

異常トラフィック測定分析手法

ネットワークリソースの浪費や品質劣化を引き起こす異常トラフィックを検知・制御する技術は、安心して快適な通信サービスを提供するために不可欠です。本稿では、パケットサンプリングにより得られるフロー測定情報を用いて、異常トラフィックを検出するためのトラフィック測定分析手法について紹介します。また、実データ評価結果を交えてその有効性も示します。

かわはら りょういち^{†1} もり たつや^{†1}
川原 亮一 / 森 達哉

はらだ しげあき^{†1} かみやま のりあき^{†1}
原田 薫明 / 上山 憲昭

こんどう つよし^{†2} いしばし けいすけ^{†2}
近藤 毅 / 石橋 圭介

NTTサービスインテグレーション基盤研究所^{†1}
 NTT情報流通プラットフォーム研究所^{†2}

パケットサンプリングによるフロー測定

インターネットトラフィックの増加とインターネットの利用形態・アプリケーションの多様化に伴い、ネットワークの効率的な運用のためにはトラフィック測定が重要となります。特に、品質劣化の要因となるネットワークリソースの浪費や、セキュリティ上の問題を引き起こす異常トラフィックを検出できる仕組みへの重要性が増しています。

このねらいのため、フローレベルのトラフィック測定が近年着目されており、異常トラフィック検出、ヘビーユーザ特定、トラフィックエンジニアリング等への応用が検討されています。ここでフローとは、発信元IPアドレス (srcIP)、着信先IPアドレス (dstIP)、発信元ポート番号 (srcPort)、着信先ポート番号 (dstPort)、プロトコル (protocol) の5つ組を同じくするパケット群のことを指します。

フロー統計情報を取得するためには、監視対象ネットワークですべてのパケットをキャプチャして解析できればよいのですが、回線速度の高速化により、スケーラビリティの問題が大きくなるため、パケットをサンプリングするフ

ロー測定法が注目されています。図1がサンプリングのイメージです。図1上段のように全パケットをキャプチャすればすべてのフロー情報 (どのフローが何パケット送出しているか) を正確に把握できますが、下段のようにN (=3) 個に1個のサンプリングをすると、処理すべきパケット数やフロー数は削減できます。その一方で必要な情報が失われる可能性があります。そこで、サンプリングの影響を考慮して異常トラフィックを適切に検出できる仕組みが必要となります。

従来のパケットサンプリングにより

得られるフロー分析手法は、これらが正常時におけるフロー統計推定を対象としていたのに対し、ここでは、サンプリングによる異常検出への影響を考慮しつつ、サンプルフロー情報を用いて異常トラフィックを検出する方法について提案します。

異常トラフィック分析の流れ

トラフィック分析の流れを図2に示します。まず、ネットワーク内の各ルータにおいて、パケットサンプリングを実施し、サンプルフロー情報を測定します。ここで、サンプルフロー情報とは、

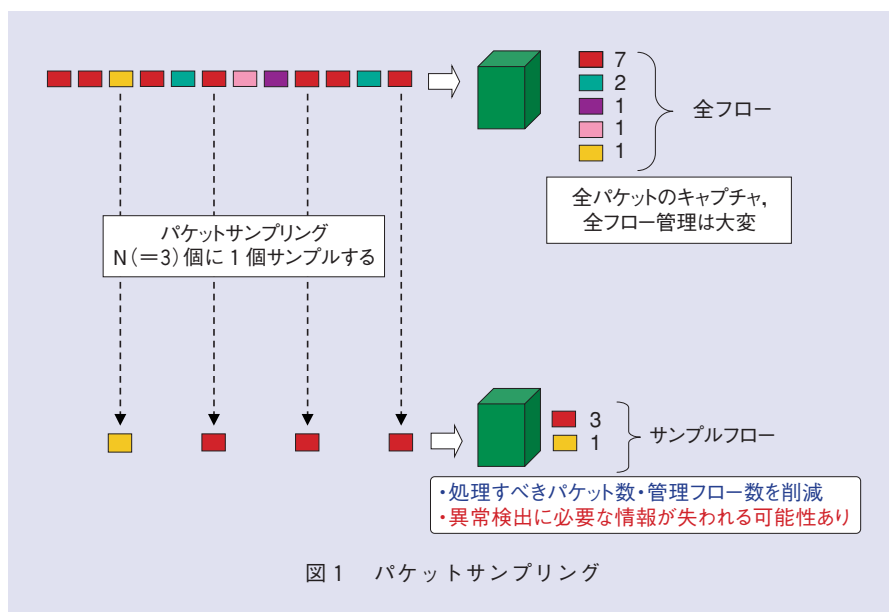


図1 パケットサンプリング

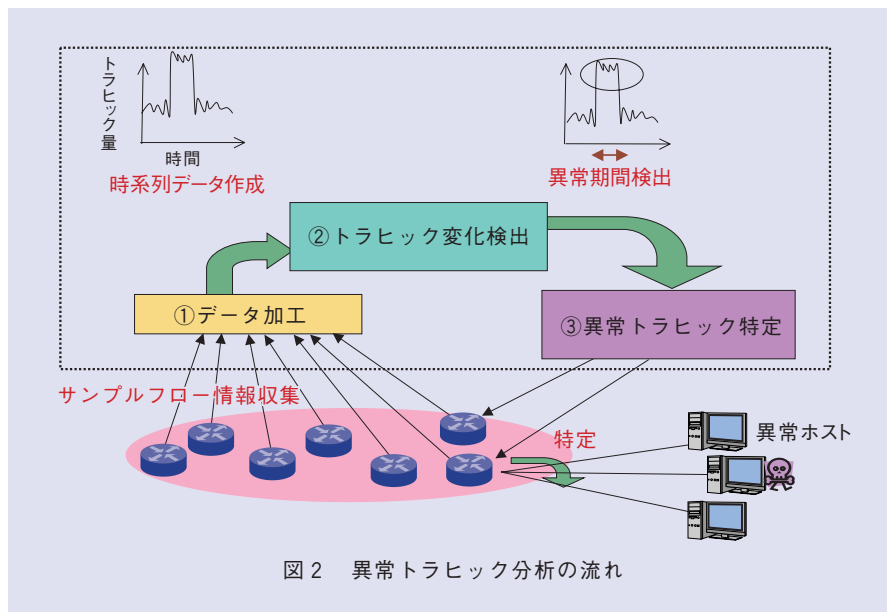


図2 異常トラフィック分析の流れ

どのフローから何パケットサンプルされたか、といった情報を指します。それからサンプルフロー情報を監視システムにおいて収集・加工します(図2①)。ここでは、あらかじめ定めたトラフィック監視単位(ルータごと、リンクごと、あるいは対地ごと)でフロー情報を集約し、トラフィック量の時系列を作成します。トラフィック量としては、単位時間当りの発生パケット数、バイト数、フロー数が考えられます。次に、この時系列データを対象に、トラフィックの急激な変化の有無をチェックし、もし変化を検出したらその異常が発生している時間区間(異常期間)を検出します(図2②)。そしてその変化を引き起こした異常トラフィックを特定します(図2③)。

次に、異常トラフィック分析において必要となる技術について説明します。

技術1：異常検出精度向上のためのトラフィック分割監視法

図2①で必要となる技術を説明する前に、サンプリングが異常検出にどのような影響を及ぼすか、実データによる評価結果を交えて説明します。あるネットワークで観測されたフロー数の時系列に対し、ネットワークスキャンを模擬した擬似トラフィックを付加したシミュレーション結果を図3(a)に示します。これより、サンプリング前にはスパイク状のトラフィック変化が見とれます。

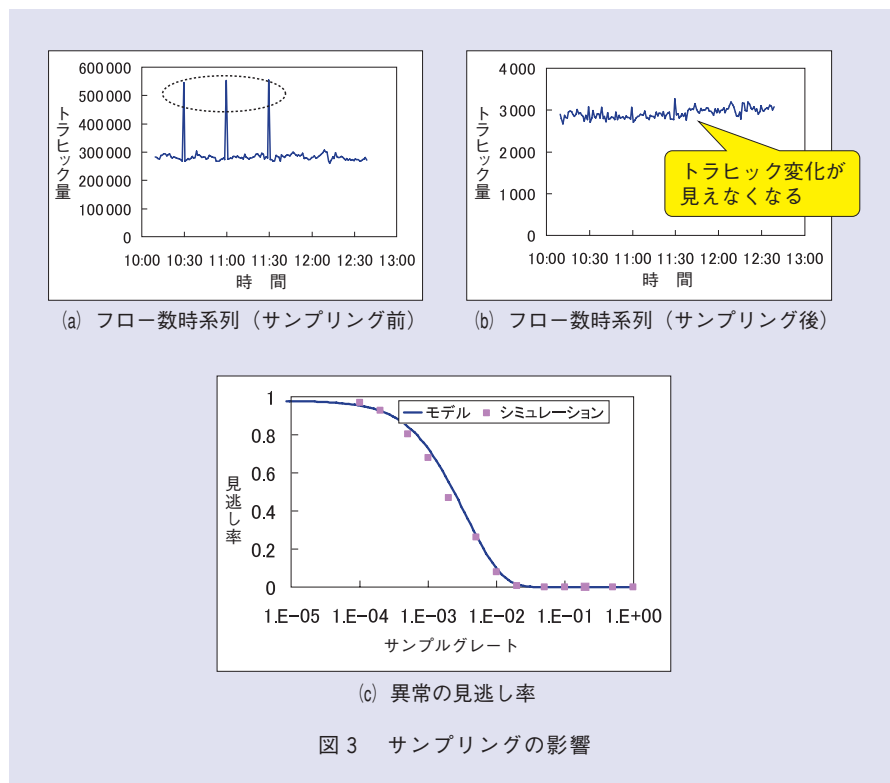
また、図3(b)に、パケットサンプリングを行った場合の結果を示します。ここでは、1000個に1個のパケットサンプリングを行いました。これより、サンプリング前には確認できた異常トラフィックが見えなくなっていることが分か

ります。この理由は、スキャンのような異常トラフィックは非常にたくさんのフローを発生するのですが、個々のフローは少ないパケット数で構成されているため、パケットサンプリングするとサンプルされにくいからです。

さらに、図3(c)に、サンプリングレート p ($=1/N$, N :サンプリング間隔)を変えたときの異常の見逃し率に関する評価結果を示します。なお、図中の「モデル」は、以下に述べる解析モデルを用いて計算した結果です。

解析モデルでは、各フローからのパケットは確率 p でランダムにサンプリングされるとします。測定区間 t でのサンプルフロー数を $N_t(p)$ とし、 $N_t(p)$ があるしきい値(後述)を超えたら異常発生と判定するとします。なお、 $N_t(p)$ は、正常時は $N_t(p) = Nn_t(p)$ 、異常発生時は $N_t(p) = Nn_t(p) + Na_t(p)$ で与え、 $Nn_t(p)$ はサンプル正常フロー数、 $Na_t(p)$ はサンプル異常フロー数とします。 $Nn_t(p)$ は、平均 $m_n(p)$ 、分散 $\sigma_n(p)^2$ の正規分布に従うとし($f(x)$ をある正常フローが x パケットから成る確率とし、 $m_n(p) = \sum \{1 - (1-p)^x\} \times f(x) \times m_n(1)$ 、 $\sigma_n(p)^2 = \Phi \times m_n(p)^c$ を計算)、一方、異常時に加わるフロー数を d 、異常フロー当りパケット数を 1 とし、 $Na_t(p)$ を平均 dp 、分散 $dp(1-p)$ の二項分布でモデル化しました。

以上のモデルで、異常検出しきい値を $m_n(p) + 3\sigma_n(p)$ としたときの異常見逃し率を評価しました。図3(c)よ



り、サンプリングレートが小さくなると見逃し率が大きくなっていることがわかります⁽¹⁾、⁽²⁾。

上記の問題を回避するため、トラヒックの分割監視法を検討しました。図2①においてサンプルフローをリンクごとといった監視単位で集約して時系列データに加工する際に、その集約トラヒックを以下の手順で分割監視します(図4)。フローを、IPアドレス等をキーに複数のグループに分類し、グループごとに時系列データを作成します。もし異常トラヒックのあるグループに集中できれば、そのグループでの異常フロー数の占める割合が増加し、異常検出精度の向上が期待されます。

実際に、図3(b)で示したサンプルフロー数データを、16分割したときの結果を図4に示します。これより、サンプル前には検知可能だった異常トラヒックが、分割することにより再び検知可能になっていることがわかります。

技術2：トラヒック変化検出法

トラヒック変化検出の際、オペレータの負担軽減のために検出しきい値を自動的に設定できる仕組みが必要となります。トラヒック変化検出法としては、本特集『大規模ネットワーク向け異常トラヒック監視システムの開発』で紹介した手法が有効です。その他の手法として、カルマンフィルタと呼ば

れる時系列解析手法を用いて、トラヒックの時系列を予測し、その予測値を元に検出しきい値を自動的に設定する方法についても検討しています⁽³⁾。

カルマンフィルタを使うメリットは、トラヒックの増加・減少割合を時系列データから抽出し、次に現れるデータがどのくらいの大きさになるかを計算できることにあります。また、カルマンフィルタを使えば、予測値と時系列データの誤差も自動的に見積もることもできます。この予測値と予測の誤差を組み合わせれば、正常データとみなせる範囲(すなわちしきい値)の設定が可能となります。これはリアルタイムにトラヒックを監視し、異常値の検出をする場合には特に有効です。

技術3：異常トラヒック特定法

技術1で分割監視したトラヒックや監視システムでの交流単位のトラヒックを対象に、技術2でトラヒック変化を検出した後、監視単位のトラヒックから異常トラヒックのみを特定する必要があります。

従来、この特定作業はオペレータが手動で異常期間のtop-Nフロー情報などを用いてドリルダウンしていましたが、次のような問題がありました。

- ・異常期間のtop-Nフローが異常原因かどうか不明
- ・DDoSなどでは異常フロー数が膨大になるためそのフローを集約する必要があるが膨大なフローの集約は困難

ここでは、トラフィック量変化を引き起こしたトラフィック＝異常トラフィックとして、そのような異常トラフィックを構成するフロー（群）を特定する技術

を説明します⁽⁴⁾ (図5)。具体的には、図5に例を示すフロー情報の5つ組の値および範囲を用いて識別 (flowIDの組で特定) します。

ここで、異常トラフィック特定の評価要素として、①正常トラフィックの巻き添え率 (後述)、②異常トラフィックのカバー率 (後述)、③異常flowID数、の3つについて考えます。異常トラフィックを特定し制御する際、正常トラフィックの巻き添えを避け (巻き添え率の低減)、異常トラフィックをできるだけ多く制御できる (大きいカバー率) 異常flowIDを特定する必要があります。一方、異常flowIDは、オペレータが識別し、最終的にはルータへのaccess control listなどの制御パラメータとして利用する情報であるため、異常flowID数は少なく保つ必要があります。

この3つの尺度を考慮した異常トラフィック特定アルゴリズムDELTA (delta traffic auto aggregator) の特徴は、次の2点です。

- ・ 正常期間、異常期間トラフィックの差分分析による異常トラフィック抽出
- ・ ツリー構造を用いた異常flowの最適集約

ここでは、srcIPのみに着目して特定する手順を図6を用いて例示します。

- ・ 手順1：出現srcIPごとに、正常・異常期間各々のトラフィック量をカウントし、差分を求めることによって出現srcIPごとの差分トラフィック量を求める。
- ・ 手順2：出現srcIPごとの差分トラフィック量と、正常期間のトラフィック量から、巻き添え率^{*1}とカ

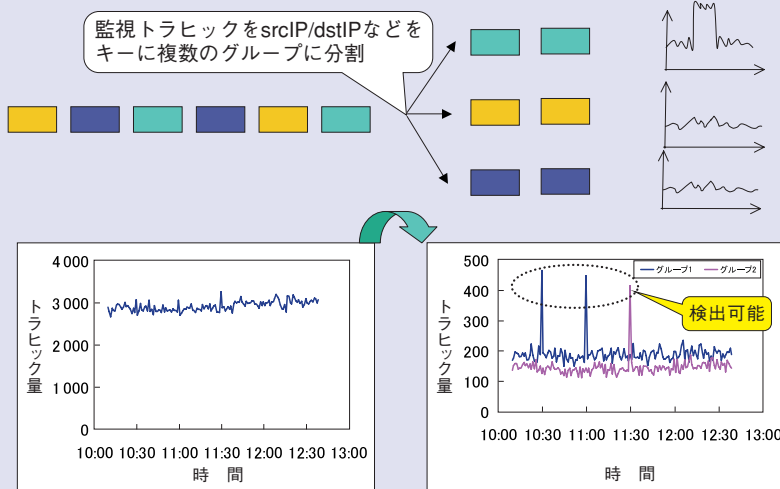
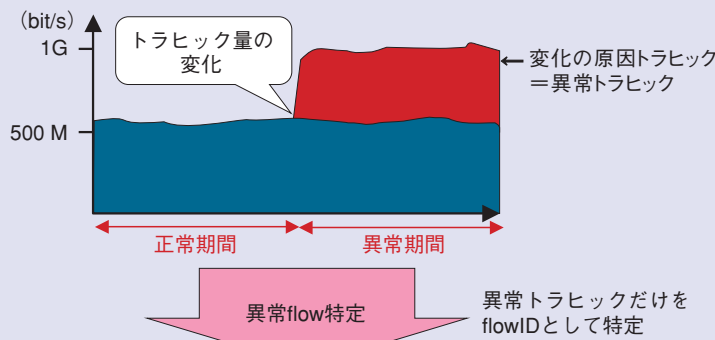


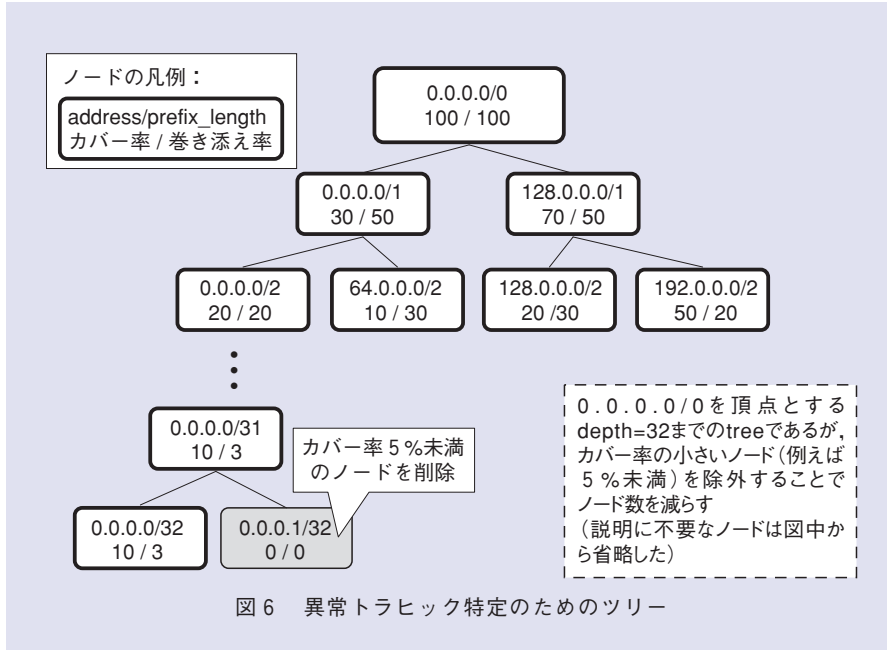
図4 トラフィック分割の効果



flowIDの例：
本提案では5つ組およびその範囲を用いてトラフィックを識別する

	src_ip	dst_ip	proto	src_port	dst_port
flow_ID (1)	10.0.0.0/8	192.168.0.1/32	6	*	80
flow_ID (2)	202.1.0.0/16	192.168.0.1/32	6	*	80
flow_ID (3)	128.0.0.0/16	192.168.0.1/32	6	*	80

図5 異常トラフィック特定技術



ラヒック分析法について必要となる各技術を紹介しました。今後は、実データを用いたさらなる有効性評価を行っていきたく考えています。

■参考文献

- (1) K. Ishibashi, R. Kawahara, T.Mori, T. Kondoh, and S. Asano : "Effect of sampling rate and monitoring granularity on anomaly detectability," 10th IEEE Global Internet Symposium 2007,Anchorage,USA, May 2007.
- (2) R. Kawahara , K. Ishibashi, T. Mori, N. Kamiyama, and S. Asano : "Detection accuracy of network anomalies using sampled flow statistics," IEEE Globecom 2007, Washington, D. C., USA, Nov. 2007.
- (3) 原田・川原・森・上山・廣川・山本 : "異常トラヒック発生検出および終了判定手法," 信学技報, Vol.106,No.420, IN2006-133, pp.115-120, 2006.12.
- (4) T. Kondoh and K.Ishibashi : "Identifying the Anomalous Traffic Using Delta Traffic," CERT FLoCon 2008, Savannah, USA, Jan. 2008.

パー率*2を計算する。

- ・手順3 : prefix長を木のdepthとしてtreeを構成し、各ノードでのカバー率、巻き添え率を計算する。
- ・手順4 : カバー率一定 (例えば、5%未満) のノードを除外し、残ノードをクラスタとする。これにより、異常トラヒックと無関係なノードがtreeから除外され、計算量が削減される。
- ・手順5 : tree上の全ノードの組み合わせに対し、以下の評価値*3を

*1 巻き添え率=異常flowIDにマッチする正常期間のトラヒック量/正常期間の総トラヒック量
 *2 カバー率=(異常flowIDにマッチする異常期間のトラヒック量 - 同IDにマッチする正常期間のトラヒック量) / (異常期間の総トラヒック量 - 正常期間の総トラヒック量)
 *3 評価値 = (αカバー率 - β巻き添え率) / 異常flowID数

最大にするノードの組み合わせをsrcIP次元の異常flowIDとして特定する。

以上の手順をdstIPやポート番号等に対しても実施し、それら出力の積に対して再度評価を行うことで、最終的な異常flowIDを特定します。

なお、本特定技術はWINDSアーキテクチャにおける分析フェーズを担う分析装置Nail-pxとして試作を行っています。本装置は本特集『インターネットトラヒック制御システム (i-TRACS) の開発』で解説するi-TRACSからの特定指示に基づき特定を実行し、特定結果のflowIDをルータへの設定情報としてi-TRACSに返答します。

今後の予定

サンプルフロー情報を用いた異常ト



(後列左から) 石橋 圭介/ 近藤 毅/
 上山 憲昭/ 森 達哉 (右上)
 (前列左から) 原田 薫明/ 川原 亮一

安心して快適な通信サービスを提供するためのトラヒック測定管理技術に関する研究を進めています。

◆問い合わせ先

NTTサービスインテグレーション基盤研究所
 E-mail kawahara.ryoichi@lab.ntt.co.jp