جامعة الموصل

كلية الهندسة

تحليل وتنفيذ نظام كشف المتطفلين باستخدام تعليم الآلة

رسالة تقدمت بها

زينة خالد ابراهيم

علوم في الهندسة الكهربائية / الهندسة الكهربائية

بإشراف

الدكتور محمد يونس ذنون

1442 هـ **2021 م**

الملخص

تطور التقنيات الحديثة مثل إنترنت الأشياء والحوسبة السحابية وشبكات التواصل الاجتماعي، أدى الى وجود أعداد هائلة من حركة مرور الشبكة والبيانات. لذلك هناك حاجة ضرورية لأنظمة كشف التطفل التي تتحكم في الشبكة وتحلل حركة المرور البيانات الواردة بشكل متكرر. في هذا البحث، تم استخدام ثلاث خوارزميات ومجموعة بيانات NSL-KDD لحساب دقة الكشف لخوارزميات التعلم الآلي لنظام كشف التطفل. وتم تقليل عدد الميزات وتحديد مجموعة من الميزات الأكثر أهمية من أجل العثور على الدقة الأفضل بين الخوارزميات المستخدمة وذلك من خلال تحديد الميزات باستخدام تقنية حذف الميزات التكرارية (RFE) وتقنية تحليل التباين (ANOVA F Test). لقد تم تصميم نظام كشف تطفل (IDS) الذي يستخدم خوارزميات التعلم الآلي، وهي خوارزمية الغابة العشوائية Random Forest(RF) و خوارزمية متجه دعم الآلة Support Vector Machine (SVM) وخوارزمية الجار الأقرب K Nearest Neighbor (KNN) لتصنيف كل نوع من الهجمات(Binary Classification) . لقد أظهرت المقارنة أداء النموذج قبل وبعد اختيار عدد محدد من الميزات. لقد تم عرض أيضًا مصفوفات التشويش وتم كذلك تصميم نموذج كشف التطفل باستخدام الخوارزميات الثلاثة لكشف جميع الهجمات (Multiclass Classification) ومقارنة الأداء. أظهرت النتائج أداء جيد لخوارزمية الغابة العشوائية عند استخدام جميع ميزات و13 ميزة وتفوقت على كل من خوارزمية متجه دعم الآلة وخوارزمية الجار الأقرب لكشف كل نوع من الهجمات (DoS ,Probe ,R2L, U2R) وعند الكشف لجميع الهجمات كان أداء خوارزمية الغابة العشوائية افضل عند تقليل عدد ميزات الى 13 ميزة حيث كانت نسبة دقة الكشف(% 99.459) بينما كان أداء خوارزمية متجه دعم الآلة افضل عند استخدام جميع الميزات بنسبة (97.047%) وكذلك بالنسبة لخوارزمية الجار الأقرب حيث كانت دقة الكشف (95.085%)

Abstract

With the use of modern technologies such as the Internet of Things, cloud computing, and social networks, massive numbers of network traffic and data are generated. So there is an essential need for intrusion detection systems for the safety net that control and manage the network and analyze the frequent incoming traffic. In this paper, three algorithms and an NSL-KDD dataset were used to calculate the accuracy of the machine learning algorithms of the Intrusion Detection System. The number of features was reduced and a set of more important features was identified in order to find the best accuracy among the algorithms used and reduce the percentage of positive errors by identifying features using Recursive Feature Elimination (RFE) and ANOVA F test. We have designed an Intrusion Detection System (IDS) that uses machine learning algorithms, namely (RF) Random Forest, Support Vector Machine (SVM), and K Nearest Neighbor (KNN) to classify each type of attack. The comparison showed model performance before and after feature selection for Random Forest, K Nearest Neighbor (KNN) and Support Vector Machine SVM algorithms. We have also introduced Confusion Matrix and an intrusion detection model designed using all three algorithms to detect all attacks and compare performance before and after feature selection.

University of Mosul

College of Engineering

Analysis and Implementation of Intrusion Detection System Using Machine Learning

A Thesis Submitted By

Zena Khalid Ibrahim

M.Sc Thesis

Electrical Engineering

Supervised by

Dr.Mohammed Younis Thanoun

1442 A.H **2021 A.D.**