جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Kawtar BOUZOUBAA

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Samedi 8 Juillet 2023 à 10H au Grand amphi à l'ENSAM de Rabat

Intitulé de la thèse

OPTIMISATION DES MODÈLES D'APPRENTISSAGE AUTOMATIQUE APPLIQUÉS AUX ATTAQUES DOS/DDOS : EVALUATION D'IMPACT DU PROCESSUS DE SÉLECTION DES CARACTÉRISTIQUES

Devant le Jury composé de :

Président :

Pr. Rachid OULAD HAJ THAMI, PES, ENSIAS, Université Mohammed V de Rabat

Directeur de thèse :

Pr. Benayad NSIRI, PES, ENSAM, Université Mohammed V de Rabat

Co-Directeur de thèse :

Pr. Youssef TAHER, PH, Centres d'orientation et de planification pédagogiques (COPE), Rabat

Rapporteurs :

Pr. Najlae IDRISSI, PES, FST, Université Sultan Moulay Slimane, Béni Mellal

Pr. Nabila RABBAH, PES, ENSAM, Université Hassan II, Casablanca

Pr. Abderrahim EL QADI, PES, ENSAM, Université Mohammed V de Rabat

Examinateur :

Pr. Mohamed ZERIAB ES-SADEK, PH, ENSAM, Université Mohammed V de Rabat

UNIVERSITE MOHAMMED V - RABAT ENSIAS

Résumé: Ces dernières années ont été marquées par une explosion du taux d'attaque en cybersécurité. Ces attaques sont de plus en plus sophistiquées et représentent une réelle menace croissante pour les individus, les secteurs privé et public qui sont de plus en plus dépendants d'Internet. Une des principales menaces informatiques auxquelles sont confrontées les organisations gouvernementales et les entreprises sont l'attaque par Déni de Service (DoS) et ses sous-variantes (DDoS, RDoS, etc.). L'impact des attaques DoS et ses sous-variantes est assez conséquent et peut ainsi entraîner pour la structure concernée: une perte financière causée par des services inaccessibles, arrêt total du système d'information, des dommages matériaux, la fermeture de l'organisation, ainsi que d'autres dégâts plus conséquents.

Les types de menaces DoS sont de plus en plus complexes, fréquentes, et parfois difficiles à combattre. Par conséquent, lutter contre ces attaques à l'aide de solutions de protection conventionnelles (traditionnelles) telles que les pares-feux logiciels/matériels et les anti-virus/antimalwares deviennent insuffisants et présentent de nombreux inconvénients (ne peuvent pas détecter à titre d'exemple en temps réel quand, où et comment les nouvelles formes des attaques DoS se produisent).

Pour remédier aux insuffisances de ces solutions conventionnelles, les départements de sécurité peuvent tirer un avantage important des grands volumes de données générés par les trafics réseaux et les cyber-attaques. À titre d'exemple, les modèles prédictifs basés sur les algorithmes sophistiqués et performants de l'apprentissage automatique (Machine Learning-ML) peuvent augmenter le taux de détection des attaques DoS/DDoS à un stade précoce (en temps quasi réel). Dans ce cadre, ce travail de thèse porte une attention particulière sur un des processus clés d'optimisation et d'amélioration de ces modèles prédictifs appliqués aux différentes variantes des attaques DoS/DDoS. Il s'agit du processus de Sélection des Caractéristiques (SC) des attaques DoS/DDoS. Cette étude de thèse a porté sur plus de cent trois importantes références scientifiques théoriques et pratiques ayant intégré le processus de sélection de caractéristiques dans des projets ML de cybersécurité.

Afin d'optimiser le processus de SC de l'attaque DoS/DDoS, nous avons exploité cette base de données bibliographiques par la mise en œuvre d'un modèle d'analyse hiérarchique sur trois niveaux d'analyse de performances. Le premier niveau d'analyse a pour objectif d'évaluer l'impact du choix parmi les quatre principales catégories de méthodes de sélection des caractéristiques : Filtre, Enveloppement, Hybride et Embarqué, sur le processus de sélection pour le cas de l'attaque DoS/DDoS. Le deuxième niveau d'analyse a pour objectif d'évaluer l'impact du choix des sous-méthodes de SC utilisées dans chaque classe étudiée dans le premier niveau sur le processus de sélection des caractéristiques DoS/DDoS. Le troisième niveau d'analyse a comme objectif d'évaluer l'impact des ensembles de données DoS-DDoS les plus utilisés et publiquement

disponibles (KDD‘99, NSL_KDD, UNSW_NB15, CIC_IDS2017, CIC_IDS2018) sur le processus de sélection des caractéristiques pour le cas de l‘attaque DoS/DDoS. Enfin, nous avons exposé les caractéristiques DoS/DDoS les plus sélectionnées quelle que soit la méthode utilisée par ensemble de données DoS/DDoS, ainsi que leurs catégories.

Mots-clés: Apprentissage automatique, DoS/DDoS attaques, processus, processus d'enveloppement, processus de filtrage, processus hybride, processus intégré, sélection de caractéristiques.

Abstract: The last few years were marked by an explosion in the rate of cyber security attacks. These attacks are increasingly sophisticated and represent a real and growing threat to individuals, private and public sectors that are increasingly dependent on the Internet.

One of the main cyber threats faced by government organizations and businesses is the Denial of Service (DoS) attack and its sub-variants (DDoS, RDoS, etc.). The impact of DoS attacks and its sub-variants is quite significant and can result in the following for the affected organization: financial loss due to inaccessible services, total shutdown of the information system, material damage, closure of the organization, as well as other important damages. The types of DoS threats are becoming increasingly complex, frequent, and sometimes difficult to combat. Therefore, fighting these attacks with conventional (traditional) protection solutions such as software/hardware firewalls and antivirus/antimalware is becoming insufficient and have many drawbacks (e.g. cannot detect in real time when, where and how new forms of DoS attacks occur).

To overcome the shortcomings of these conventional solutions, security departments can take significant advantage of the large volumes of data generated by network traffic and cyber attacks. For example, predictive models based on sophisticated and powerful Machine Learning (ML) algorithms can increase the detection rate of DoS/DDoS attacks at an early stage (in near real time). In this context, this thesis focuses on one of the key processes for optimizing and improving these predictive models applied to different variants of DoS/DDoS attacks. This is the Feature Selection (FS) process of DoS/DDoS attacks. This dissertation study reviewed more than one hundred and three important theoretical and practical scientific references that have incorporated the feature selection process into ML model-based cyber security projects. In order to optimize the DoS/DDoS attack FS process, we have exploited this bibliographic database by implementing a hierarchical analysis model on three levels of performance analysis. The first level of analysis aims to evaluate the impact of the choice among the four main categories of FS methods: Filter, Wrapping, Hybrid and Embedded, on the selection process for the DoS/DDoS attack case. The second level of analysis aims to evaluate the impact of the choice of FS submethods used in each class studied in the first level on the DoS/DDoS FS process. The third level of

analysis has the objective of evaluating the impact of the most used and publicly available DoS-DDoS datasets (KDD'99, NSL_KDD, UNSW_NB15, CIC_IDS2017, CIC_IDS2018) on the feature selection process for the DoS/DDoS attack case. Finally, we outlined the most selected DoS/DDoS features regardless of the method used per DoS/DDoS dataset and their categories.

Keywords: DoS/DDoS attacks, embedded process, feature selection, filtering process, hybrid process, machine learning, wrapping process.