A Dissertation for the Degree of Ph.D. in Engineering

GAMPAL: A General-Purpose Anomaly Detection Mechanism for Internet Backbone Network without Labeled Data

January 2025

Graduate School of Science and Technology,
Keio University

Taku Wakui

別表５
(3)

主　論　文　要　旨

No.1

報告番号	㊙甲 乙 第　　号	氏　名	和久井　拓

主 論 文 題 名：

GAMPAL: A General-Purpose Anomaly Detection Mechanism for Internet Backbone Network without Labeled Data

(GAMPAL：インターネットバックボーンを対象としたラベル付きデータを必要としない汎用異常検知手法)

（内容の要旨）

様々なネットワークを相互接続するインターネットバックボーンは，多様な異常事象の影響を受ける可能性がある．また，膨大なトラフィックが観測され，平常時でも変動が激しく，トラフィックの監視のみで異常事象を検知するのは難しい．

既存研究では，ネットワーク規模に対するスケーラビリティへの考慮や，トラフィックの局所性を考慮した方式が少ない．また，ラベル付きデータを用いて，異常の特徴をモデルに学習させる方式が多いが，ラベル付きデータは生成が難しく，また，検知対象がラベルで定義された異常に限定されてしまう．

本学位論文は，異常なトラフィックの挙動を汎用的に検知する GAMPAL（General-purpose Anomaly detection Mechanism using Prefix Aggregation without Labeled data）を提案する．GAMPAL はトラフィック情報として，5 タプルが共通するパケットを集約したフロー情報を採用する．膨大なフローを分析するため，BGP（Border Gateway Protocol）の経路表を用いたフロー集約方式 Prefix Aggregation を定義する．Prefix Aggregation はトラフィックの局所性に着目し，BGP における経路情報である AS path の一部が同じフローを空間的に集約する．この集約単位を PA（Prefix Aggregate）とする．また，時間的な集約のため，5 分間のタイムスロットでバイト数を集約したフローサイズの時系列データを生成する．この時系列データで，集約単位毎に予測モデルを構築する．GAMPAL は，フローサイズの予測値と実測値を比較することで，ラベル付きデータを使わずに，汎用的に異常を検知する．

本学位論文は，GAMPALv1 と GAMPALv2 を提案している．v1 では，トラフィックの様々な周期性を考慮し，LSTM-RNN（Long Short-Term Memory-Recurrent Neural Network）を用い，フローサイズの時系列データから，1 つ次のタイムスロットのフローサイズを予測するモデルを構築した．また，フローサイズの予測値と実測値の差分評価のために，2 つの時系列データの相違の割合を評価する評価指標 NSD (Normalized Summation of Difference) を定義した. 評価では, 実社会で発生した異常事象を題材に, 1 日分のフローサイズの予測値と実測値から PA 毎に NSD を求め, 全 PA の平均を算出

主　論　文　要　旨

した．その結果，異常のあった日は NSD の平均が大きいことを確認した．

一方，v1 には 4 つの課題があった．モデル学習と予測値算出に要する計算量が膨大であること．モデル学習に時間のセマンティクスを活用できていないこと．検知の閾値設定が困難なこと．そして，異常が検知された PA や時間帯を特定できないことである．

1 つ目と 2 つ目の課題に対し，v2 は各タイムスロットから時間に関する特徴量を抽出した時間特徴量とフローサイズの関係をモデル化した．また，この方式に適した RFR（Random Forest Regressor)をモデルとして採用した．モデルへの入力の次元数は 288 次元から 7 次元に削減され，また RFR は LSTM-RNN よりも計算量が少ない．また，3 つ目と 4 つ目の課題に対し，v2 はモデルの予測結果から予測範囲を定義し，予測範囲と実測値の比較による異常検知を提案した．

v2 の評価でも，v1 の評価で使用した異常事象を対象に，予測範囲を算出し実測値と比較した．その結果，観測点のある大学ネットワークを宛先とするフローが属する PA では，異常のあった日のみ，実測値が予測範囲を逸脱し異常として検知された．また，ラベル付きデータを用いて生成したデータセットによる精度評価では，再現率は 93.1%を記録した．評価方式やデータは異なるが，再現率は v1 の 81.8%から向上した．また，v1 では，モデル学習と予測値算出に，GPU を用いて約 4 日要していたが，v2 では 8 コア CPU を用いて約 23 分であった．以上の評価で，GAMPAL の精度と実用性の改善を確認した．

別表６
（３）

Keio University

Thesis Abstract

No. 1

Registration Number	☑ “KOU” □ “OTSU” No. *Office use only	Name	Taku Wakui

Thesis Title

GAMPAL: A General-Purpose Anomaly Detection Mechanism for Internet Backbone Network without Labeled Data

Thesis Summary

The Internet backbone is a network that interconnects various networks and can be affected by various anomalies. In addition, since enormous traffic is observed and changes rapidly even during ordinary times, detecting anomalies only by monitoring traffic is difficult.

In existing research, only a few methods consider scalability to the network and locality of traffic. In addition, most methods learn anomaly features using labeled data, which is difficult to generate and limits the detection target to anomalies defined by labels.

This dissertation proposes GAMPAL (General-purpose Anomaly detection Mechanism using Prefix Aggregation without Labeled data), which detects anomalous traffic behavior generally. GAMPAL uses flow information, the traffic data aggregating a packet sequence with a common 5-tuple. To analyze enormous flows, GAMPAL defines prefix aggregation, a flow aggregation method using the routing table in BGP (Border Gateway Protocol) and PA (Prefix Aggregate), a unit of aggregated flows based on prefix aggregation. Focusing on the locality of Internet traffic, flows are aggregated spatially based on the part of the AS (Autonomous System) path by prefix aggregation. In addition, to aggregate flows temporally, time series data of the flow size is generated by aggregating the number of bytes in a 5-minute time slot. Using this time-series data, prediction models are established for each PA. GAMPAL detects anomalies generally without labeled data by comparing predicted and observed flow sizes.

This dissertation proposes GAMPALv1 and GAMPALv2. Considering various periodicities of traffic, v1 uses LSTM-RNN (Long Short-Term Memory-Recurrent Neural Network). The prediction model takes time-series data of flow size and predicts the subsequent flow size. In addition, to evaluate the difference between the predicted and observed flow size, v1 defined NSD (Normalized Summation of Difference), which evaluates the ratio of the difference between two time-series data. In the evaluation, anomalies in the real world were used as targets, and the NSDs were calculated for each PA based on the predicted and observed flow size for one day. Then, the average of the NSDs for all PAs was calculated, and it was confirmed that the average NSD was larger on days with anomalies.

On the other hand, v1 has four issues. Its computational complexity for training and predicting is enormous. The semantics related to time cannot be utilized to train the model. Determining the detection threshold is difficult. V1 cannot identify when and at which PAs an anomaly is detected.

To address the first and second issues, v2 proposes to model the relationship between time features extracted from each time slot and the flow size. In addition, RFR (Random Forest Regressor), suitable for this method, is adopted as the model. The number of dimensions of input to the model is reduced from 288 to 7, and RFR has less computational complexity than LSTM-RNN. To address the third and

Thesis Abstract

fourth issues, v2 defines a prediction range based on the prediction results of the model. V2 detects anomalies by comparing the prediction range and the observed flow sizes.

In evaluating v2, the prediction range is calculated for the anomalies used in evaluating v1 and compared with the observed values. As a result, the observed values are out of the predicted range, and v2 detects anomalies in the PAs, to which the flows are destined for the university network, where the observation point is located only on the days when anomalies occur. In addition, the accuracy evaluation using a dataset generated with labeled data shows a recall of 93.1%. Although the evaluation methods and data differ, the recall improves from 81.8% in v1. In addition, GAMPALv1 requires four days on a GPU to train the model and predicts flow size for one day, while the GAMPALv2 takes approximately 23 minutes on an 8-core CPU. These evaluations confirmed the improved accuracy and practicality of GAMPAL.