社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－CSEC－33 (6)
2006／5／12

アノマリコネクションツリーを用いたサイレントワームの早期検知手法の提案

川口 信隆* 重野 寛* 岡田 謙一*

本論文では，イントラネットやLAN内におけるサイレントワームの検知手法を提案する．既存の検知手法の多くは，ワームに感染したホストが起すアドレススキャン等の異常なネットワーク活動を検知する．このため，予め脆弱性を持つホストのリストを利用して，静かに感染活動を行うワームを検知することは難しい．本論文ではこのようなワームを検知するために，アノマリコネクションツリーメソッド (Anomaly Connection Tree Method, ACTM) を提案する．ACTMは多くのワームの感染活動に見られる2つの特徴を検知に利用する．1つ目は，感染コネクションをエッジ，ホストをノードとするツリーが構築されることである．2つ目は，ワームが次の感染先ホストを選択する時，自身が感染しているホストがどのホストと頻繁に通信を行うかということを考慮しないことである．シミュレーションにより，ACTMがワームの感染活動の初期段階（感染ホスト数が全ホスト数の数パーセント程度）で検知を行える事を示す．

ACTM: Fast Detection Method of Silent Worms using Anomaly Connection Tree

Nobutaka KAWAGUCHI* Hiroshi SHIGENO* Kenichi OKADA*

In this paper we propose a novel worm detection method that can detect silent worms in intranet and local area network. Most existing detection methods use aggressive activities of worms as a clue for detection and are ineffective against worms that propagate silently using a list of vulnerable hosts.To detect such worms, we propose Anomaly Connection Tree Method (ACTM). ACTM uses two features present to most worms to detect worms. First is that the worms's propagation behaviour is expressed as tree-like structures. Second is that the worm's selection of infection targets does not consider which hosts its infected host communicates to frequently. Through the simulation results, we have shown that ACTM can detect the worms in an early stage.

1 はじめに

Staniford らは近年フラッシュワームという新しい形態のワームが現れる可能性について言及している [3] [4]．フラッシュワームは脆弱性ホストのアドレスリストを持ち，このリストを利用して高速かつ効率的な感染活動をするワームである．フラッシュワームは無差別的なアドレススキャンを行わない点で，Blaster，CodeRed といったこれまでのワームと大きく異なる．我々はこのフラッシュワームのようにアドレスリストを持ち，且つ個々の感染ホストが攻撃するホスト数を控え，異常なネットワーク活動を抑制するワームをサイレントワームと定義する．このワームは，他のワームに比べて少ないアクティビティで効率的な感染活動を行う．アドレススキャンに伴うトラフィックの異常性を検知に利用する既存手法では，サイレントワームの感染活動を早期に検知するのは難しい．

そこで本論文では，イントラネットやLAN内における未知のサイレントワームの感染活動を効率的に検知する手法として，アノマリコネクションツリーメソッド（Anomaly Connection Tree Method, 以下ACTM) [1] を提案する．ACTMは，通常のネットワーク活動では発生頻度が低いコネクション AC (Anomaly Connection) のツリーを検出することで，ワームの存在を検知する．ACTMは，サイレントワームを含む多くのワームの感染活動が示す2つの特徴を利用する．1つ目は，ワームの感染活動は感

* 慶應義塾大学 理工学部 情報工学科
Department of Instrumentation(Information), Faculty of Science and Technology, Keio University

染ホストをノードするツリー構造を取ることである．2つ目はワームが攻撃ホストを選択する時，自身が感染しているホストがどのホストと頻繁に通信を行うかということを考慮しないことである．

以下，第2章ではサイレントワームについて論じる．第3章ではACTMを提案し手法の詳細について述べる．第4章ではシミュレーション実験を通じてACTMの評価と考察を行う．そして，第5章を本論文のまとめをする．

2 サイレントワーム

Staniford らが提起したフラッシュワーム [3] [4] は，予め何らかの手段で脆弱性をもつホストのアドレスリストであるヒットリストを作成し，このリストを元に効率的かつ高速な感染活動を行うワームである．ヒットリストはワーム間で共有され，攻撃先ホストを決定する際に用いられる．アドレススキャンによる感染活動と異なり，ヒットリストを用いた感染活動は高い確率で成功する．

我々はフラッシュワーム同様にヒットリストを持ち，且つ個々の感染ホストの他ホストへの感染試行回数を数回程度にすることで，ネットワークベースのIDSによる検知を困難とするワームをサイレントワームと定義する．本論文ではその中でも特に，サーバホストだけではなくクライアントホストも有するネットワークサービスの脆弱性を利用することで，1つのイントラネットやLAN中の全脆弱性ホストに感染しようとするサイレントワームを対象とする．代表的な脆弱性には，BlasterやSasserが利用したMicrosoft WindowsのRPCサービスの脆弱性などがある．

3 アノマリコネクションツリーメソッド

本章では，イントラネットワーク内で発生したサイレントワームを検知するための手法としてアノマリコネクションツリーメソッド(ACTM)を提案する．

3.1 コネクションモデル

本論文では，2つのホスト間の通信を抽象化してコネクションと呼ぶことにする．このようなコネクションは実際にはTCPコネクションである場合や，UDPによる一連のパケットのやり取りである場合もある．ACTMはコネクションを何らかの手段で検出できる事を前提する．コネクションは送信元，送信先アドレス，ポート番号などで識別できるものとする．ACTMは，接続元ホスト，接続先ホストが共に検知対象ネットワーク内に存在するコネクションをIC(Internal Connection)と呼ぶ．ICの中で，正規のネットワーク活動により発生するICをLC(Legistimate Connection)，ワームの感染活動により発生するICをWC(Worm Connection)と呼ぶ．次に，ICのうち，ある閾値を基準として発生頻度が高いICをNC(Normal Connection)，発生頻度が低いICをAC(Anomaly Connection)とする．つまりACは，通常のネットワーク活動下における通信頻度が一定数以下の相手ホストとのICと定義される．NC，ACの分類はあくまで過去のコネクション発生の履歴から求めた頻度によるため，LC，WCとは異なる概念である．

ACTMは個々のICを，接続元ホスト，接続先ホストで分類する．1つのホストを共有するIC，例えばホストAからBへのICとホストAからCへのICは区別される．また同じホスト間のコネクションであっても向きが異なれば，異なるコネクションとして区別される．

3.2 検知アルゴリズム

3.2.1 アルゴリズムの概要

ACTMは，ICからACを抽出し，ACをエッジ，ホストをノードとするツリー構造を検出することで，ネットワーク中のワームの存在を検知する．ACTMは以下に述べるワーム感染活動の2つの特徴を検知に利用する．

1. ワームは自身を再帰的に感染ホストにコピーし感染ホストはさらに他のホストに再感染を行う．よってワームの感染活動は感染ホストをノード，WCをエッジとしたツリーとして表現できる．
2. ワームは，自身が感染しているホストがどのホストと頻繁に通信するかを行うかを考慮せずに，感染活動を行う．

一方，一般に，ワームに感染していないホストは，ネットワーク中の全ホストの一部のホストとのみ頻繁に通信する [5] [6]．例えば，あるホストの通信のうち80%の宛先ホストは全ホスト中の20%のホストに集中するなど，通信頻度に偏りがある．

ACTMは，ネットワーク内でパケットキャプチャリング等を通じて，ネットワーク内で発生するICを観測する．次に，観測したICをACまたはNCに分類する．そして，ACに分類された複数のACを連結することで，アノマリコネクションツリー(ACツリー)を検出する．

ACTMはワームが発生していない一定期間に観測したLCのうち発生頻度が閾値以上であるLCをNCと判断する．そして，NC以外のICをACと判断す

る．前述の通りワームに感染していないホストのLCの相手先ホストは一部ホストに集中するため，LCの多くはNCに分類される．一方で，ワームの特徴(2)で述べた通りワームは，自身の感染ホストの通信先ホストの特性を考慮せずに感染活動を行う．このため，ACTMはWCの多くをACと判断する．

図1に，複数のホストが複数のAC，NCを確立している様子を示す．この図では{A,B,D,E,H}，{C,F,G,I}の2つのACツリーが検出される．LCがACである確率はWCがACである確率よりも相対的に低い．このため，通常のネットワーク活動時にツリー状になったn個のACが検出される確率は，ワームが発生した場合にツリー状になったn個のACが検出される確率と比べて相対的に低い．そこでACTMは閾値を超えるサイズのACツリーが検出された場合，ワームがツリー中に存在すると判断する．なお本論文ではツリーサイズを，ツリーに含まれるホスト数とする．

一方，WCがNCとして検出される確率も一定量ある．このため，ワーム発生に起因するACツリーは，1つではなく複数のツリーに分割された形で検出される．このような場合，ある1つのツリーから一定距離内に平均的なツリーサイズよりも大きいACツリーが複数検出されるという傾向がある．そこでACTMは複数の距離が近いACツリーを集約したことで，VACツリー(Virtual ACツリー)という仮想的なACツリーを検出する．そしてACツリー同様，閾値を超えるVACツリーが検出された場合もACツリーの場合と同様に，ワームがVACツリー中に存在すると判断する．なお，ACツリー間の距離の定義などは3.2.4で述べる．

ACTMは(1)学習フェイズ，(2)検知フェイズの2つのフェイズを持つ．学習フェイズではネットワークを観測し，検知に用いられるACツリー，VACツリーの閾値を決定する．またそのために，ACと判断されるべきICのリストであるACリストを生成する．このフェイズではワームは存在していないものとする．検知フェイズではリアルタイムにAC，VACツリーを検出していき，閾値を越えるツリーが検出された場合，ワームがネットワーク中に存在すると判断する．

3.2.2 ACリスト作成アルゴリズム

ここでは学習フェイズで実行するACリスト作成アルゴリズムについて述べる．学習フェイズでは一定期間ネットワークを観測しICログを収集する．そして個々のホストごとに，ネットワーク中のその他のホストを接続先とするIC数の頻度分布のリストを作成する．このリストをClistと呼ぶ．Clistにはネットワーク中の全ホストが登録される．接続先ホストは，IC数が多い順に並べられる．図2にあるホストXのClistの例を示す．ここで，CListの上位FR($0 \leq FR \leq 1$)のホストをFhostと呼ぶ．そして，FhostとのICをNC，Fhost以外のホストとのICをACとする．また，全IC数に対するNC数の比率を，CR(Fhost's Connection Rate)とする．ワームはランダムに感染試行先ホストを選ぶため，あるWCがACと判断される確率は1.0-FRとなる．一方，LCがACと判断される確率は1.0-CRとなる．尚，ホストごとにCListの内容は違うため全ホストで同一のFR値を用いたとしても，CRの値は各ホストによって異なる．

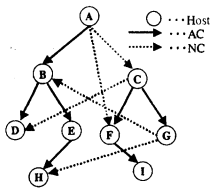

図1: ACツリーの例

接続先ホスト	IC数	
A	90	2 Fhosts (=10*FR)
B	70	CR=0.8 (=160/200)
C	13	
D	10	
E	5	
F	5	
G	4	
H	3	
I	0	
J	0	
合計10ホスト	合計 200IC	

図2: ホストXのCList（FR=0.2）

図2では，全接続先ホスト数が10台，FRが0.2の場合を示している．ホストXのFhostはホストA，Bとなる．よってホストXと，ホストA,Bとの通信はNC，それ以外のホストとのICはACとなる．また，CRは0.8(= (90 + 70)/200)となる．よってWCがACである確率は0.8，LCがACである確率は0.2となる．

最後にACTMは，個々の接続元ホストと，ACと見なされるICの接続先ホストの対応付けをACリストとしてまとめる．あるICがACであるかNCで

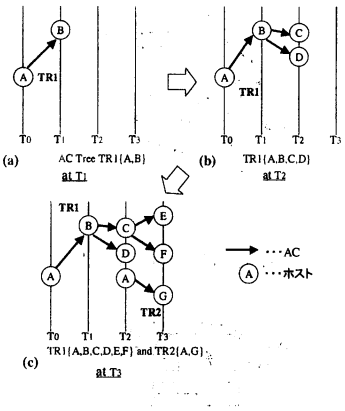

図 3: AC ツリーの構築

あるかを判断する際には AC リストを用いる.

3.2.3 AC ツリー検出アルゴリズム

ACTM は新しく検出した AC をすでに検出した AC ツリーに連結することで，より大きな AC ツリーを探索していく．図 3 に AC が連結され，より大きなツリーが検出される様子を示す．この図では AC は接続元ホストを始点，接続先ホストを終点とした矢印で表現される．検出時間は終点が示す時間となる．以下ホスト X が通信開始ホスト（始点）ホスト Y が通信先ホスト（終点）で時刻 Z に検出された AC を $AC^{Z}_{X,Y}$ と表記する．図 3(a) の AC は $AC^{T_1}_{A,B}$ と表記する.

ある AC X を検知した場合，AC X の始点が過去 T_{limit} 以内に検出した AC Y の終点または始点と一致する時，ACTM は AC X を AC Y に連結する．T_{limit} を接続限界時間と呼ぶ．AC を複数のツリーに AC に接続できる場合は，最も大きいツリーに属する AC を選択するものとする.

図 3 に，$T_2 - T_1 \leqq T_{limit} < T_3 - T_1$ であるときの AC ツリーの検出の様子について示す．図 3(b) では，時刻 T_2 に検出された $AC^{T_2}_{B,C}$,$AC^{T_2}_{B,D}$ は共に，時刻 T_1 に検出された $AC^{T_1}_{A,B}$ に連結される．これは，検知間隔時間 $T_2 - T_1$ が T_{limit} 以内であるためである．これによりツリー TR1 のサイズは 2 から 4 へと増える．一方，条件を満たす連結相手が無い場合，新しく検出された AC は，新しい AC ツリーの最初の AC となる．例えば図 3(C) では，時刻 T_3 に検出された $AC^{T_3}_{A,G}$ は既存のツリー TR1 の $AC^{T_1}_{A,B}$ とは連結されず，新しいツリー TR2 として認識される．

アルゴリズムに接続限界時間を導入するのは，長期間をかけて大きくなる AC ツリーの検知を制限するためである．接続限界時間により，ワームが存在していない時に必要以上に大きい AC ツリーが検出されることを制限することで，ワームの侵入により急速に成長する AC ツリーと，通常のネットワーク活動により比較的ゆっくりと成長するツリーの区別が可能であると考えられる.

3.2.4 VAC ツリー検出アルゴリズム

前述の通り，WC が AC として検出される確率はであり，1-FR は NC として検出される確率は FR である．よって FR が十分小さいときには，WC が AC として検出される確率は NC として検出される確率よりも高い．しかし，WC であっても確率 FR で NC と見なされる．このため，ワームの感染活動により成長していく AC ツリーが途中で NC と見なされた WC によって分断される．結果的に，図 4 に示すように，大規模な 1 つの AC ツリーが検出されるのではなく，NC によって隔てられた複数の AC ツリーが検出される．図 4 では 2 つの NC によって，TR1 から 2 つのツリー TR2，TR3 が分離され，3 つのツリーが検出される．NC によって元々のツリー（図 4 では TR1) 成長は妨げられ，ツリーサイズが閾値を超えるのに時間がかかるため，検知が遅れることになる.

ここで，NC によって分離された 2 つの AC ツリー (TR1 と TR2, TR1 と TR3) の距離を，ツリー間を最短で連結する NC のパスの長さと定義する．ワームの感染活動に伴い発生する AC ツリーが NC によって 2 つに分離される場合，ツリー間の距離が d となる確率 X は $X = FR^d * (1.0 - FR)$ となる．FR=0.2 の場合，$d \geq 2$ になる確率は 0.04 となる．よってワームにより発生した AC ツリーは，NC によって分離されたとしても，比較的近い距離内 (d < 2) に密集する確率は高いと言える．よって，ある AC ツリーから一定距離内に存在する AC ツリーを集約し 1 つのツリーとして扱うことで，NC により分離されたツリー構造を高い確率で復元して検出できる．そこで ACTM は，任意の AC ツリーをセンターツリーと定め、センターツリーから NC によって分離され，且つセンターツリーの近傍の AC ツリーを集約することで，VAC ツリーを検出する．そしてこの VAC ツリーのサイズが閾値を超えた場合に，ワームの存在を判断する．図 4 では，TR1 をセンターツリー，TR2，TR3 を近傍ツリーとする VAC ツリー VTR1 が検出される.

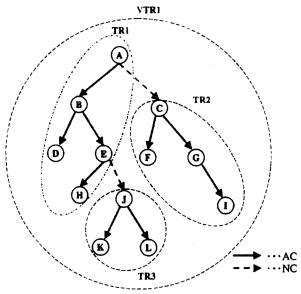

図 4: $V_d = 1, V_n = 2$ のときの VAC ツリー VTR1

このアルゴリズムは全ての AC ツリーに対して適用される．つまり，AC ツリーが N 個あれば，各ツリーをセンターツリーとした VAC ツリーが計 N 個検出され，ツリーサイズがチェックされる．

ここで，VAC ツリーに集約する近傍ツリーを決定するためのパラメータとして V_d,V_n を定義する．V_d は，VAC ツリーに集約される近傍ツリーのセンターツリー (図 4 では TR1) からの最大距離 (図 4 では 1)，V_n は近傍ツリー数の上限 (図 4 では 2) を示す．センターツリー TR1 から距離 V_d 以内に S_d 個の近傍ツリー $NTR_i, (1 \leq i \leq S_d$, 但し $sizeof(NTR_i) \geq sizeof(NTR_{i+1}))$ が存在する場合 VAC ツリー VTR1 のサイズ $sizeof(VTR1)$ は

$$sizeof(VTR) = sizeof(TR1) + \sum_{i=1}^{V_n} sizeof(NTR_i) \quad (1)$$

となる．但し $S_d < V_n$ の場合は，$V_n = S_d$ とする．この式からわかるようにサイズが大きい NTR_i を優先して VAC ツリーに加える．これは，ワームの感染活動に起因する AC ツリーを優先して集約するためである．サイズが小さい AC ツリーはワームの感染活動の有無に関わらず多数検出されるため，これらを集約してもワーム検知に役立てられない．

尚，ある NC が 2 つの AC ツリー TR1，TR2 を連結するには，3.2.3 の AC ツリー検出アルゴリズムでの AC の連結条件と同様，TR1，TR2 中の AC1,AC2 が NC1 と始点または終点を共有し，接続間隔時間が接続限界時間以内である必要がある．TR1 と TR2 の距離が 2 以上の場合は複数の連続した NC によって連結されるが，このときの AC と NC，又は NC 同士の連結条件も同様である．

4 評価

4.1 シミュレーション条件

本シミュレーションでは，ネットワークとしてイントラネットを想定し，ネットワーク内の全ホストが脆弱性ホストとする．各ホストはファイアーウォールなどに妨げられずに他のどのホストとも自由に通信を行えるとする．時間は TU を単位時間として用いる．通常のネットワーク活動は指数分布に従って平均 10TU ごとに，各ホストは他のホストへ LC を確立するものとした．検知フェイズでは 1 つのサイレントワームが何らかの手段でネットワーク内の 1 台のホストに感染し，そこから全ホストをターゲットとした感染活動を行う．ワームは感染コネクションとして TCP を用いるものとする．表 1 にシミュレーションパラメータを示す．本シミュレーションでは断りがない限り，FR を 0.2 に，全ホストで CR は 0.8 に設定した．また 1 つのワーム感染ホストが行う感染試行回数は 2 とする．現状の殆どのワームがほぼ無制限に感染活動を行う事を考えると，この値は小さく，検知側にとってはより厳しい条件である．感染インターバルは連続する 2 つの WC の開始時間の間隔の平均値を意味する．またあるホストにワームが感染してから第一回目の WC を行うまでの時間間隔も同様とする．実際の感染インターバルはこの値の ±20% の範囲の乱数値をとるものとした．

表 1: シミュレーションパラメータ

ホスト数	1000 台
FR / CR	0.2 / 0.8
通常コネクションインターバル	平均 10TU
感染試行回数	最大で 2 回
感染インターバル	可変 (1-20TU)
V_d / V_n	1 / 2

学習フェイズではワームが存在しない状態でシミュレーションを 10000TU 時間実行し，この時検出される AC ツリーと VAC ツリーの最大値を TH_{AC}, TH_{VAC} とした．検知フェイズではシミュレーション実行 1000TU 時間後にサイレントワームを 1 台のホストに感染させる．ワームはすぐに感染活動を開始する．そして，閾値を越える AC，VAC ツリーが生成された時のワーム感染ホスト数，検知時感染ホスト数を測定する．結果はシミュレーション 20 回

の平均値である．この条件の下で検知時感染ホスト数を比較する。比較対象手法としては，ウイルススロットル [2] と AC カウント手法の 2 つを用いる．

ウイルススロットル [2] は，新しい IC が確立される度に接続開始パケット (TCP SYN パケットなど) を，送信元ホストのキューにプッシュする．その一方で，一定時間ごとにキューからパケットをポップする．大量のパケットによりキューが溢れた場合，ワームが存在すると判断する．なお，この評価実験ではウイルススロットルは，接続開始パケットから IC が AC か NC であるかを判定し，AC の接続開始パケットのみをキューにプッシュするものとした．これによりキューサイズは小さくなり，検知速度が向上する．

AC カウント手法は一定時間（検知ウィンドウ）以内の過去に検出された AC の合計値を求め，この値が閾値を超えた場合ワームがネットワーク内に存在すると判定する．ACTM 同様，閾値は学習フェイズで検出された AC の合計数のうち最も大きいものを用いる．

4.2 シミュレーション結果

図 5 に，ACTM，AC カウント手法における検知時感染ホスト数を示す．ワームの感染インターバルが 14TU より短い場合，ACTM は AC カウント手法よりも早い段階での検知を実現している．よってこの範囲内では，ワームの感染活動の特徴であるツリー構造の利用が検知時間の短縮に貢献していると言える．一方で，感染インターバルが 14TU を越えると AC カウント手法による検知のほうが早くなる．これは感染インターバルが長くなるにつれ，TH_{AC}, TH_{VAC} が大きくなるからである．このとき，NC が AC ツリーの成長を妨げる可能性は増え，VAC ツリーでも分断された全ての WC に起因する AC ツリーを集約することが難しくなる．

通常，多くのワームのコネクション頻度は通常ネットワーク活動の数倍〜数十倍である．感染インターバルが通常コネクションインターバル以上であるワームは極めて遅いワームと言える．このため通常コネクションインターバルの 1.4 倍程度までの感染インターバルにおいて，AC カウント手法よりも早期の検知を実現している ACTM の有効性は高い．特に感染インターバルが通常コネクションインターバル 10TU より短い場合，ACTM は全ホストの 5%程度が感染した段階でワームを検知できる．

また，スケールの関係で図 5 には示されていないが，ウイルススロットルにおける検知時の感染ホスト数

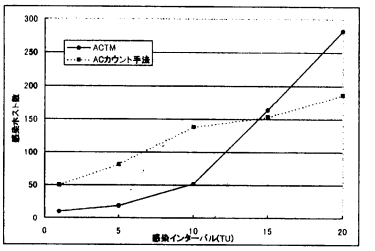

図 5: 検知時感染ホスト数の比較

は感染インターバルが 1TU で 519 台，10TU で 720 台となり，ACTM に比べて大きい．これは，サイレントワームは従来のワームと異なり個々の感染ホストが数回程度しか感染活動を行わないため，キューが溢れるほどパケットが溜まる可能性が低いためである．

5 まとめ

本論文では，イントラネットワークや LAN 内でのサイレントワームの検知を行うために ACTM を提案した．ACTM は，(1) ワームの感染活動をツリー構造として表現できる，(2) ワームが通常のネットワーク活動での発生確率が低いコネクションを多数確立する，という 2 つの特徴を利用した検知を行う．評価実験を通じて，ACTM は感染インターバルが通常コネクションインターバルより短いワームを，全ホストの 5 パーセント以下が感染した段階で検知できる事を示した．

参考文献

[1] Nobutaka Kawaguchi, Yusuke Azuma, Shintaro Ueda, Hiroshi Shigeno, Kenichi Okada, ACTM: Anomaly Connection Tree Method to detect Silent Worms, in Proc of The 20th IEEE International Conference on Advanced Information Networking and Applications, to be appear, 2006.

[2] M. Williamson, Throttling Viruses: Restricting Propagation to Defeat Malicious Mobile Code, Technical Report HPL-2002-172, 2002.

[3] S. Staniford, et.al., How to Own the Internet in Your Spare Time, in Proc of the 11th USENIX Security Symposium, 2002.

[4] S. Staniford, et.al., The Top Speed of Flash Worms, in Proc of WORM 2004, pp.33-42, 2004.

[5] William Aiello, et.al., Analysis of Communities Of Interest in Data Networks. in Proc of Passive and Active Measurement Workshop 2005, March 2005.

[6] Ruoming Pang, et.al., A First Look at Modern Enterprise Traffic, in Proc of the Internet Measurement Conference (IMC) 2005.