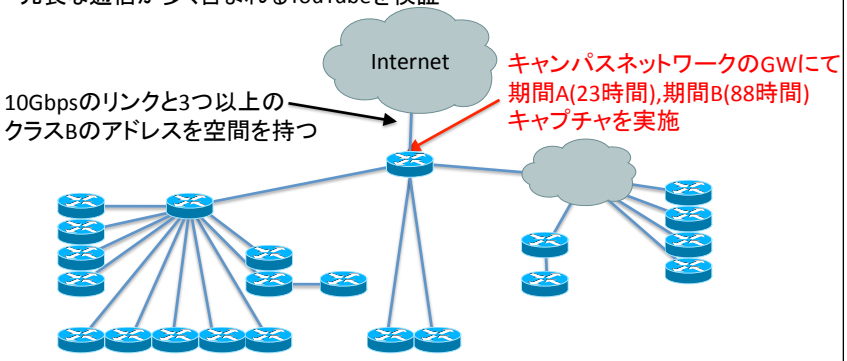
11

3.重複ビデオトラフィックの解析

12

キャンパスネットワークにおける YouTubeトラフィック解析

- In-Networkでのデータキャッシュの実現可能性の検討
  - 冗長な通信が多く含まれるYouTubeを検証

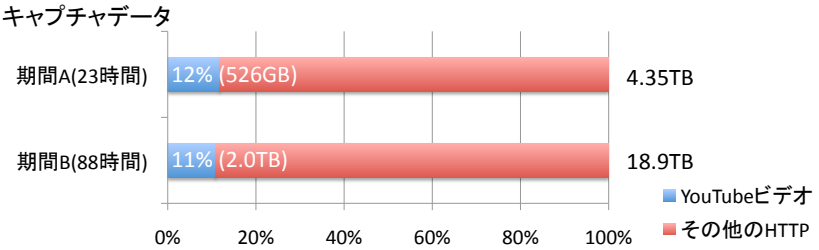


期間	開始時刻	終了時刻	総キャプチャ時間
期間A	2010/12/15 14:40	2010/12/16 13:45	23 時間 5 分間
期間B	2011/05/16 06:01	2011/05/19 21:57	3 日 15 時間 56 分間

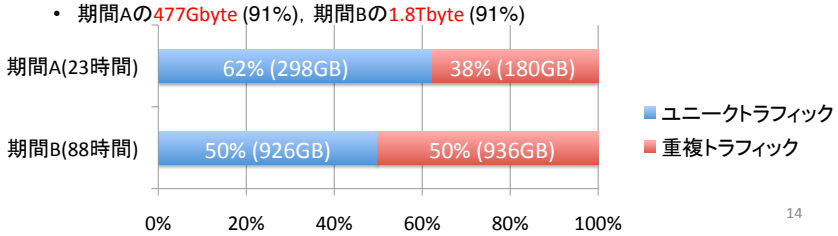
13

トラフィック解析の結果

- キャプチャデータ

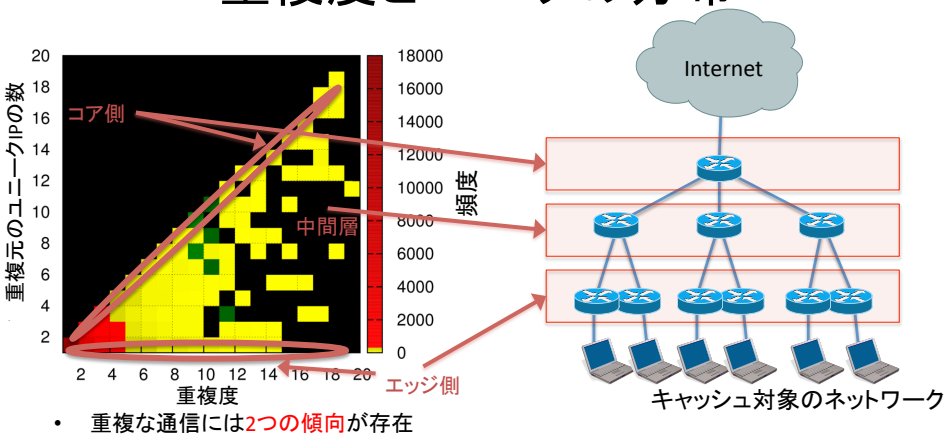


- 解析対象
  - エラー，206 “Partial Content”，途中からのアクセスを除外
    - 期間Aの477Gbyte (91%)，期間Bの1.8Tbyte (91%)



14

重複度とユーザの分布



- 重複な通信には2つの傾向が存在
  - 1つ/少数のユーザが何度もアクセスすることにより発生する重複
  - 多数のユーザが1回/数回ごとアクセスすることにより発生する重複
- 重複トラフィックの発生傾向によってキャッシュの配置
  ネットワークコア側,中間層,エッジ側に効率よく分散すると,
  高いキャッシュ効果が期待できる

15

4.重複ビデオトラフィック削減手法の検討

- In-Networkキャッシュシミュレーション
- 5つのキャッシュ手法の概要
- シミュレーション結果

16

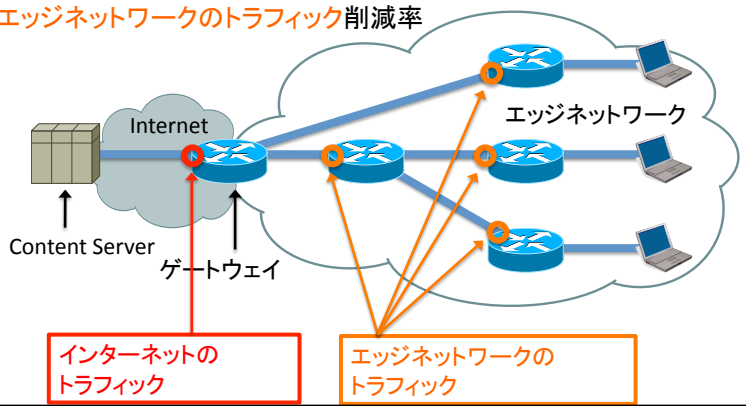
In-Networkキャッシュのシミュレーション

- In-Networkキャッシュのトラフィック削減効果の検討
- トラフィック解析で得たYouTubeのトラフィックデータを使用しIn-Networkキャッシュのシミュレーションを実施
  - class B x 3 のアドレス空間を持つネットワーク
    - netmask 8, 256のIPごとに集約されているエッジネットワーク
  - キャッシュ戦略
    - 全ての動画をキャッシュ対象
    - Least Recently Usedのキャッシュアルゴリズム
    - パケットのハッシュ化により，トラフィックは0になると仮定
  - 使用データ
    - 期間B，88時間のキャプチャデータを使用
    - フローレベルでのシミュレーションを実施するため，データリンク層での重複を除き，42.2%が重複トラフィックであるデータを使用

17

In-Networkキャッシュのトラフィック削減率

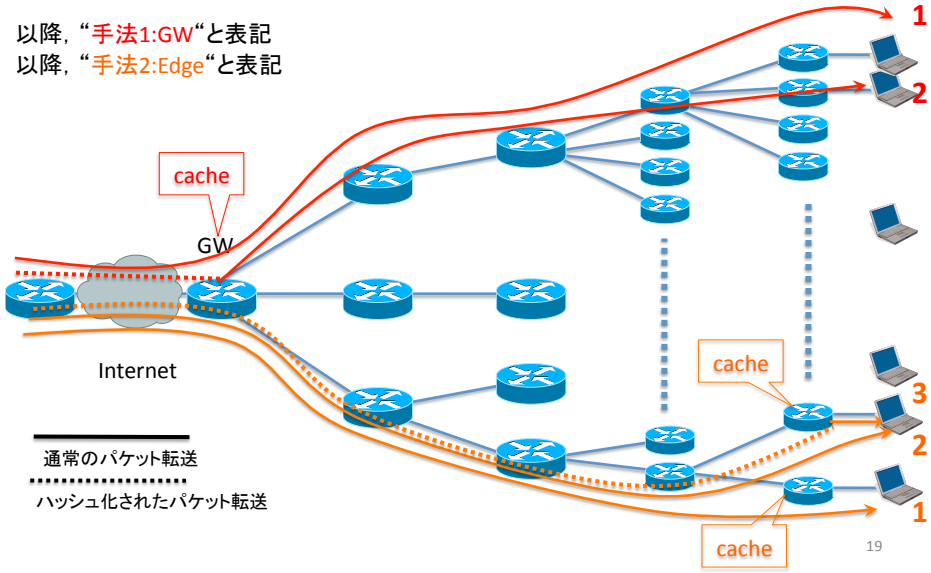
- シミュレーション内容
  - 5つのキャッシュ手法 (後述)
  - 3つのキャッシュサイズの条件 (後述)
- シミュレーション評価
  - インターネットのトラフィック削減率
  - エッジネットワークのトラフィック削減率



18

手法1: GW，手法2:エッジ最外部ルータでのキャッシュ

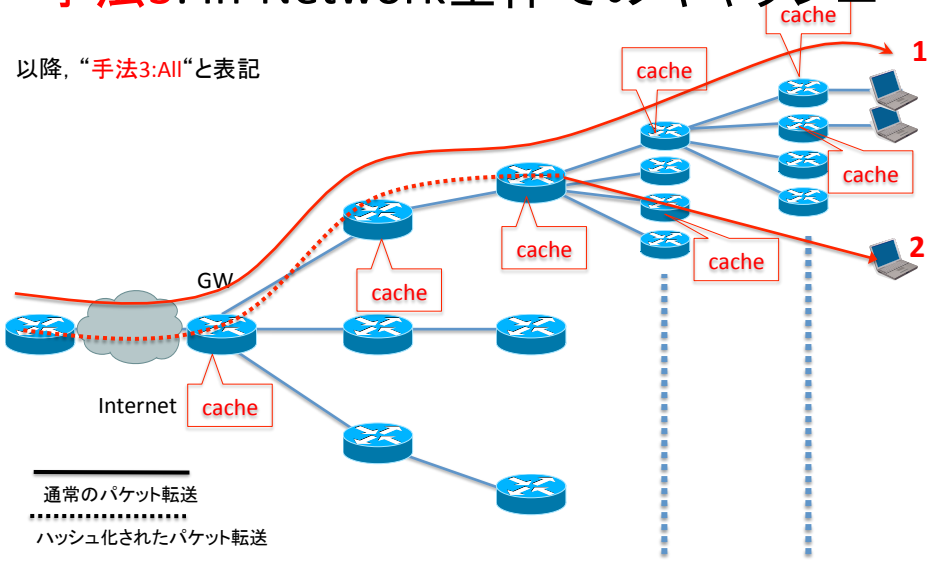
以降，“手法1:GW“と表記
以降，“手法2:Edge“と表記



19

手法3: In-Network全体でのキャッシュ

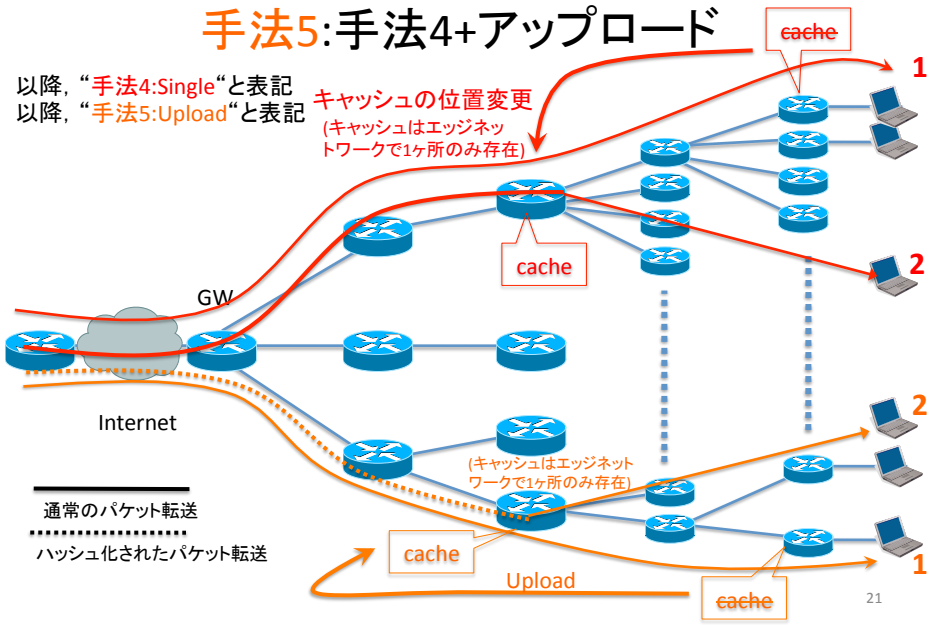
以降，“手法3:All“と表記



20

手法4:手法3+協調キャッシュ
手法5:手法4+アップロード

以降，“手法4:Single“と表記
以降，“手法5:Upload“と表記



21

3つのキャッシュサイズの条件

- 3つのキャッシュサイズの条件でシミュレーション
  - (a) 容量総和制限を設けない場合
    - 手法3: All の削減効果が最も良い事を検証
  - (b) 容量総和制限を設け，均等にキャッシュサイズを割り振った場合
    - 手法1: GW の削減効果が流入トラフィックにおいて良い事を検証
  - (c) 容量総和制限を設け，最適にキャッシュサイズを割り振った場合
    - トラフィックに応じたキャッシュが可能な手法5: Uploadの削減効果が最も良い

- 容量総和制限:
  ルータ上でキャッシュする -> ルータが多い方が効率が良い
  同条件で各手法の削減率を検討するため
  $\sum$ (各ルータのキャッシュサイズ)　を全ての手法で等しくする

22

(a)容量総和制限を設けない場合

In-Network手法	インターネットのトラフィック削減率	エッジネットワークのトラフィック削減率
手法1: GW	42.2%	0%
手法2: Edge	25.8%	25.8%
手法3: All	42.2%	30.2%
手法4: SIngle	34.7%	17.7%
手法5: Upload	42.2%	18.9%(Upload Traffic+6.6%)

- (a) 容量総和制限を設けない場合，インターネット/エッジネットワーク共にトラフィック削減率は，手法3：All が最も高い
  - 他の手法と比較して最大4.7倍のキャッシュサイズを使用
- キャッシュサイズの容量総和を等しくした場合でのトラフィック削減効果の検討を行う

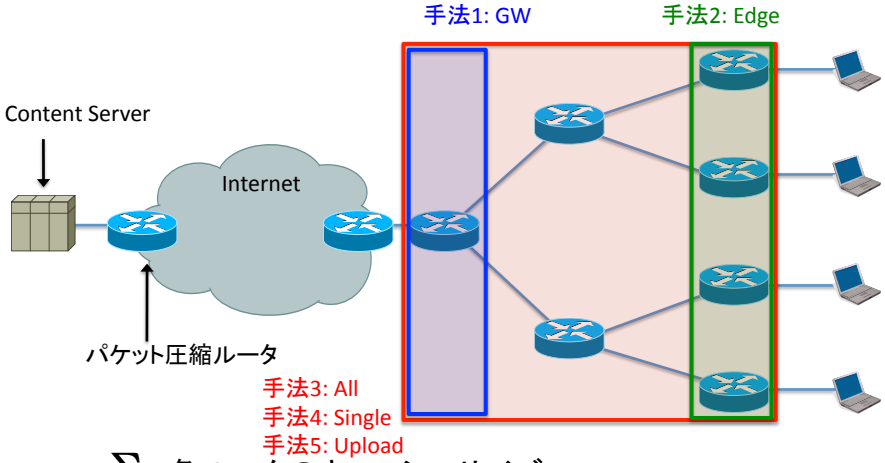
23

(b,c)容量総和制限を設ける場合

- (a) 容量総和制限を設けない場合，インターネット/エッジ トラフィック共に削減率は，手法3：All が最も高い
  - 手法1:GW, 手法4:Single, 手法5:Uploadに比べて4.6-4.7倍
    手法2:Edgeと比べて3.6倍のキャッシュサイズを使用
- キャッシュサイズの容量総和を等しくし，以下の2つの場合でのトラフィック削減効果の検討を行う
  - (b)容量総和制限を設け,均等にキャッシュサイズを割り振った場合
  - (c)容量総和制限を設け,最適にキャッシュサイズを割り振った場合

24

(b)容量総和制限を設け，均等にキャッシュサイズを割り振った場合



$\sum$ (各ルータのキャッシュサイズ) = 926GB
= ユニークトラフィックサイズ

25

(b)容量総和制限を設け，均等にキャッシュサイズを割り振った場合(期間B)

In-Network手法	インターネットのトラフィック削減率	エッジネットワークのトラフィック削減率
手法1: GW	42.2%	0%
手法2: Edge	10.6%	10.7%
手法3: All	10.1%	10.3%
手法4: SIngle	9.3%	8.2%
手法5: Upload	12.2%	8.7%(Upload Traffic+3.0%)

- (b) 容量総和制限を設け，均等にキャッシュサイズを割り振った場合
  - インターネットのトラフィック削減率は，手法1：GW が最も高い
  - エッジネットワークのトラフィック削減率は，手法2：Edge が最も高い
  - (a) 容量総和制限を設けない場合の削減率である，インターネット:42.2%, エッジネットワーク:30.2%にはどれも達していない
- (c) 容量総和制限を設け，最適にキャッシュサイズを割り振った場合を考える
  - 最適なキャッシュ = 同一ビデオのキャッシュの存在はエッジネットワーク内で1つ
    - 手法1: GW，手法4: Single，手法5: Upload

26

(c)容量総和制限を設け，最適にキャッシュサイズを割り振った場合(期間B)

In-Network手法	インターネットのトラフィック削減率	エッジネットワークのトラフィック削減率
手法1: GW	42.2%	0%
手法4: SIngle	34.7%	17.7%
手法5: Upload	42.2%	18.9%(Upload Traffic+6.6%)

- (a) 容量総和制限を設けない場合の削減率である，インターネット:42.2%，エッジネットワーク:30.2%にはどれも達していないものの，キャッシュサイズが約1/5であることを考慮すると手法5: Upload は高い削減率を示す
- ただし，手法5:Uploadでは最大キャッシュ使用容量として一時的に4.4%余分にキャッシュを使用し，また6.6%のアップロードトラフィックが発生する．
- キャッシュの最適配置は，ゲートウェイとエッジに，重みを多くキャッシュ容量を分布させることであった．

27

5.結論

28

本研究におけるまとめ

1. **ネットワークエッジにおけるビデオ共有サービストラフィックの重複度の解析**
   - YouTubeトラフィックがWebトラフィックの約1/8を占めている
   - YouTubeトラフィックの重複が50%存在している
   - YouTubeトラフィックの重複には2つの傾向が存在

2. **ネットワークエッジにおける重複トラフィック削減のためのIn-Networkキャッシュの検討**
   - 提案手法である，In-Networkキャッシュ手法により，インターネットのトラフィック(問題1,2)だけでなく，エッジネットワークのトラフィック(問題3)にも高い削減効果が得られるキャッシュ手法を提案・評価した
   - キャッシュに総和量制限を設け，最適配置にした場合，手法5:UploadのIn-Networkキャッシュが最も効率的
   - キャッシュの最適配置は，ゲートウェイとエッジに，重みを多くキャッシュ容量を分布させることであった．

29

今後の課題

- トラフィック解析
  - 重複トラフィック解析の汎用化
    - YouTube以外のトラフィックの重複解析
  - 重複トラフィックのリアルタイム解析
    - 事前にキャプチャされたデータではなく，リアルタイムに解析を実行
- In-Networkキャッシュ
  - In-Networkキャッシュにおける双方向のキャッシュ移動
    - 手法5でのキャッシュのアップロードにダウンロード方向のキャッシュ移動を追加し，削減効果の検証を行う
  - 実環境での実装と検証
    - ClickやOCTEON上で実装をおこない，実環境での検証
  - 仮想化環境での検証
    - 実装したシステムをスライスの1つとして展開

30