كشف الأنشطة الضارّة في بيئة إنترنت الأشياء باستخدام تقنية هجينة في تعلّم الآلة

إعـداد

أناس نوفان البطاينة

إشراف

الدكتور محمود العمري

الملخص

إنترنت الأشياء من أحدث التطورات التكنولوجية، ويعزو هذا المصطلح إلى عملية ربط الأجهزة على شبكة واحدة. مما يؤدي إلى ظهور أنواع عديدة من البيانات النافعة و الضارة. الحجم الهائل من البيانات التي تنتج عن الحساسات والسيرفرات والأجهزة غالباً ما يحتوي على بيانات غير مفيدة تؤثر بشكل سلبي على أداء الشبكة وتخفض مستوى أمانها , ففي هذه الرسالة تم طرح استخدام خوارزميات التجميع (Clustering) مع خوارزميات تعلم الآلة للحصول على نتائج أفضل في تمييز البيانات الضارة من الجيدة. تم تطبيق الخوارزميات على مجموعة بيانات من بيئة افتراضية لإنترنت الأشياء(DS2OS traffic traces) تم الحصول عليها من موقع الانترنت Kaggle. حيث استخدم لغة البرمجة بايثون لأنها تسهل وتسرع عمل خوارزميات تعلم الآلة بشكل كبير. تمت مقارنة خوارزميتي التجميع SOM و K-Means بالإضافة إلى خمسة مصنفات XGboost, SVM, NB, K-NN, RF استخدمت المقاييس التالية للتحقق من أداء الطريقة المقترحة F1 Score, Accuracy with 8-fold Cross Validation, Confusion Matrix . وقد أظهرت النتائج أن مصنف XGboost هو الأفضل في كافة المقاييس عند إجراء التجارب قبل تطبيق خوارزميات التجميع بنسبة 94.49%

دقة و نسبة 93.39% في F1 Score, عند تطبيق مجمّع K-Means و لوحظ ارتفاع في دقة كافة المصنفات ولكن بقيت XGboost في المقدمة بنسبة دقة 96.59% و 96.49% في F1 Score , وعند المقارنة مع خوارزمية التجميع خريطة التنظيم الذاتية (SOM) ظهرت فاعلية كبيرة في الأداء و تحسنت دقة نتائج المصنفات بشكل كبير. حيث أن أفضل نتيجة تم الحصول عليها عند دمج مصنف XGBOOST مع خوارزمية التجميع SOM وكانت الدقة 99.97% و 99.49% F1 Score .

Malicious Activities Detection in IoT Environment Using a Hybrid Technique in Machine Learning

Prepared by:

Enas Nofan Al.bataineh

Supervised by:

Dr.Mahmoud Omari

Abstract

The current study is a new trend in tackling the issue of Internet of things which is considered the most recent developed technology because it connects all types of machines on the same network. However, there are a huge number of different types of traffic including both normal and malicious. This huge amount of data generated from sensors, servers and clients which usually contain malicious traffic that negatively impact on the performance of the network and degrades its security. This study proposes a method of using clustering algorithms along with machine learning to achieve better results to differentiate between malicious and good traffic. To do so, python programming language was used for the implementation due its ease of use in machine learning algorithms. The algorithms were applied to a virtual IOT

environment traffic dataset (DS2OS traffic traces) which are obtained from Kaggle website, whereas the clustering algorithms used are SOM and K-Means with the machine learning classifiers XGBOOST, KNN, SVM, NB, and Random Forests. The metrics that were used to verify the work proposed are F1 Score, Confusion Matrix and Accuracy with 8-fold Cross Validation, XGboost achieved the best results in overall metrics before applying any clustering to the encoded dataset at 94.49% accuracy and 93.39% F1 Score, the introduction of clustering algorithms showed an enhancement in the accuracy of the classifiers, but XGboost remained the highest with K-Means the accuracy was 96.59% and 96.49% for F1 Score. It is worth mentioning that the best results of 99.97% accuracy and 99.49% F1 Score were obtained when SOM and XGBOOST were combined together.