5ZC-03

クラウドサービスへのオフローディングを活用した機械学習によるNIDSの開発

栗原 優斗 †　　中村 純哉 ‡
豊橋技術科学大学 †　豊橋技術科学大学 ‡

1　はじめに

ネットワーク型侵入検知システム（NIDS）は，ネットワークを流れるトラフィックを監視し，攻撃通信を検知するシステムである．サイバー攻撃を早期に気づくことができるため，攻撃の被害を抑えるために有効な手段である．近年では，未知の攻撃に対しても検知が可能である，機械学習に基づくNIDSが研究されている．しかし，機械学習に基づくNIDSは動作のために高い計算能力を備えたコンピュータが必要であり，導入できる環境が限られている．本研究では，高い計算能力を必要とする処理をクラウドサービスへオフローディングする機械学習に基づくNIDSのプロトタイプを作成し，その性能を評価する．

2　システム構成

本研究では，多田らが提案した機械学習に基づくNIDSの分散処理フレームワーク [2] をもとに，Amazon Web Services (AWS) 上に機械学習に基づくNIDSを構築した．処理件数の変化に応じて適切な量のリソースを配分するために，それぞれのクラウドサービスにはストレージサイズやインスタンス数の動的スケーリングが可能であるマネージドサービスを用いた．図1に，構築したNIDSの構成を示す．Amazon EMRは分散処理フレームワークを実行するためのマネージドサービスであり，インスタンス数やストレージのサイズのオートスケーリングが可能である．機械学習モデルによる推論のための特徴量を抽出するために利用した．Amazon Sagemakerは，AWS上で機械学習モデルを利用するための一連の流れを実行できるフルマネージドサービスである．訓練済みの分類モデルをデプロイすることにより，推論のためのエンドポイントとして利用した．Amazon MSKは，ストレージのオートスケーリングが可能なメッセージキューのフルマネージドサービスである．オンプレミス環境からクラウド環境へのセッション情報の送信処理から，Amazon Sagemakerでの推論処理までの間の非同期処理を実現するために利用した．AWS Lambdaはイベントの発生に応じてプログラムを実行可能なサービスであり，イベントの発生件数に応じて同時実行数のオートスケーリングが可能である．Amazon MSKとAmazon Sagemakerの間のデータの中継に利用した．

本システムにおける処理の流れは，次のよ

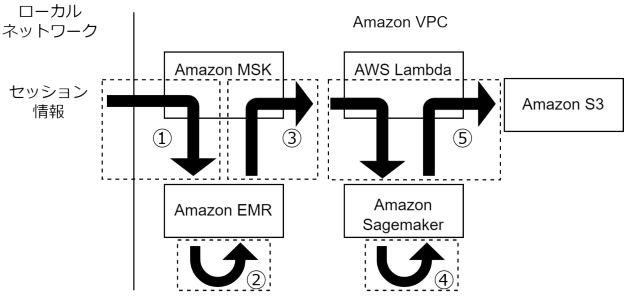

図1　提案システムの構成

Development of ML-based NIDS with offloading to cloud services
† Yuto Kurihara, Toyohashi University of Technology
‡ Junya Nakamura, Toyohashi University of Technology

うになる．まず，ローカルネットワークで得られたセッション情報を基本特徴量として Amazon MSK の Kafka Broker に送信する．次に，Amazon EMR 上に構築した Apache Spark にて機械学習モデルの推論に必要な特徴量を抽出し，完全な特徴量として Amazon MSK に書き戻す．完全な特徴量は AWS Lambda によって Amazon Sagemaker の推論エンドポイントに送られ，事前に訓練した機械学習モデルで攻撃通信かどうかを分類する．分類結果は，AWS Lambda を通じて Amazon S3 に保存される．

3 評価実験

本研究で作成した機械学習に基づく NIDS の処理性能を評価するために，図 1 に示す各区間及び全体でのスループットとレイテンシを計測する．本実験では，スクリプト作成のための時間的都合及び後述するデータセットにすでに特徴量抽出済みのデータが含まれていることから，ネットワークトラフィックからセッション情報を抽出する部分，Amazon EMR での特徴量抽出の部分を除いて実装をした．計測のためのセッション情報には，Amazon VPC 上の Amazon EC2 インスタンスから送信した UNSW-NB15 データセット [1] を用いた．Amazon Sagemaker の推論エンドポイントには，UNSW-NB15 データセットで事前に訓練したランダムフォレストのモデルを用いた．

レイテンシの計測結果を表 1 に，スループットの計測結果を表 2 に示す．表 1 より，処理件数が増加すると，Amazon EMR から AWS Lambda への通信のレイテンシが非常に長くなっていることがわかる．区間 2 は，Amazon EMR での特徴長抽出処理を実装していないため，レイテンシが 0.00 となっている．それ以外の区間は，処理件数の増加に対するレイテンシの増加は区間 3 と比較して緩やかである．また，表 2 より，データの流量が区間 3 から約 5300 session/s に抑えられていることがわかる．このことから，本研究で作成した NIDS では区間 3 での通信が全体の処理性能のボトルネックとなっており，最大スループットは約 5000 session/s であると考えられる．考えられる原因としては，AWS Lambda でのコールドスタートが多く発生していることが挙げられる．そのため，想定される処理件数に応じて AWS Lambda の関数インスタンスをプロビジョニングすることによって改善する可能性がある．

表 1　各区間でのレイテンシ [ms]

送信速度 [session/s]	区間 1	区間 2	区間 3	区間 4	区間 5	全区間
1	2264.42	0.00	1074.22	107.42	0.18	3446.24
10	2021.44	0.00	1062.33	115.85	0.23	3199.85
100	1798.38	0.00	1063.10	168.62	1.31	3031.41
1000	1674.28	0.00	1110.24	223.92	12.50	3020.93
10000	4595.60	0.00	52332.52	237.48	63.46	57229.06

表 2　各区間でのスループット [session/s]

送信速度 [session/s]	区間 1	区間 2	区間 3	区間 4	区間 5	全区間
1	0.98	1.03	1.03	1.03	1.03	0.98
10	9.90	10.31	10.28	10.32	10.32	9.86
100	93.78	97.61	97.26	97.67	97.67	93.40
1000	961.70	1000.66	996.98	1001.04	1001.04	957.65
10000	9236.70	10367.73	5281.70	5313.61	5313.61	4965.88

4 おわりに

本研究では，クラウドサービスへのオフローディングを活用した機械学習に基づく NIDS のプロトタイプを開発し，その処理性能を評価した．評価では，Amazon EMR から AWS Lambda への通信がボトルネックとなり，最大スループットは約 5000 session/s であった．より多くのデータの処理を可能にするためには，Amazon EMR から AWS Lambda への通信のレイテンシを抑えることが必要である．

参考文献

[1] N. Moustafa and J. Slay. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *Proc. MilCIS 2015*, pp. 1–6, 2015.

[2] 多田, 中村, 大村, 小林. 機械学習ベース NIDS 構築のための分散処理フレームワーク. 情報処理学会論文誌, 60(9):1448–1465, 2019.