氏　　名	Romain Thibault FONTUGNE
学位（専攻分野）	博士（情報学）
学位記番号	総研大甲第1456号
学位授与の日付	平成２３年９月３０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	Increasing Reliability in Network Traffic Anomaly Detection
論文審査委員	主　査　准教授　福田　健介 教授　中村　素典 教授　山田　茂樹 准教授　計　宇生 准教授　鯉渕　道紘 教授　菅原　俊治　（早稲田大学）

論文内容の要旨

Network traffic anomalies stand for a large fraction of the Internet traffic and compromise the performance of the network resources. Detecting and diagnosing these threats is a laborious and time consuming task that network operators face daily. During the last decade researchers have concentrated their efforts on this problem and proposed several tools to automate this ta-sk. Thereby, recent advances in anomaly detection have permitted to detect new or unknown anomalies by taking advantage of statistical analysis of the traffic.

In spite of the advantages of these detection methods, researchers have reported several common drawbacks discrediting their use in practice. Indeed, the challenge of understanding the relation between the theory underlying these methods and the actual Internet traffic raises several issues. For example, the difficulty of selecting the optimal parameter set for these methods mitigates their performance and prevent network operators from using them. Moreover, due to the lack of ground truth data, approximate evaluations of these detection methods prevent to provide accurate feedback on them and increase their reliability.

We address these issues, first, by proposing a pattern-recognition-based detection method that overcomes the common drawbacks of anomaly detectors based on statistical analysis, second, by providing a benchmark tool that compares the results from diverse detectors and ground truth data obtained by combining several anomaly detectors.

The proposed pattern-recognition-based detector takes advantage of image processing techniques to provide intuitive outputs and parameter set. An adaptive mechanism automatically tuning its parameter set according to traffic fluctuations is also proposed. The resulting adaptive anomaly detector is easily usable in practice, performs a high detection rate, and provides intuitive description of the anomalies allowing to identify their root causes.

A benchmark methodology is also developed in order to compare several detectors based on different theoretical background. This methodology allows researchers to accurately identify the differences between the results of diverse detectors. We employ this methodology along with an unsupervised combination strategy to combine the output of four anomaly detectors. Thereby, the combination strategy

increases the overall reliability of the combined detectors and it detects two times more anomalies than the best detector.

We provide the results of this combination of detectors in the form of ground truth data containing various anomalies during 10 years of Internet traffic. Thus, with these results researchers can rigorously evaluate their anomaly detectors and increase their reliability. Since the availability of these results on the Internet in December 2010, researchers from many countries (e.g. Japan, France, China, Italy, Czech Republic, Brazil, and U.S.) have manifested their interests in our work, and the website hosting the results has received 236 visits for the single month of June 2011.

博士論文の審査結果の要旨

本博士論文は「Increasing Reliability in Network Traffic Anomaly Detection (ネットワークトラフィック異常検出の信頼性向上に関する研究)」と題し，英文で書かれている．インターネットバックボーン中には正常な通信に相当するパケットデータと，各種の異常と判断すべきパケットデータ(ウィルス・ワーム，フラッシュクラウド，機器の設定ミス，故障など)が存在するが，通常は大多数の正常なデータ中に少数の異常なデータが含まれ，これらを効率よく検出することが重要となっている．博士論文では，異常を効率的に検出するために，時間を含むパケット特徴量空間を2次元画像として扱い，その中に現れる異常を画像処理(エッジ検出)によって発見する異常検出器の設計・実装・評価を行っている．また，複数の理論的バックグラウンドが異なる異常検出器を組み合わせることで，その検出精度を高めるための，ベンチマークアーキテクチャの設計・実装・評価を行っている．

論文は八章と付録Aから構成され，第一章，第二章では研究の目的，背景と関連研究について外観し，既存の手法の問題点を指摘しながら，本研究の必要性について述べている．

続いて，第三章では，提案手法の検証に用いるMAWIデータセットと呼ばれる，日米インターネット回線の10年にわたるトラフィックデータを用いて，既存の異常検出器の出力とその精度についての解析を行っている．

第四章では，画像処理に基づく異常検出器の基本アルゴリズムの設計・実装・評価について述べている．基本アルゴリズムでは，ネットワークを流れるパケットを時間・特徴量空間の二次元空間上の点として捉え，異常なトラフィックが描く二次元空間上の軌跡をハフ変換と呼ばれるエッジ検出アルゴリズムを用いて検出する．基本アルゴリズムでは，既存手法(PCA, KL-based, Sketch-gamma)に比べて効率よくワームトラフィックを検出可能であることが示された．

第五章では，基本アルゴリズムの拡張と自動パラメータチューニングについて議論している．主な拡張は，ハフ変換を用いるための画像を任意の特徴量空間から構成すること，パケット数そのものではなく画像中の点の数に着目することでハフ変換を適用する際の最適画像サイズを動的に変更することである．それにより，基本アルゴリズムと比較して検出率の向上が可能となった．

第六章では，複数の異常検出器を組み合わせるためのベンチマークアーキテクチャについて述べている．異常検出器の結果は時間・特徴量出力(検出イベント)の粒度が異なるため，単純に結果を比較することは困難である．提案手法では，検出イベントをノード，イベント間に共通するパケットを重みとするリンクからなるグラフを構成し，そのグラフ構造から稠密な部分をコミュニティマイニングによって検出する．抽出されたコミュニティを，複数の投票方式によって異常判定することで，最終的に複数の検出器による異常イベントを効率よく発見することが可能となる．評価には前述のMAWIデータセットを用いているが，検出器のタイプと異常のタイプにより精度が異なることから検出器の組み合わせが性能改善に効果があることが示された．

第七章，第八章では，議論および研究成果をまとめ，今後の課題を提示した．

なお，本研究の成果として，出願者は電子情報通信学会論文誌1篇，英文論文誌2篇，査読付国際会議論文5篇，査読付Student workshop2篇の研究発表を行っている．

以上要するに，本論文は高速大容量化するバックボーンネットワーク中での異常トラフィックの検出手法ならびにそのベンチマーク手法を提案・評価することで，その新規性・有用性を示しており，今後のインターネットの信頼性向上に貢献するところが大変大きい．よって，本論文は博士の学位請求論文として合格と認める．