セッション情報を用いた トラフィックの解析に関する研究

松本智[†1] 佐藤聡[†2] 新城靖[†2]
中井央[†3] 板野肯三[†2]
正村雄介[†2] 吉田健一[†4]

1. はじめに

ネットワークシステムの利用形態が多様化している．そのため，ネットワークシステムを流れるネットワークトラフィック（以下，**トラフィック**）はより複雑になっている．それに伴って，ネットワーク管理者がネットワークシステムの利用状況を知ることが困難となってきている．

ネットワークシステムの利用状況を知るために，パケットキャプチャを行う手法がある[3)]．この手法は，パケット情報を基にした情報しか取得できないため，OSI 参照モデルの Layer4 セッション情報に基づいたトラフィックの情報を正確に取得することができない．また，専用の解析機器の導入が必要であり金銭的コストが必要となる．

大学や企業等における大規模なネットワークシステムでは，外部ネットワークシステムとの境界点にネットワークシステムの防衛を目的としたファイアウォール装置が設置されていることが多い．このようなファイアウォール装置は，Layer4 のセッション情報に基づきトラフィックの管理を行っているものが多い．

そこで本研究では，そのようなファイアウォール装置のセッション情報を基にしたトラフィック情報が記録されているファイアウォールログに着目した．ファイアウォールログを解析することによって，セッション情報を基にしたトラフィック情報を得ることで，既存手法であるパケットキャプチャを用いて解析した結果との違いを比較調査する．それによって，パケット情報を基にしたトラフィック情報とセッション情報を基にしたトラフィック情報から得られる情報の違いを明確にし，それぞれの有効性を示す．

2. 集約フローと異なり数

本研究では集約フローと異なり数という概念を用いてトラフィックを解析する。ネットワークシステムを介するトラフィックにおいて，5-tuple（SIP：送信元 IP アドレス，DIP：宛先 IP アドレス，PR：IP プロトコルタイプ，SPT：送信元ポート番号，DPT：宛先ポート番号）が一致するものを同一のフローであるとみなすことができる．文献[3)1)]では，これら 5-tuple のうち，任意の項目について同一であるフローの集合を集約フローと定義している．集約フローにおいて，5-tuple のうち固定した項目以外の項目についての異なり具合を**異なり数**という．トラフィックを集約フローに分類し，その 5-tuple の固定した項目の値と固定しなかった項目の異なり数を計測することにより，その集約フローの振る舞いを分類することができる[3)]．

3. ファイアウォールログを用いたトラフィックの解析

集約フローを得るために，本研究ではファイアウォールログを用いる．ファイアウォールログには，トラフィックについての 5-tuple の情報が Layer4 のセッション単位で記録されている．格納されている 5-tuple については，セッションを認識しているかしていないかの違いはあるものの，パケットキャプチャを行った時に収集できる情報のものと変わらない．そのため，その情報を用いることによって集約フローを求めることができ，その 5-tuple の組み合わせごとに上述した手法を用いて異なり数を計測することができ，トラフィックの特徴抽出を行うことが可能であると考えられる．さらに、パケットキャプチャでは得られないセッション情報を用いて、多角的にトラフィックの解析を行う。

3.1 計測実験

セッション情報を基にしたトラフィック情報についてもパケット情報を基にした方法と同様の基準で集約フローに分類し両者を比較することで，それぞれの情報を解析した際に取得できるトラフィックに関する情報の違いを示す計測実験を行った．

実験にはファイアウォール装置として，Juniper Networks の SSG350M セキュア・サービス・ゲートウェイ（ScreenOS 6.1.0r4.0）を利用した．この装置は，筑波大学キャンパスネットワーク内において学生宿舎ネットワークとキャンパスネットワークの境面に設置されている．また計測実験の比較対象として，キャンパスネットワークに設置されているパケットキャプチャを行って集約フローを求める装置である Aggregated Flow Mining（AFM）装置[2)] を用いて収集したデータを利用した．二つの装置は図 1 に示す箇所に

†1 筑波大学第三学群情報学類
†2 筑波大学システム情報工学研究科
†3 筑波大学図書館情報メディア研究科
†4 筑波大学大学院ビジネス科学研究科

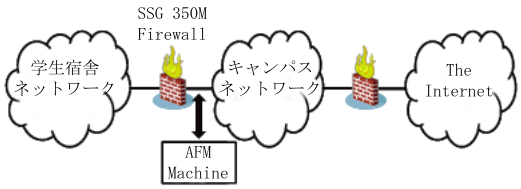

図 1 ネットワーク概要

設置されている．これら二つの装置を用いて 2009 年 10 月 22 日 22 時から 2009 年 10 月 23 日 22 時までの 24 時間のトラフィック情報を収集した．

3.2 評　　価

得られたトラフィック情報について，ファイアウォール装置では 15 分ごとに計測できたセッション情報を，AFM 装置では 1000 パケットごとに計測できたパケット情報をそれぞれ SIP を固定した集約フローに分類した．その結果セッション情報を基にした集約フローとして 98732 フロー，パケット情報を基にした集約フローとして 554626 フロー収集することができた．

それぞれ得られた集約フローについて，計測区間ごとの DIP，DPT の異なり率を求めた結果を図 2，3 に示す．異なり率とは，1 つの集約フローについて，ある項目がどの程度異なっているかを示している．集約フローにおいて異なり率が高いということは，その集約フローは多様な通信を計測したことになる．

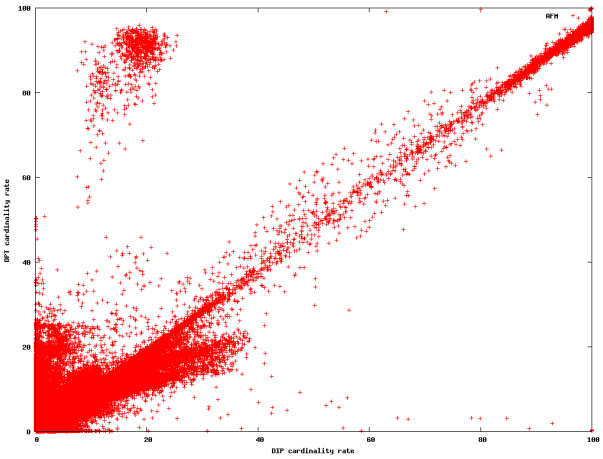

図 2 AFM 装置の異なり率

これら二つの結果から，パケット情報を基にした結果では異なり率が低く，またセッション情報を基にした結果では異なり率が高く分布することが確認できた．転送量の多いセッションは，パケット情報では多くの計測期間にまたがってそのセッションが計測されるため異なり数が低くなる傾向があるためと考えられる．一方セッション情報では 1 度だけ計測されるため，セッションの通信量にかかわらず異なり数に反映されるためであると考えられる．

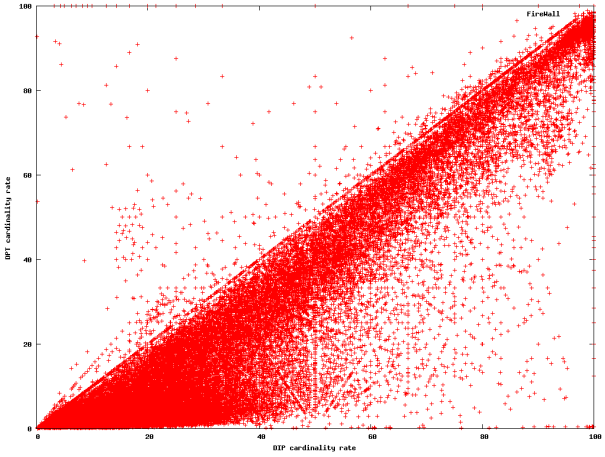

図 3 FW 装置の異なり率

4. おわりに

本論文では，セッション情報を基にした集約フローとパケット情報を基にした集約フローについて，得られる特徴の違いを調査した．セッションあたりの転送量が多い通信は，それぞれの集約フローにおいて異なり数の出方が大きく異なった．これは，大量のデータ転送を多数のホストと行う通信を分類するうえで一定の効果があるといえる．

今後，様々な組み合わせの集約フローについても同様の比較実験を行い，パケット情報を基にしたトラフィック情報とセッション情報を基にしたトラフィック情報から得られる情報の違いをより明確にしたい．

参　考　文　献

1) K.Yoshida, S.Katsuno, A.Ano, K.Yamazaki and M.Tsuru: Stream mining for network management, *IEICE Trans.Commun. vol.E89-B*, pp. 1774–1780 (2006).
2) Y.shomura, A.Sato, K.Yoshida, S.matsumoto and K.Itano: A case study of Analyzed Method for Number of Varieties in Frequently Found Flows in real network environment, *IEICE Technical Report. vol.109 no.262 IA2009-53*, pp. 39–44 (2009).
3) Y.Shomura, Y.Watanabe and K.Yoshida: Analyzing the Number of Varieties in Frequently Found Flows, *IEICE Trans.Commun. vol.E91-B*, pp.1896–1905 (2008).