氏　　名	Kriangkrai Limthong
学位(専攻分野)	博士(情報学)
学 位 記 番 号	総研大甲第 1796 号
学位授与の日付	平成27年9月28日
学位授与の要件	複合科学研究科 情報学専攻 学位規則第6条第1項該当
学 位 論 文 題 目	Multi-timeline Based Real-time Anomaly Detection in Network Traffic
論 文 審 査 委 員	主　査　准教授 福田 健介 教授 計 宇生 准教授 鯉渕 道紘 教授 菅原 俊治 早稲田大学 特任教授 山田 茂樹 国立情報学研究所

(別紙様式 2)
(Separate Form 2)

論文内容の要旨
Summary of thesis contents

The volume of traffic in both core and access networks has exponentially increased every year over the past few decades. The computer attacks also have increased in sophisticated techniques to evade existing intrusion detection systems. It is rather difficult for daily network operators and administrators to inspect every single packet or flow for discovering anomalies. Therefore, the need to automatically detect attacks and unusual incidents in computer networks is of crucial importance for nowadays operations.

An effective system that could expeditiously detect a broad range of anomalies would enable administrators to prevent serious consequences of anomalies related to network security, availability, or reliability. For over a decade, many researchers have been studying to improve techniques for anomaly detection by proposing and applying plenty of methods from simple to sophisticated ones. Unfortunately, most of the studies are batch processing techniques, and many of them are not fairly flexible to detect a vast variety of anomalies caused by threats or accidents.

In this study, we proposed a detection system using microscopic to macroscopic designs for real-time anomaly detection. The key idea of the proposed system is that the system learns network traffic from multiple timelines rather than a single timeline of input data employed by most conventional detection systems. The advantages of the proposed system are 1) improving on detection performance over the single timeline, 2) flexibility in applying the proposed system to various types of networks or protocols, 3) robustness to incorrect training data or manipulating data by attackers, 4) performance improvement with weighted multiple timelines, and 5) real-time detectability for anomalies caused by threats or accidents. We also performed a series of experiments to examine the proposed system by employing three standard machine learning algorithms, namely multivariate normal distribution, k-nearest neighbor, and one-class support vector machine. In our experiments, we extracted nine key features on account of several selected attacks from a testbed data set. We examined capabilities of the proposed system in many aspects including detection performance, robustness, learning rate, time consumption, different volume of background traffic, time of anomaly occurrence, and weighting for old data.

The experimental results show that the proposed system with machine learning algorithms effectively detected several of anomalies caused by threats or accidents. Our experiment also indicates that the multi-timeline technique outperforms both conventional real-time and a combination of single and multi-timeline. The proposed system shows a robust capability to learn from incorrect training data or manipulating data by attackers. Moreover, two of the three algorithms with the proposed system could learn from training data in reasonable time. The proposed system cannot only enable network administrators to detect novel types of attacks but can be used to identify abnormal behavior of their networks in real time as well.

(別紙様式 3)
(Separate Form 3)

博士論文の審査結果の要旨
Summary of the results of the doctoral thesis screening

本博士論文は「Multi-timeline Based Real-time Anomaly Detection in Network Traffic（マルチタイムラインに基づくリアルタイムネットワーク異常検出)」と題した，インターネットにおけるローカルエリアネットワークでのリアルタイムネットワークトラフィック異常検出に関する研究内容である．

従来のネットワークトラフィック異常検出では，ルールベースに基づくパターンマッチ手法や正常なトラフィックからのずれを統計的処理に基づいて検出する手法が知られている．本論文は後者に関する提案であるが，従来手法の多くは，過去・未来・現在の全てのデータを用いて検出を行うバッチ的な入力データもしくは，過去のデータおよび現在のデータから検出を行うリアルタイム的な入力データを使用している．本研究では，リアルタイム性・アクシデント等の異常データに対する頑強性・検出精度の向上を目指した，マルチタイムラインに基づく入力データを考慮した，特徴量の時系列に基づくリアルタイムネットワークトラフィック異常検出システムを提案している．

論文は８章から構成され，第１章の研究の背景，第２章の関連研究に続いて，第３章では異常検出システムのデザインおよび入力データの検討，第４章ではシステムの実装について述べている．提案のキーとなるアイディアは，過去のトラフィックデータを学習データとして使用する際に，直近の数ステップの特徴量データではなく，過去の同時刻の特徴量データをもとに学習を行う点にある(マルチタイムライン)．これにより，学習のための数ステップのトラフィックに何らかの人為的・機械的な異常があった場合にも誤学習の効果を弱めることが可能となる．また，過去のデータに対して線形重み付けを行うことでシステムの性能向上をはかることを可能としている．第５章のデータセットの収集では，大学の教育用計算機室環境において得られた正常トラフィックデータの収集方法，および異常データとして用いる既存の異常トラフィックについて述べている．第６章では，提案手法および既存手法に関して，９つの性能評価を行っている．これらは，システムの速度・性能に関する評価，パラメータ依存性，特徴量抽出の妥当性，重み付きの学習の効果，既存手法との性能比較，バックグラウンドトラフィック量の影響等が含まれる．主な結果として，既存の入力方式であるリアルタイム的な入力データおよびリアルタイム＋マルチタイムラインの組み合わせ手法と比較して，マルチタイムラインに基づく手法がリアルタイム性を持ち頑強かつより高い検出性能を有することを示している．第７章では，第６章の評価結果をもとに，提案システムの妥当性およびその限界について議論している．第８章では結論と今後の課題について述べている．

なお研究成果として，出願者は主著で電子情報通信学会英文論文誌論文を１編，査読付国際会議論文を２編発表している．

以上要するに，本論文はインターネットトラフィックの異常検出において，従来手法と比較して，リアルタイム性を持ち，アクシデント等の異常データに関する頑強性のある，検出精度の高い技術の実現に貢献するところが大きい．よって，本論文は博士の学位請求論文として合格と認められる．