

平成 21 年 6 月 9 日現在

研究種目：若手研究(B)
 研究期間： 2007～2008
 課題番号： 19700076
 研究課題名(和文) インターネット上の組織の相対的なトラフィック量推定に関する研究
 研究課題名(英文) Modeling of Traffic Demand of Node in Large-scale Network
 研究代表者
 福田健介(FUKUDA KENSUKE)
 国立情報学研究所・アーキテクチャ科学研究系・准教授
 研究者番号：90435503

研究成果の概要：

本研究課題では、インターネット上のノード(組織や個人)の生成するトラフィック量がどのような分布で特徴付けられるかを、複数のインターネットトラフィックデータから解析した。その結果、ノードのトラフィックは既存の指数分布モデルとは異なり、裾の広い対数正規分布でモデル化する必要があることが明らかになった。また、そのような分布が、トラフィック量の指数的增加により説明可能であることを明らかにした。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	2,000,000	0	2,000,000
2008年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総計	3,200,000	360,000	3,560,000

研究分野：総合領域

科研費の分科・細目：情報学 計算機システム・ネットワーク

キーワード：情報ネットワーク，インターネットトラフィック

1. 研究開始当初の背景

負荷分散，輻輳制御，P2P，経路制御等の，インターネット上で使用される各種プロトコルの性能を大規模ネットワークで評価を行うためには，一般にトポロジやネットワークの負荷を何かしらのモデルを仮定することが一般的である。しかしながらインターネットはネットワークのネットワークと呼ばれることからわかるように，その瞬間瞬間のトポロジや負荷のスナップショットを取ることが原理的に不可能である。近年，ネッ

トワークトポロジの研究において，インターネットトポロジは，従来仮定されてきたランダムグラフよりも，多様性・頑強性の高いスケールフリーネットワークやスモールワールドネットワークとして特徴付けられることがわかってきた。これらのモデルの示唆する点としては，ネットワークトポロジの統計性の違いが上位サービスの性能に大きな影響を及ぼすことが挙げられる。しかしながらネットワークトポロジの統計的性質が明らかになったとしても，ネットワークトポロジ上のノードにおいて，各ノードが送受信する

トラフィック量に関する知見がなければ、各種シミュレーションにおいて、正しいシミュレーション結果が得られるとは言い難い。例えば下図のようなネットワークトポロジを考えてみると、各ノードのトラフィックがほぼ一様(もしくは指数的)であるか、大きな偏りがあるかを知ることが直感的にも理解できる。

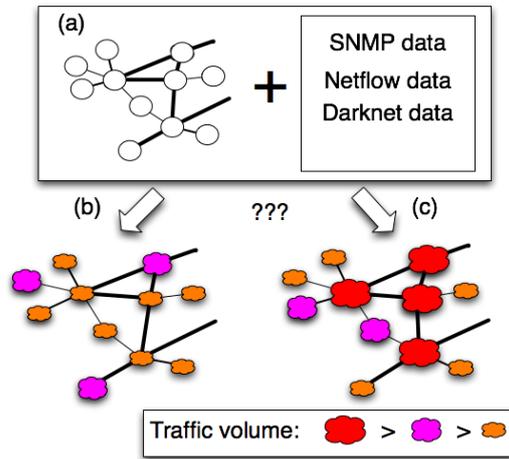


図1: ノードレベルトラフィック分布

2. 研究の目的

本研究では、背景で述べた、「ネットワークトポロジ中の各ノードのトラフィック需給」について着目した。すなわち、(1) ネットワーク上のノードが果たして、均一なトラフィック需給を持つのか、もし偏りがあるとするれば、それは、正規分布のような裾の短い分布となるのか、もしくは、さらに裾の広い分布として特徴付けられるのかを明らかにすることである。(2) 同様に、ノードごとのトラフィック需給の分布が明らかになった場合に、その時間発展を予測可能な単純な物理モデルを明らかにすることを目的とする。

3. 研究の方法

インターネット上のネットワークトポロジを考える際には、複数のレベルの抽象化が考えられる。一つは広域経路制御の単位であるASレベル、もう一つはAS内のノードレベルである。本課題では、主としてAS内のノードレベルトラフィック分布を同定するために、3種類のトラフィックデータを収集・解析した。

- (1) 学術情報ネットワーク(SINET)の加入機関単位(大学、短大、高専、研究機関等)の年間トラフィック量(SNMPにて収集)
- (2) 国内ISP 1社の加入者(ユーザ)ごとの1日当たりの使用トラフィック量(sampled

netflowにて収集)。データ収集期間は、2004年、2005年、2006年、2007年のそれぞれ1週間である。

(3) 国内設置の/18サイズのダークネットへ到着したトラフィック(tcpdumpにて収集)。ダークネットは経路広報されているものの実際にはホストの存在していないネットワークであり、到着するパケットは全て異常なものである。

4. 研究成果

(1) ノードレベルトラフィックの解析

図1はSINETの組織レベルでのトラフィック量を累積グラフで表現したものであり、このグラフは各ノードの入出力トラフィックがどのような偏りを持つかを表している。注目すべき点は水平軸方向が対数スケールになっていることである(なおトラフィック量は規格化してある)。すなわち、ノードレベルトラフィックは、非常に幅の広い分布となっており、既存の一般モデルや指数モデルとは大きく異なっている。このような幅の広い分布はいくつか知られているが、本研究では、ベキ分布、ならびに対数正規分布によるフィッティングを行った。図を見ればわかるように、トラフィック量は、ネットワークトポロジの統計的性質として現れるベキ分布よりも、対数正規分布によるフィッティングの方が当てはまりが良い。つまり、トラフィックの場合、大多数のノードが少量のトラフィックで特徴付けられ、少数のノードが多量のトラフィックで特徴付けられるモデルとは異なることを意味している。

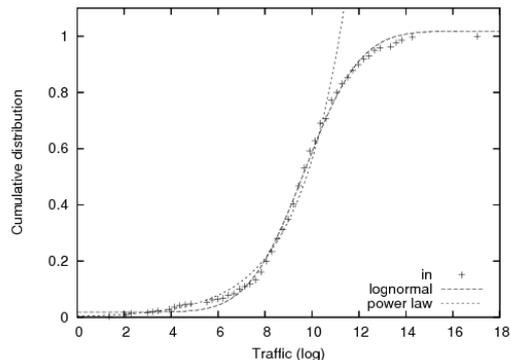


図2: ノードレベルトラフィック(組織単位)

次にISPで得られた各加入者単位でのトラフィック使用量をプロットしたものを図2に示す。この図より加入者単位のトラフィック分布もまた、上り下り共に裾の広い対称的な分布となることがわかる。しかしながら図を見ると、2005年のデータには2つの山があり、2008年のデータではその山の差があまりはっきりしなくなっていることが読み取れる。つまり、2005年のユーザのトラフィック分布

は、2つのタイプのユーザから構成されていると言える。前者のピークは従来のサーバ・クライアント型のサービス(web, mail等)を主として利用していることから、既存の大多数のユーザはこのタイプ(CSタイプ)に分類される。それに対して後者のピークは主としてP2Pサービスを利用している少数のユーザに相当する(P2Pタイプ)。P2Pタイプのユーザはユーザ数で見るとたかだか4-5%にすぎないが、水平軸が対数表示であることからわかるように、使用トラフィック量の約75%に対応するとの結果を得ている。同様に、2005年および2008年の二つの図を比較してみるとCSタイプのピークは右方向にシフトしているものの分布の形自体には変化がないことが読み取れる。これは、トラフィックの成長モデルが時間発展に対して、同一の関数形で表現可能であることを意味している。それに対して、P2Pタイプのピークは2008年にはピークが顕著ではなくなっている。これは、ユーザ数の減少というよりは、CSタイプの裾が右方向にシフトしているため、ピークがわかりにくくなっているものと考えられる。また、図中の右側に相当する大量のトラフィックには、アクセスリンクによるボトルネックが存在する。そのため現状の技術では、これ以上大量のトラフィックをユーザが生成することは原理的にないため、将来の観測では、この分布にある種の歪みが生じることになると予想できる。実際、P2Pタイプの最頻値はほとんど変化がない。

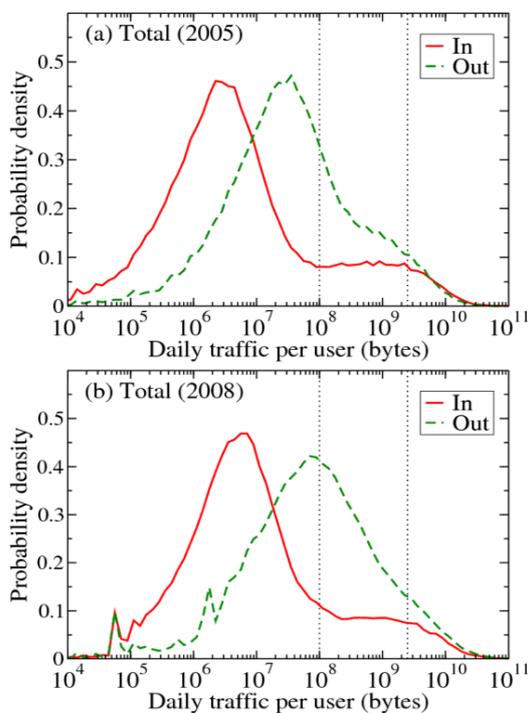


図3: ノードレベルトラフィック分布(加入者単位)

併せてダークネットに到着したパケットの送信元アドレスに基づいてASレベルで集

約したパケット数の分布を調査した。その結果、分布は、裾の長い分布となるものの、その関数形は対数正規分布よりもベキ分布に近いとの結果が得られた。これは、ウイルス等に感染しているホスト数が、AS内のホスト数に比例していると考えれば、AS内のホスト数に関係した指数となると言える。しかしながら、感染ホストがパケットを送信するタイミングに大きな偏りがあるため、分布が対数正規分布からずれたものになっていると考えられる。ダークネットに関するさらなる解析は今後の主要な課題である。

(2) トラフィック生成モデル

上記解析結果より、AS内ノードレベルトラフィックは対数正規分布で表現可能であることが明らかとなった。本稿では、そのようなトラフィック量の分布を生成する単純なモデルを導入する。これは、multiplicative growth modelとして経済学では良く知られているモデルである(経済学では、所得の分布や会社の規模の分布を説明するモデルとなっている)。

初期状態ではノード i (ノード数 n) は一様乱数で振られた初期トラフィック量 $p_i(0)$ を持つ。これらのノードが時間発展の度に、一様乱数で割り当てられた成長率 $b_i(t) (> 1)$ を持つとすると、1ステップ後のトラフィック量は、 $p_i(1) = p_i(0) * b_i(0)$ と表すことができる。すなわち時間ステップ k では、 $p_i(k) = \prod (b_i(k)) * p_i(0)$ となる。両辺の対数を取ると、 $\log(p_i(k)) = \sum \log(b_i(k)) + \log(p_i(0))$ である。このようなトラフィックを持つノードが n 個存在する状況を考えると、この対数を取ったものの分布は正規分布となる。よって、オリジナルの分布は対数正規分布である。

このモデルの前提はトラフィックの成長が指数的事であることだけであるが、我々の観測では、2種類のノードレベルトラフィックではその成長は指数的事であることが確認されており ($b \approx 1.3$)、このモデルの前提条件は満たされていることがわかった。

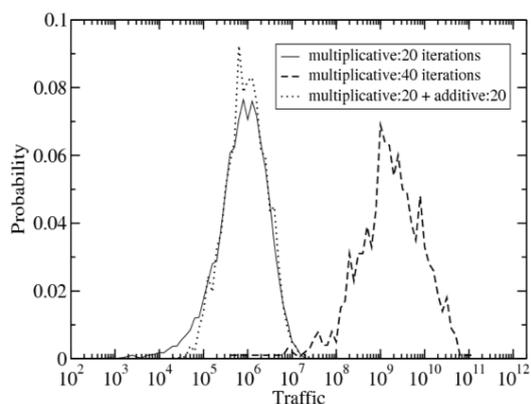


図4: シミュレーション結果

実際にシミュレーションでトラフィックの分布を調べた結果を図4に示す。シミュレーションパラメータは、トラフィックの初期値：1~1000の乱数、ノード数：1000、増加率1.0~2.0とした。図中の3つのプロットは下記の3つのシナリオに対応する。(1)20回の指数増加モデルの反復、(2)40回の指数増加モデルの反復、(3)20回の指数増加モデルの反復の後、20回の加算的増加モデル($p_i(t) = p_i(t-1) + a_i(t-1)$)の反復。

実験結果を見ると、シナリオ1でまず実際のトラフィック分布で現れるのと同じような裾の長い釣り鐘状の分布が現れることがわかる。これは、トラフィック成長の反復が比較的少なくてもトラフィック量は対数正規分布に近いものが得られることを表している。また、シナリオ2では、時間発展とともに、分布の最頻値が右方向へシフトしていくことがわかる。これは、加入者レベルトラフィックの時間発展で得られた結果と似たものとなっている。最後にシナリオ3であるが、これは、将来的にトラフィック量の成長が緩やかになりほぼ加算的になった場合の想定シナリオである。この結果は、分布の左側部分のトラフィックの少ない部分には多少の違いがあるが、おおむねシナリオ1の結果とほとんど同じであることがわかる。つまり、将来的にトラフィックの成長が止まったとしても、インターネットトラフィックの分布が一度対数正規分布になったとすると、他の分布に変化する可能性が少ないことを表している。

以上のシミュレーション結果より、現在、および近未来のトラフィック量分布を必要とするシミュレーションでは、その分布を対数正規分布で表現することに妥当性を与えるものであると結論づけられる。

現在のインターネットトラフィックを生成している主要なアプリケーションは、P2Pアプリケーションによるものである(我々の調査ではバックボーントラフィックの80%程度がP2Pと推定)。それに対して、ビデオストリームやgoogle mapに代表されるような広帯域のネットワークを必要とするアプリケーションの需要が高まっていることも指摘されている。ISP加入者トラフィックのデータにせよ、学術加入機関のトラフィックにせよ、多数を占めるライトユーザは、P2Pトラフィックの生成には関与していないが、今後はこれらの広帯域利用アプリケーションの普及とともに、徐々にではあるが、トラフィックを生成するノードへと変化していくと予想されるため、シナリオ3で示した楽観的な状況になる見込みは今のところ低いと思われる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1件)

- ① 福田 健介, 佐藤 進也, 明石 修, 廣津 登志夫, 栗原 聡, 菅原 俊治, ネットワークボロジの次数情報に着目したサーバ・クライアント負荷分散方式の提案と評価, コンピュータソフトウェア, vol. 24, pp. 78-87, 2007. 査読有

[学会発表] (計 3件)

- ① 福田 健介, 廣津 登志夫, 明石 修, 菅原 俊治, 異常パケットトレースアドレス局所性に関する解析, 情報処理学会全国大会, 2007年3月14日, 査読無
- ② Kensuke Fukuda, Towards Modeling of Traffic Demand of Node in Large Scale Network, Proc. IEEE ICC2009, pp. 214-218, 北京, 2008. 査読有
- ③ Kenjiro Cho, Kensuke Fukuda, Hiroshi Esaki, Akira Kato. Observing Slow Crustal Movement in Residential User Traffic, Proc. ACM CoNEXT2008, p. 12, マドリッド, 2008, 査読有

[図書] (計 1件)

- ① 福田 健介, P2P 教科書, インプレス社, pp.280-297, 2007

[その他]

ホームページ等

<http://www.fukuda-lab.org/>

6. 研究組織

(1) 研究代表者

福田 健介 (FUKUDA KENSUKE)

国立情報学研究所・アーキテクチャ科学研究系・准教授

研究者番号：90435503