وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم البرمجيات

كشف التطفل بالاعتماد على التعلم الفوقي

رسالة مُقَّدمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
البرمجيات

من قبل

زهراء هاني سالم حمدون

بإشراف
أ.د. صفوان عمر حسون خليل

الخلاصة

في ظل تزايد التهديدات السيبرانية والتطور المستمر للهجمات على الشبكات وأنظمة المعلومات، يعد تطوير أنظمة كشف التطفل (IDS) باستخدام تقنيات الذكاء الاصطناعي أمراً حيوياً. تهدف هذه الدراسة إلى تطوير نظام كشف تطفل يعتمد على التعلم الفوقي، وذلك لتعزيز كفاءة ودقة كشف الهجمات وتقليل معدلات الإنذارات الكاذبة. يتم تحقيق ذلك من خلال دمج تقنيات التعلم الآلي والتعلم العميق والتعلم الفوقي، النظام تكون من مرحلتين ،حيث تم توظيف عدد من خوارزميات التعلم الالي كمتعلمين اساسين وخوارزمية تعزيز التدرج الأقصى XGBoost كمتعلم فوقي في المرحلة الاولى للتنبؤ بالهجوم، والمرحلة الثانية لكشف نوع الهجوم باستخدام خوارزميات الشبكة العصبية التلافيفية(CNN) ، والشبكة العصبية المتكررة(RNN) ، والذاكرة الطويلة والقصيرة المدى (LSTM) في المرحلة الأساسية، بينما استخدمت تقنية التعلم الفوقي باستخدام الانحدار اللوجستي (LR) كمتعلم فوقي للتنبوء وتحسين دقة التصنيف.

تعتمد منهجية البحث على تحليل ومعالجة مجموعتي بيانات NSL-KDD و IoTID20 لاختبار أداء النماذج المقترحة، حيث تضمنت الدراسة مراحل أساسية شملت جمع وتحليل البيانات من خلال معالجة القيم المفقودة وحذف الميزات الأقل ارتباطاً بالهدف عن طريق حساب معامل الارتباط والانحراف المعياري لايجاد الميزات الأقل ارتباطا وازالتها ،واختيار الميزات الأكثر تأثيراً باستخدام خوارزمية الغابة العشوائية، وبناء النماذج عبر تطوير خوارزميات تعلم آلي وتعلم عميق متبوعة بتطبيق التعلم الفوقي لدمج تنبؤات النماذج وتحقيق تحسين في الأداء، كما تم ضبط المعلمات باستخدام البحث الشبكي والبحث العشوائي لتحسين دقة الكشف، وتقييم الأداء بالاعتماد على مقاييس الدقة ومعدل الإنذارات الكاذبة ومعامل F1 والاستدعاء.

تم اعتماد مجموعتي بيانات أساسيتين في الدراسة، حيث استخدمت مجموعة بيانات NSL-KDD التي تحتوي على أنواع مختلفة من الهجمات لاختبار أداء التصنيف الثنائي ومتعدد الفئات، بينما تم اعتماد مجموعة بيانات IoTID20 المتخصصة في كشف التطفل على شبكات إنترنت الأشياء لاختبار أداء النماذج في البيئات الذكية.وتم تقسيم مجموعتي البيانات الى بيانات تدريب بنسبة ٧٠٪ وبيانات اختبار بنسبة ٣٠٪ ، وقد تم تنفيذ النظام المقترح باستخدام لغة بايثون وعلى منصة Google Colab باستخدام

مكتبات تعلم الآلي والتعلم العميق مثل TensorFlow و Keras و Scikit-learn وXGBoost ، كما تم تشغيل النماذج على وحدة معالجة الرسومات لتحسين كفاءة التدريب وتسريع عملية المعالجة.

أظهرت النتائج أن نماذج التعلم الفوقي حققت أداءً متفوقاً مقارنة بالنماذج التقليدية، حيث بلغت دقة التصنيف 99.03٪ بعد ضبط المعلمات الفائقة مقارنة بـ 98.09٪ عند استخدام نماذج التعلم العميق فقط، كما ساهم التعلم الفوقي في تقليل معدل الإنذارات الكاذبة وتحسين استجابة النظام للهجمات الجديدة، مما يعكس فاعلية النظام في التعامل مع التهديدات السيبرانية المختلفة. بناءً على هذه النتائج، توصي الدراسة بتبني تقنيات التعلم الفوقي في أنظمة كشف التطفل المستقبلية، لا سيما في بيئات إنترنت الأشياء، لتعزيز دقة التصنيف وتقليل تأثير الهجمات السيبرانية. يمثل هذا البحث خطوة مهمة نحو تطوير أنظمة كشف التطفل الذكية، حيث يفتح آفاقاً جديدة للاستفادة من تقنيات التعلم الفوقي والذكاء الاصطناعي في تعزيز أمن الأنظمة الرقمية في ظل تصاعد التهديدات السيبرانية.

Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Software

Intrusion detection based on meta learning

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Master in
Software**

by

Zahraa Hani Salim Hamdoon

Supervised by

Prof. Dr. Safwan Omar Hasson Khalel

2025 A.D. **1446 A.H.**

Abstract

With the increasing cyber threats and the continuous development of attacks on networks and information systems, the development of intrusion detection systems (IDS) using artificial intelligence techniques is vital. This study aims to develop an intrusion detection system based on meta-learning, in order to enhance the efficiency and accuracy of attack detection and reduce false alarm rates. This is achieved by integrating machine learning, deep learning and meta-learning techniques. The system consists of two stages, where a number of machine learning algorithms were employed as base learners and the XGBoost maximum gradient boosting algorithm as an over-learner in the first stage to predict the attack, and the second stage to detect the type of attack using convolutional neural network (CNN), recurrent neural network (RNN), and long short-term memory (LSTM) algorithms in the basic stage, while the meta-learning technique using logistic regression (LR) was used as an over-learner to predict and improve classification accuracy. The research methodology is based on analyzing and processing the NSL-KDD and IoTID20 datasets to test the performance of the proposed models. The study included basic stages that included collecting and analyzing data by processing missing values and deleting features that are less related to the target by calculating the correlation coefficient and standard deviation to find and remove the less related features, selecting the most influential features using the random forest algorithm, and building models by developing machine learning and deep learning algorithms followed by applying meta-learning to integrate model predictions and achieve performance improvement. The parameters were also adjusted using grid search and random search to improve detection accuracy, and performance was evaluated based on accuracy measures, false alarm rate, F1 coefficient, and recall. Two main datasets were adopted in the study, where the NSL-KDD dataset containing different types of attacks was used to test the performance of binary and multi-class classification, while the IoTID20 dataset specialized in detecting intrusions on Internet of Things networks was used to test the performance of the models in smart environments. The two datasets were divided into 70% training data and 30% testing data. The proposed system was implemented using the Python language and, on the Google Colab, platform using machine learning and deep learning libraries such as TensorFlow, Keras, Scikit-learn, and XGBoost. The models were also run on a graphics processing unit to improve training efficiency and speed up the processing process. The results showed that the meta-learning models achieved superior performance compared to traditional models, as the classification accuracy reached 99.03% after

adjusting the hyperparameters compared to 98.09% when using deep learning models only. Meta-learning also contributed to reducing the false alarm rate and improving the system's response to new attacks, reflecting the effectiveness of the system in dealing with various cyber threats. Based on these findings, the study recommends the adoption of meta-learning techniques in future intrusion detection systems, especially in IoT environments, to enhance classification accuracy and reduce the impact of cyber-attacks. This research represents an important step towards the development of intelligent intrusion detection systems, as it opens new horizons for leveraging meta-learning and AI techniques to enhance the security of digital systems in the face of escalating cyber threats.