トラヒック特性に基づく近似機能を有する空間分割型パケットキャプチャシステムに関する研究

M2005MM017 中村 陸
指導教員　野呂 昌満

1　はじめに

ネットワークの安定運用のためには, 悪意のあるホストからの攻撃や, ネットワーク機器の動作を監視・診断することが不可欠である．監視・診断の方法としてパケットを収集し, その中身から状態を把握するパケットキャプチャシステムが広く使われている．

しかし, 高速なネットワークでは膨大な量のパケットが短時間で流れることがある．この場合, パケットキャプチャシステム内での転送速度やディスクの書き込み速度, 記憶装置の容量といった制限により, 全てのパケットをキャプチャし, 監視・診断を行うのは困難である．

この問題を解決するために, ネットワークのインターフェースでフィルタリング技術 [1][2] を用いて監視・診断に必要なパケットだけを選択してキャプチャする方法が有効になる．この方法を用いればパケットキャプチャシステムが処理するトラヒック量を減らす事ができる．しかし, 高速なネットワークでフィルタリングを行うにはフィルタリング速度も高速でなければならない．

一方, 多数のフィルタに対して高速なフィルタリングが行える方式として空間分割型のパケット分類方式がある [1][2]．しかし, フィルタの数が多くなると多量のメモリが必要となり, 高コストになる．

この問題の解決法としてフィルタリングポリシを変更して必要なメモリ量を減らす手法 [4] がある．しかし, フィルタリングポリシを変更すると本来キャプチャする必要のないパケット (**ノイズ**と呼ぶ) までキャプチャることがある．ノイズを低減するにはどのフィルタをどのように変更するかが重要となる．変更するフィルタの決定法として分類空間におけるフィルタの変更量に基づく決定法 [4] がある．しかし, この方式ではパケットの分布に偏りがあるとノイズが上昇してしまう．

そこで本稿では, 以下の特徴を持つパケットキャプチャシステムを提案する．

特徴 1 パケット空間の各格子点に対するパケット出現頻度を推定する機能を実現する．これにより, フィルタを変更により生じるノイズの量を推定可能となる．

特徴 2 パケット出現頻度に従ってフィルタリングポリシを変更する機能を実現する．これにより, メモリを任意の容量以下に抑えつつ, ノイズの低減が可能となる．

2　空間分割型パケット分類器

2.1　機能概要

パケット分類器は, パケットのヘッダの値に従い,IP パケットを識別する [3]．ここで, ヘッダは送信先 IP アドレス, 送信先ポート番号など, それぞれ長さ一定の複数のフィールドからなる．このうち, パケット分類器で識別に用いるフィールドを**キー**と呼び, 各キーが満たすべき条件の記述を**フィルタ**と呼ぶ．フィルタの集合を**フィルタリングポリシ**と呼ぶ．

本稿ではパケット分類器を次のように動作するものとする．すなわち, フィルタリングポリシとパケットが与えられた時, パケットの各キーとフィルタリングポリシの各フィルタとを照合する．そして, あるパケットに対し, 全てのキーが対応する条件との照合に成功するフィルタが 1 つ以上あればそのパケットをキャプチャする．

2.2　パケット空間を用いたパケット分類器の表現法

パケット分類問題は, 計算幾何学的なアプローチにより多次元空間における点位置決定問題として捉える事ができる [1][2]．この問題ではまず, キーの数を N とした時, それぞれのキーを軸とする N 次元空間 (**分類空間**と呼ぶ) を考える．全てのパケットをこの N 次元空間の 1 点で表す．分類空間において, フィルタは条件を満たす全ての点を包含する空間 (**フィルタ領域**と呼ぶ) として表す．また, フィルタリングポリシは全てのフィルタ領域の和集合となる．分類空間においてパケットを示す点を含むフィルタ領域が存在するとき, 照合に成功するといい, パケットをキャプチャする．

送信元 IP アドレス (以下 SrcIP,$0 \leq \mathrm{SrcIP} \leq 2^{32}-1$) と送信先 IP アドレス (以下 DstIP,$0 \leq \mathrm{DstIP} \leq 2^{32}-1$) をキーとする分類の例を示す．このとき, 以下の 2 つのフィルタ F1,F2 がある時, 分類空間は図 1 となる．

F1:$4 \leq \mathrm{SrcIP} \leq 14$ かつ $10 \leq \mathrm{DstIP} \leq 16$

F2:$8 \leq \mathrm{SrcIP} \leq 16$ かつ $3 \leq \mathrm{DstIP} \leq 12$

上の例では簡単のため SrcIP と DstIP のみを用いたがキャプチャ条件で使用する全てのキーに対し軸を作成し, パケットがフィルタ領域に包含されるか調べる．

SrcIP の軸では,F1 は [4,14] で F2 は [8,16] の区間となる．F1,F2 の重なり具合により全区間を F1 のみの区間 [4,7],F1,F2 の区間 [8,14],F2 のみの区間 [15,16], 空の区間 [0,3][17,2^{32}-1] の 5 区間に分割できる．重なるフィルタの集合が等しい区間の集合を**等影区間**と呼ぶ．

2.3　フィルタ領域の近似手法

我々が先に提案した高速パケットキャプチャシステム [3] では分類空間のある軸の各座標と, その座標を含む等影区間を示す識別子との対応表 (分類表) をオンチップのメモリに格納し, パケット到着時にはその表を探索するのみで高速にパケットを分類する．この分類表を少ないメモリ容量で格納するために, 近似を用いて表のエントリ数を一定以下に抑える [4]．

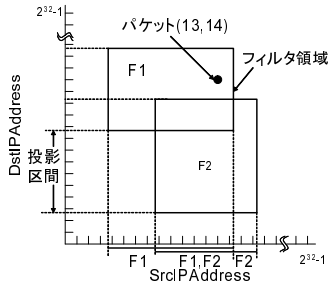

図 1　分類空間の例

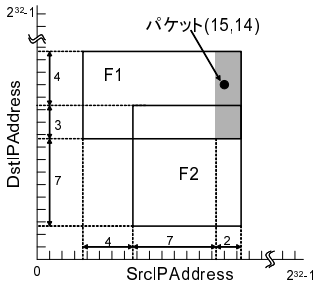

図 2　F1 を右に拡大

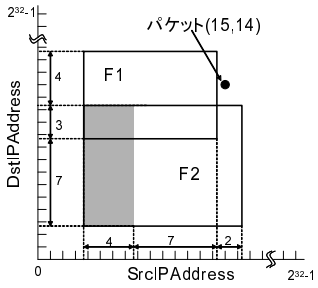

図 3　F2 を左に拡大

2.3.1　分類条件の近似

分類空間から作られる分類表のエントリ数は等影区間の数に依存する．近似的空間分割型パケット分類器では等影区間の数が減るようにフィルタ領域を変形することで分類表のデータ量を小さくする．

ある軸で等影区間の組 A と B を 1 つの等影区間にするには, それぞれが持つフィルタ集合を等しくすれば良い．そこで, まず A が持っていて B が持っていないフィルタの集合を求める．このフィルタ集合を B の等影区間に拡大する．同様に B が持っていて A が持っていないフィルタの集合も A 側に拡大する．このように拡大した各フィルタ領域の拡大部分を**変化領域**と呼び, 2 つの等影区間を 1 つの等影区間とするようにフィルタ領域を拡大する操作を**フィルタ領域の近似**と呼ぶ.

フィルタ領域の近似を一回行うと, 投影区間の数が 1 つ減る．そのため, フィルタ領域の近似を繰り返すことで等影区間の数を一定以下に抑えることができる [4].

フィルタ領域の近似を行うと, 元はフィルタ領域が無かった場所にフィルタ領域が拡大されることがある．この領域に現れるパケットがノイズとなる.

2.3.2　容量に基づく変化領域決定法

部分空間の容量に基づく決定法 [4] では, 部分空間に含まれる格子点の総数を部分空間の**容量**と定義する．そして, 変化領域の容量が最小となるフィルタ領域を近似する．もし, 全ての格子点にパケットが等確率で出現するならば, 変化領域が小さいほど, パケットの到着する確率が低くなるので, ノイズの低減が可能である.

図 1 の分類空間で投影区間を 1 つ減らす例を示す．この分類空間では以下の 2 つ変化領域が考えられる.

D1:F1 を右に拡大したときの変化領域 (図 2 の網掛部)
D2:F2 を左に拡大したときの変化領域 (図 3 の網掛部)
これらの変化領域のうち, 変化領域の容量が小さい D1 が近似するフィルタ領域として選択される.

3　提案システムの実現法

3.1　提案システムの概要

2.3 節で述べた近似法では全ての格子点においてパケットが等確率で出現するとした．しかし, 実際のネットワークでは格子点によってパケットの出現確率が異なる．そのため, 変化領域の容量が小さくても大量のノイズが発生することがある.

2.3 節の例では図 1 の F1,F2 のフィルタに対し,D1 の変化領域が選択された．ここで, 送信元アドレスが 15 かつ送信先アドレスが 14(図 2 の格子点 (15,14)) のパケットが多数到着した場合を考える．この場合,D1 は D2 より容量は小さいが多量のノイズが発生する.

そこで, 提案システムでは事前にキャプチャしたトラヒック特性に基づいてフィルタリングポリシを変更する．ここで, 変更した結果できるフィルタリングポリシを**近似フィルタリングポリシ**と呼ぶ.

提案システムでは格子点の重みをその格子点に対するパケットの出現頻度に比例した値として定義する．また, 部分空間に含まれる格子点の重みの総和を部分空間の**重み付き容量**と呼ぶ．部分空間の重み付き容量が小さいことはその部分空間にパケットが出現する確率が低いことを意味する．この重み付き容量の最も小さい変化領域を持つフィルタを選択し, 近似する．この計算法を**重み付き容量に基づく変化領域決定法**と呼ぶ.

提案システムは図 4 に示すようにフィルタコンパイラ, パケット分類器, 近似フィルタリングポリシ生成器, ノイズアナライザからなる．フィルタコンパイラでは第 2 章で述べたパケット分類器のための分類表を作成する．パケット分類器は分類表に従いパケットをキャプチャする．ノイズアナライザはキャプチャしたトラヒックからノイズを分離し分析する．フィルタリングポリシ生成器ではノイズアナライザで分析したトラヒック特性に基づいて近似フィルタリングポリシを生成する.

3.2　近似フィルタリングポリシ生成器

事前にキャプチャしたトラヒックから各格子点毎のパケットの出現頻度を推定し, パケット出現頻度が高い格子点を変化領域に含まないよう近似するフィルタを決定する．ここで, 格子点毎のパケット出現頻度は次の式で表される単位時間当たりのパケット到着数とする.

$$\text{パケット出現頻度} = \frac{\text{パケット到着数}}{\text{観測時間}} \qquad (1)$$

ネットワークの構成やサービスに変化が無ければ, キャプチャされたトラヒックが持つ特性が今後も続くと考えられるので, 過去にパケット出現頻度が高かった格子点は今後もパケット出現頻度が高いと考えられる.

重み付き容量に基づく変化領域決定法では以下の手順により, 近似する変化領域を決定する.

Step1　全てのフィルタの組み合わせについて近似する

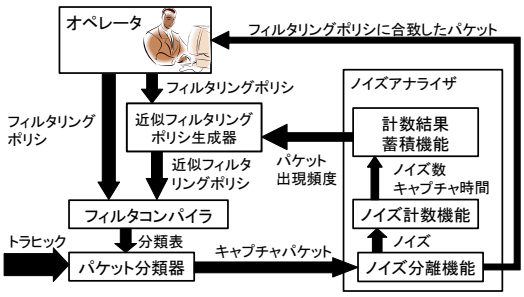

図 4　提案システムの構成図

とどのような変化領域になるか求める.

Step2　変化領域に含まれる格子点を求める.

Step3　変化領域に含まれる格子点のパケット出現頻度の合計を求め, 重み付き容量とする.

Step4　重み付き容量が最小の変化領域を選択する.

変化領域の重み付き容量が小さいことは, その領域に出現するパケットが少ないことを意味する. 従って, 重み付き容量に基づく変化領域決定法を用いることでノイズの低減が可能となる.

図 1 の分類空間において,SrcIP 軸の投影区間を 1 つ減らす例を示す. ここで, パケットが SrcIP=15 かつ DstIP=14 の点 (図 2 の格子点 (15,14)) のパケット出現頻度を α, 他の点のパケット出現頻度を β(ここで,$\alpha = 100\beta$) とする. この時, 変化領域は容量に基づく変化領域決定法と同様に D1(図 2),D2(図 3) が考えられる. D1 の容量は 14 である. このうち点 (15,14) の格子点の重みは α で, 他の格子点は β である. したがって,D1 の重み付き容量は $\alpha + 13\beta$=113β となる. 一方,D2 の容量は 40 で, この領域に含まれる格子点の重みは全て β である. したがって,D2 の重み付き容量は 40β となる. この結果, 重み付き容量の小さい D2 が選択される.

ここではフィルタが 2 つのみの場合を挙げたが, フィルタが 3 つ以上ある場合は全てのフィルタの組み合わせについて同様の操作を行い, 変化領域の重み付き容量が最も少なくなるようフィルタ領域を近似する.

3.3　ノイズアナライザ

ノイズアナライザはノイズ分離機能, ノイズ計数機能, 計数結果蓄積機能から成る. ノイズ分離機能ではキャプチャしたトラヒックからノイズを分離する. ノイズ計数機能では各格子点のパケット出現頻度を求める. 計数結果蓄積機能では過去に計測したノイズの計数結果と新たに計測したノイズの計数結果を合わせ, パケット出現頻度を求める. 以降, ノイズ計数機能の詳細について述べる.

ノイズ計数機能ではキャプチャしたトラヒックから, 各格子点のパケット出現頻度を求める. 各格子点でパケット出現頻度を求めるには, 各格子点についてキャプチャを行った観測時間とパケット到着数を記録する必要がある. しかし, 全ての格子点について記録を行うとデータ量が膨大になる. そこで, 提案方式では次の 2 つの方式を用いてデータ量を削減する.

(1) 軸毎のパケット出現頻度を用いたデータ削減法

この方式では, パケット到着数を軸毎に独立して記録する. ここで, 軸毎のパケット到着数を**軸パケット到着数**と呼ぶ. すなわち,SrcIP が a かつ DstIP が b の点 P にパケットが到着したとき, SrcIP が a の軸パケット到着数と DstIP が b の軸パケット到着数をそれぞれ 1 増やす. そして, 点 P のパケット到着数を SrcIP が a の点の軸パケット到着数と DstIP が b の点の軸パケット到着数の積とする.

(2) セル毎のパケット出現頻度を用いたデータ削減法

この方式では投影区間によって分類空間を分割し, 分割された部分空間毎にノイズを保存する. ここで, このように分割された部分空間のことをセルと呼ぶ. 図 1 分類空間においてセルは, 図 5 の S1〜S18 の 25 個の領域となる. フィルタ領域の近似を行う際には, 変化領域は必ず 1 つ以上のノイズ領域を組み合わせた領域となるので, 変化領域に含まれるノイズ領域のノイズ出現頻度の合計が変化領域の重み付容量となる.

4　評価実験

実験により, パケット出現頻度を用いた近似によるノイズ低減効果を確認した. snort [6] のルールからランダムに 25 個ずつ抜き出し, サンプルフィルタリングポリシ $P^1_{25}, P^2_{25}, P^3_{25}, P^4_{25}$ を作成した. 同様に 50 個ずつ抜き出し, サンプルフィルタリングポリシ $P^1_{50}, P^2_{50}, P^3_{50}, P^4_{50}$ を, 100 個ずつ抜き出し, サンプルフィルタリングポリシ $P^1_{100}, P^2_{100}, P^3_{100}, P^4_{100}$ をそれぞれ作成した.

サンプルトラヒックに MAWI [7] で公開されているトラヒックアーカイブを用いた.

4.1　実験 1

近似したフィルタリングポリシを用いてトラヒックをキャプチャすることによって発生するノイズの量を調べる. 実験ではフィルタリングポリシとして P^1_{25} を用いる. P^1_{25} から分類表を作成すると,18k バイトとなる. これを 10 k バイト以下になるように近似した場合と 5k バイト以下になるように近似した場合について実験を行う. また, 近似方式として次の 3 つを用いる.

近似方式 1　容量に基づく変化領域決定法

近似方式 2　重み付き容量に基づく変化領域決定法 (軸パケットを用いたデータ削減法を使用)

近似方式 3　重み付き容量に基づく変化領域決定法 (セル毎のパケット到着数を用いたデータ削減法を使用)

実験は次の手順で行う.

1. サンプルトラヒックのキャプチャ
2. 過去のデータと合わせてキャプチャ結果の分析
3. 分析結果に従い近似フィルタリングポリシの生成
4. 3. で生成したフィルタリングポリシを使用した翌日のサンプルトラヒックのキャプチャ
5. 2.〜4. を繰り返す

このとき, 手順 1. でキャプチャされたパケットのノイズの量を調べる. ここで, ノイズの量を示す値としてノイズ率を次のように定義する.

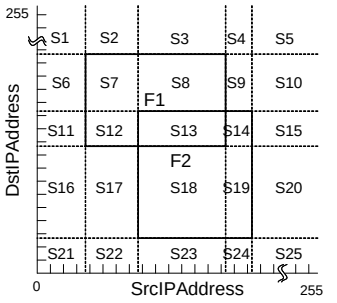

図 5　分類空間をセルに分割した例

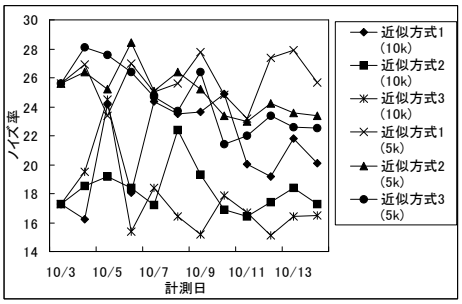

図 6　ノイズ率の変化

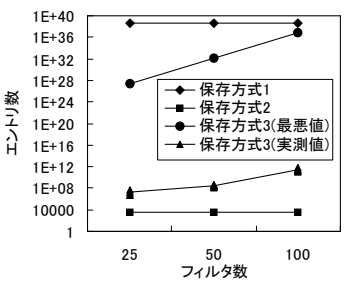

図 7　エントリ数の変化

$$\text{ノイズ率} = \frac{\text{ノイズの数}}{\text{廃棄すべきパケットの数}} \times 100[\%] \quad (2)$$

4.2　実験 2

パケット出現頻度の保存に必要なメモリの量を調べる．フィルタリングポリシとして P_{25}^1〜P_{25}^4, P_{50}^1〜P_{50}^4, P_{100}^1〜P_{100}^4 を用いる．また, パケット出現頻度の保存法として次の 3 つの保存方式を用いる．

保存方式 1　格子点毎にパケット到着数を保存

保存方式 2　軸パケット到着数を保存

保存方式 3　セル毎にパケット到着数を保存

これらの保存方式を用いてパケット出現頻度を表として保存し, このとき必要となるエントリ数を調べる．

4.3　結果と考察

実験 1 の結果を図 6 に示す．横軸はサンプルパケットをキャプチャした日付で縦軸はノイズ率である．実験 2 の結果を図 7 に示す．ここで, 横軸はフィルタの数, 縦軸はエントリ数で, それぞれ対数座標となっている．また, 保存方式 3(最悪値) とはフィルタリングポリシの全てのフィルタ領域の境界が異なる場合に必要となるエントリ数で, 保存方式 3(実測値) とはサンプルフィルタリングポリシを用いた場合に必要なエントリ数で, それぞれのフィルタ数で 4 種類のフィルタリングポリシを用いた場合のそれぞれの値をプロットし, 平均値を結んだ．

図 6 の 14 日のノイズ率をみると, 分類表を 5k バイト,10k バイトのどちらに近似した場合においても, 近似方式 2, 近似方式 3 は従来方式である近似方式 1 よりも低い．従って, パケット出現頻度を用いた近似によりノイズを低減可能であるといえる．

図 6 において, 近似方式 2 と近似方式 3 の 14 日のノイズ率を比較すると, 近似方式 3 の方がノイズ率が低い．また, 図 7 の保存方式 3(実測値) をみると, フィルタ数 25 の場合のエントリ数は 5×10^6 程度の少量のエントリ数でパケット出現頻度を保存できることがわかる．従って, フィルタ数 25 の場合はセル毎にパケット到着数を保存する方式がパケット出現頻度の保存に適している．

しかし, 図 7 の保存方式 3(実測値) のフィルタ数が 50,100 の場合のエントリ数を見ると, 膨大なエントリ数が必要となる．そのため, フィルタ数が 50 以上の場合に保存方式 3 を用いることは困難である．保存方式 2 では, フィルタ数 50,100 の場合においても,4000 程度のエントリ数で保存できる．また, 図 6 をみると, 近似方式 2 のノイズ率は近似方式 1 のノイズ率より低い．従って, フィルタ数が 50 以上の場合は軸パケット到着数を保存する方式が適しているといえる．

5　おわりに

トラヒック特性に基づく近似機能を有する空間分割型パケットキャプチャシステムを提案した．実験を行い, 提案システムによりノイズが減らせることを確認した．

謝辞

日頃, 御指導ならびに御助言を頂いた, 蜂巣吉成講師, 張漢明助教授, 野呂昌満教授, 並びに名古屋工業大学の片山喜章助教授, 高橋直久教授に心より感謝致します．

参考文献

[1] T.V.Lakshman and Dimitrios Stiliadis：High-speed policybased packet forwarding using efficient multi-dimensional range matching, SIGCOMM,pp.203-214,1998.

[2] N.Takahashi:　A systolic sieve array for real-time packet classification, Journal of IPSJ,Vol.42,No.2,pp.146-166,2001.

[3] 加藤和夫, 大須賀怜, 高木淳史, 片山喜章, 高橋直久：スケーラブルパケットキャプチャシステムの実現, 日本ソフトウェア科学会インターネット技術ワークショップ WIT 2003,pp.32-39,November 2003.

[4] 大須賀怜, 片山喜章, 高橋直久：近似機能を有する空間分割型パケット分類器, 情報処理学会論文誌,Vol.47,No.4,pp.1195-1208,April 2006.

[5] 中村陸, 片山喜章, 高橋直久:トラヒック特性に基づく近似機能を有する空間分割型パケットキャプチャシステム, 信学技報,vol.106,no.151,IN2006-48,pp.79-84,July 2006.

[6] Sourcefire：The open source network intrusion detection system, `http://www.snort.org/`.

[7] WIDE Project：Mawi working group traffic archive, `http://tracer.csl.sony.co.jp/mawi/`.