様式 C-19

科学研究費補助金研究成果報告書

平成２４年　５月１９日現在

機関番号：１１３０１
研究種目：　基盤研究(B)
研究期間：　2008～2010
課題番号：　20300023
研究課題名（和文）
多解像度観測量に対応した分散型ネットワーク異常検知システムの構築
研究課題名（英文）
Distributed Network anomaly Detection using Multiresolutional Observables
研究代表者
根元　義章（NEMOTO YOSHIAKI）
東北大学・大学院情報科学研究科・理事
研究者番号：60005527

研究成果の概要（和文）：
ネットワークを流れるパケットを送受信端末毎にグループ化したフローに対し、複数のアルゴリズムにより抽出した観測量を利用することにより、ネットワーク管理上許されていない異常な通信を高精度に識別・検知出来るトラヒック評価方式を提案した。これらの観測量を分散配置された観測点から収集することにより、未知のウイルスの検知も可能な分散型ネットワーク異常検知システムを構築する成功した。

研究成果の概要（英文）：
A network anomaly detection system has been developed. This system can achieve high detection accuracy by using feature values extracted with plural algorithms from network flows of which packets are aggregated based on their IP addresses and port numbers. The system can higher detection rate with feature values collected from distributed observation points.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	3,500,000	1,050,000	4,550,000
2009 年度	1,400,000	420,000	1,820,000
2010 年度	2,000,000	600,000	2,600,000
年度			
年度			
総　計	6,900,000	2,070,000	8,970,000

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：情報システム、セキュア・ネットワーク

１．研究開始当初の背景

近年、インターネットは電子政府などをはじめとする多くの社会・経済基盤に利用され、その役割の重要性は高まる一方である。これに伴い、インターネットには健全で堅固であることが強く要望されている。誰もが情報通信技術の利便性を享受できる安全・安心な情報化社会の発展のためには、インターネットの安全で安定した運用技術の確立が世界的な最重要課題である。

そのようなネットワーク運用・管理技術を確立するために、ネットワークトラヒックの絶対量に影響を受けないネットワークトラヒック状態評価方式である相関係数発生確率行列を提案した。この方式に様々なアルゴリズムにより抽出された観測量を入力し解析することにより、ネットワークの異常状態を高精度に検知するシステムの構築を目指す。

２．研究の目的

本研究では、既に提案している相関係数発生確率行列によるネットワーク状態評価方式を様々な観測方式に対応可能なネットワーク評価システムとして拡張し、効率的な異常通信の範囲と原因の特定を実現する分散型ネットワーク異常検知システムの構築を行う。更に、本研究で提案するネットワーク異常検知方式が、ネットワークにおける異常状態の範囲、ホストやフローなど異常として検知可能な単位を明らかにすることを目的とする。

３．研究の方法

（１）トラヒックデータベースの作成

ネットワークトラヒックを発生させる最小単位である個々のホストや通信の最小単位である端末間のフローを観測単位とし、それらのデータベースを作成する。ホスト単位の場合は、一定時間間隔に観測されたパケットヘッダの集合を作成し、ネットワークパケットの数値化方式の検討過程で効率よく扱えるようなデータベースを作成する。フロー単位のデータベースは、送受信端末のＩＰアドレス、ポート番号の組み合わせが一致するパケットの集合を抽出し、ＴＣＰによる通信の場合はシーケンス番号に基づきパケットを整列する。ＵＤＰによる通信の場合は、観測時刻により整列されることとする。

（２）相関係数算出単位の検討

（１）で作成したホスト単位のトラヒックデータベースを利用し、相関係数を算出する単位を、単一ホストのトラヒック、複数のホストをまとめたトラヒックと変化させ、それぞれにおける検知特性や精度の変化を検討する。これにより、異常検知システムの分散化やホストのグルーピングなどの方針を立てる。

（３）相関係数算出方法の検討

積率相関と順位相関をそれぞれ利用し、異常トラヒックの検知精度や特性の違いを検討する。積率相関係数では、パケット数の単位時間当たりの変動が少ない場合、有意に係数の変化を観測することができずに、未検知となっていた異常トラヒックが存在し得ることが予想される。一方で、順位相関係数を用いた際は、単位時間当たりのパケット数の変動が少ない場合でも他の観測種別のパケット数の大小関係により順位が変動するため、パケット数の絶対値に影響を受けずに異常トラヒックを検知可能と考えられる。この予想を（１）で作成したデータベースを利用し検証する。

（４）フロー単位での状態評価

（２）と（３）の検討においては、単位時間に観測されたパケットの種別を利用し、ネットワークの状態評価を行っている。この評価方法の場合、異常状態を引き起こしている通信を特定することが困難である。そこで、通信端末毎に抽出したフローを評価対象として異常通信の検知・特定方式の検討を行う。具体的には、（１）で作成したフロー単位のデータベースに様々なアルゴリズムによる数値化方式を適用し、複数の特徴量を抽出する。それら特徴量に相関係数を利用した異常検知方式、パターン認識技術を応用した通信種別特定方式を適用し、ネットワーク管理上許されていないアプリケーションの利用などの不正を含んだ通信の検知を可能とする異常通信検知方式を検討する。

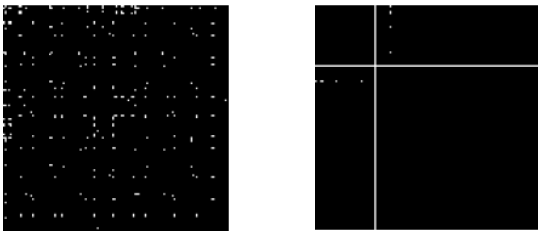
（５）分散型異常検知のための交換情報の検討

（２）～（４）で検討した異常検知方式により発見した異常通信を複数の観測点で交換することにより、高精度な異常検知システムの構築を行う。その際、分散配置された異常検知システム間でやり取りすべき情報の種類やそれら情報の統合方法を検討し、分散型異常検知システムを完成する。

４．研究成果

（１）相関係数観測単位の変化による異常検知精度

LAN 全体のトラヒックとホスト毎に分割したトラヒックからそれぞれ相関係数を算出し、その発生確率行列を図示する（図１）。相関係数発生確率行列を算出したトラヒックには、ある一つのホストへのスキャンが含まれている。LAN 単位での相関係数算出では、スキャンのトラヒックが他の通常通信に埋もれてしまい、発生確率行列に有意な変化を見ることが出来ない。その一方で、ホスト単位での相関係数算出では、適切にスキャントラヒックを検知出来ることが確認されている。このことから、管理目的に応じた適切な単位での観測が必要であることが明らかとなった。



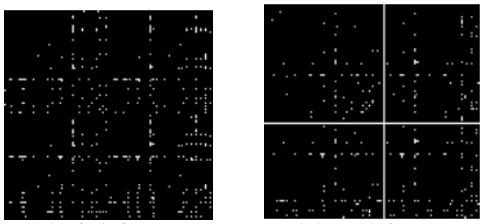
(a)LAN 単位　　(b)ホスト単位

図１：LAN とホスト単位で算出の相関係数発生確率行列

（２）相関係数算出方法の違いによる検知特性の変化

ネットワークトラヒックをホスト単位に分割し相関係数を算出することで、LAN 全体のトラヒックを対象とした場合に比べ、小規模な攻撃の検知が可能であることが判明した。しかし、ホスト単位での観測では、解析対象のトラヒックの絶対量が減少するため、有意に相関係数の変化を観測することが困難となる場合があることも明らかとなった。そこで、積率相関ではなく順位相関を利用した相関係数発生確率行列での異常トラヒック検知を検証した。

図２に設定ミスによって発生した OSFP のルーティングパケットを検知した際の相関係数発生確率行列を示す。ここで利用したデータはホスト単位での観測量データベースうである。白色の線は発生確率が 0.01 未満の箇所である。順位相関を用いた場合、積率相関には存在しない異常を表す直線が表れていることが分かる。このことから、順位相関を利用することにより、積率相関では検知不可能だった異常トラヒックを検知することが可能となった。



(a)積率相関　　(b)順位相関

図２：積率相関と順位相関によるルーティングパケットの検知

（３）フロー単位での異常トラヒック検知

（１）と（２）では、ホスト単位での検知となっている。ネットワークにおいては、一つのホストが複数の端末と同時に通信するため、ホスト単位での異常状態検知では安全性の確保は不十分である。そこで、フロー単位での異常検知、その通信種別の識別を検討した。

①パケットサイズの遷移パターンによる通信識別

フローを構成するパケットのサイズを観測順に並べベクトル化した観測量にクラスタリング等の識別モデルを適用し、そのフローのアプリケーション種別を識別した。この識別により、ネットワーク管理上利用の許されていない通信を検知出来ることになる。

本研究においては、パケットサイズが比較的小さいものに、それぞれの通信の特徴が現れると仮定し、パケットサイズの逆数を利用することを提案した。パケットサイズの逆数

	oLVQ1		K-means		Hierarchical	
	A [%]	th	A [%]	th	A [%]	th
HTTP	17.17	0.0012	-	-	5.54	0.0020
HTTPS	20.69	0.0053	46.24	0.0039	79.19	0.0079
IMAP	97.03	0.033	97.49	0.039	99.24	0.0055
POP3	97.38	0.0093	97.38	0.0081	98.59	0.0047
POPS	85.84	0.011	85.84	0.011	86.02	0.012
RTSP	-	-	-	-	0.37	0.0007
Share	94.88	0.67	97.21	0.71	98.06	0.0017
SMTP	91.14	0.014	66.81	0.010	94.68	0.018
SSH	81.50	0.020	84.59	0.023	85.65	0.010
Winny	93.51	0.11	52.81	0.015	84.99	0.024
Overall	79.03		73.05		84.69	

を利用したアプリケーション種別の識別結果を表１に示す。

表１：パケットサイズの逆数によるアプリケーション識別精度

表より、80%以上の精度でフローのアプリケーション種別の識別が実現出来ていることが分かる。

②パケットデータの遷移パターンによる識別

通信のデータ内容をパケット毎に不可逆変換を施し、その遷移を距離遷移ベクトルとしてモデル化することを提案した。その距離遷移ベクトルを利用したアプリケーション識別の識別精度を表２に示す。表２より、95%以上の精度で識別できることが分かる。

表２：距離系列ベクトルによる識別

	学習データ	テストデータ	学習データ	テストデータ	学習データ	テストデータ
アプリケーション	正解率	正解率	正解率	正解率	正解率	正解率
winny	89.63%	83.33%	93.10%	85.71%	75.00%	65.52%
share	100.0%	98.86%	99.31%	98.70%	99.43%	99.34%
http	99.87%	99.95%	99.72%	99.84%	99.86%	99.91%
pop3	91.63%	90.78%	96.20%	97.31%	99.52%	100.00%
netbios	44.12%	48.00%	79.17%	88.89%	78.95%	73.33%
skype	73.91%	75.00%	72.22%	78.26%	92.31%	94.12%
ntp	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
dns	18.75%	100.00%	40.00%	100.00%		
TOTAL	98.38 %	98.40 %	99.05%	99.15 %	99.55%	99.45 %

③メッセージの送受信タイミングによる識別

①②で提案した方式では、パディングなどによる改竄のリスクが残されている。そこで、パケットサイズやそのデータ内容ではなく、メッセージに送受信タイミングの遷移をベクトル化する方式を検討した。時間軸の観測量は、計算機の負荷状況やネットワーク遅延などのノイズ的な要素が混入する可能性があり、不安定な観測量であると予想されるが、80%程度のアプリケーション識別精度を確保すること可能となった。考慮するメッセージ数による識別精度の変化を図３に示す。

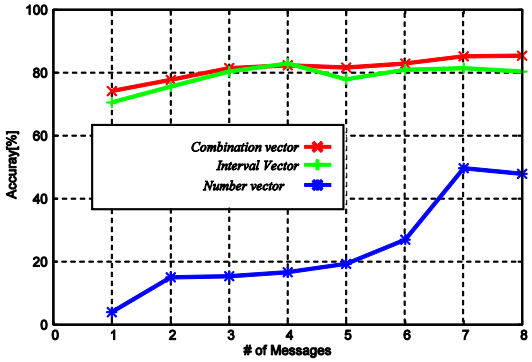


図３：メッセージの送受信タイミングによる識別

図３より、メッセージの送受信タイミングにおいても、アプリケーション種別に独特のタイミングが存在し、それによるアプリケーションの識別や異常通信の検知が可能であると結論付けられる。

（４）分散型異常通信検知システムの構築

研究成果（３）までで検討したトラヒックの数値化手法を利用し、異常通信の特徴を表し得るシグネチャを自動生成し、分散配置された異常検知システムと共有することにより、新種のウイルスなどを検知可能な分散型異常検知システムを構築した。シグネチャ生成に利用した数値化手法は、研究成果（３）②のパケットのデータ内容の不可逆変換に基づいたものである。分散配置された観測点から、この方式に基づき数値化されたフローを収集し、クラスタ分析を適用することにより、新種のウイルスの検知が実現出来ることが明らかとなった。収集されたフローが新種のウイルスであると判断する上での仮定は、ウイルスがインターネット上に出現する際は、その感染活動により、複数のネットワークで高い類似性を持った通信が多数観測されるということである。この仮定に基づき、複数のネットワークで観測されたフローが、一定以上の規模もクラスタを生成した場合、それがネットワークウイルスであると判断し検知することとした。この検知アルゴリズムにより、まだ検知用のシグネチャが作成される前の新種のウイルスであっても、90%以上の正確性で検知出来ることが明らかとなった。

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４件）

１．　Y.Waizumi、Y. Tsukabe,、H. Tsunoda, Y.Nemoto 、 K.Tanaka 、 ”Network Application Identification Based on Communication Characteristics of Application Messages” 、 Journal of Communication And Computer、No.8、2011年、111-119、査読有

２．　和泉勇治、阿部康一、根元義章、”メッセージの遷移パターンに基づくネットワークアプリケーション識別システムの試作”、電子情報通信学会論文誌 D、J93-D、2010年、2257-2267、査読有

３．　K. Simkhada, T. Taleb, Y. Waizumi, A. Jamalipour, Y. Nemoto、“Combating against internet worms in large-scale networks: an autonomic signature-based solution”、 SECURITY AND COMMUNICATION NETWORKS、2009 年、11-28、査読有

４．　H.Tsunoda, K.Ohta, A.Yamamoto, N.Ansari, Y.Waizumi, Y.Nemoto 、“Detecting DRDoS attacks by a simple response packet confirmation mechanism”、Computer Communications、No. 3、2008年、3299-3306、査読有

〔学会発表〕（計４件）

１．　Y. Waizumi, Y. Tsukabe, H. Tsunoda, and Y. Nemoto,”Network Application Identification Based on Communication characteristics of Application Messages”,Proc. of WCSET 2009, No. 6, 708-713、2009 年 12 月 26 日、Bangkok、タイ、査読有

２．　Y. Waizumi, T. Sato, and Y. Nemoto, ”A New Traffic Pattern Match-ing for DDoS Traceback Using Independent Component Analysis”,Proc. of WCSET 2009, No. 60,701-707、2009年12月26日、Bangkok、タイ、査読有

３．　S. Yagi, Y. Waizumi, H. Tsunoda, Y. Nemoto,”A Reliable Network Application Identification Based on Transition Pattern of Payload Length”, Proc. of IEEE Globecom 2008, 2008 年 12 月 2 日、New Orleans, 米国、査読有

４．　S.Yagi, Y.Waizumi, H.Tsunoda, A.Jamalipour,N.Kato,Y.Nemoto, ”Network Application Ident-ification Using Transition Pattern of Payload Length”, IEEE WCNC 2008, 2008 年 4 月 2 日、Las Vegas 、米国、査読有

〔産業財産権〕

○出願状況（計１件）

名称：Network Failure Detection Method and Network Failure Detection

発明者：Yuji Waizumi, Hitoshi Tsunoda, Yoshiaki Nemoto

権利者：Tohoku University

種類：特許

番号：US 2009/0265784 A1

出願年月日：May 8, 2008
国内外の別：国外

６．研究組織
(1)研究代表者
根元　義章（NEMOTO YOSHIAKI）
東北大学・理事
研究者番号：60005527

(2)研究分担者
和泉　勇治（WAIZUMI YUJI）
東北大学・大学院情報科学研究科・准教授
研究者番号：90333872

角田　裕　（TSUNODA HIROSHI）
東北工業大学・工学部情報通信工学科・
講師
研究者番号：30400302