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

Classification of Networking Threats based on Modified SIEM Platform and Machine Learning Techniques

**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 Doctor of Philosophy in
Computer Science**

By

Zeyad Safaa Younus Tlayea

Supervised by

Prof. Dr. Mafaz Mohsin Khalil Hasan

2025 A.D. **1447 A.H.**

Abstract

The rapid development of networks and the heavy dependence on them by countries and organizations have raised their exposure to cyberattacks. Many traditional security systems are available to protect systems and network resources against cyberattacks, but they operate separately, leading to fragmented and imprecise evaluation of network resources' security posture. Thus, more effective defensive solutions are required. Security Information and Event Management, known as SIEM, acts as one of the leading security solutions used to increase the information security level and data protection through centrally managing network devices. Small and medium organizations rely on open-source SIEM solutions due to financial and resource limitations. However, open-source SIEMs face some difficulties, like log analysis complexity and the probability of alert overloading that reduces the accuracy rate and increases the number of false alerts. Additionally, SIEM solutions from open-source platforms do not provide machine learning (ML) capabilities as standard built-in functionalities. Where ML techniques can assist in connecting events and identifying attacks to improve any security defense system. These difficulties increase the workload on cybersecurity teams, which need more precise and reliable security solutions.

In this thesis, new hybrid security systems called integrated SIEM systems have been proposed based on the integration between open-source SIEM with suggested hybrid intelligent models to improve attack detection accuracy. Implementing the proposed systems includes three phases. The first phase represents building a virtual network environment that simulate a real environment using VMware Workstation. Then, Wazuh, an open-source SIEM platform, is used as a central server to monitor network devices, collect and analyze logs to detect threats, and perform file integrity monitoring (FIM) to detect unauthorized file changes. After that, Wazuh is integrated with IDS tool, namely Suricata, to enhance its ability to detect threats. In addition, Wazuh's rules have been modified to improve Wazuh's ability in analyzing logs and classifying the detected threats. Then, many attacks (DoS, SQL injection, brute force, network scan, and file modification) are performed against network devices to evaluate Wazuh's ability to detect threats based on the modified rules. In the second phase, a new dataset is constructed from historical data collected by Wazuh and preprocessed for both binary and multiclass classification. Subsequently,

hybrid intelligent models based on feature selection using the mutual information (MI) method, with or without feature reduction through Principal component analysis (PCA) or independent component analysis (ICA), and training and classification using long short term memory (LSTM), namely (MI-LSTM, MI-PCA-LSTM, and MI-ICA-LSTM) with various configuration, are suggested and trained on the new dataset and validated for their effectiveness in accurately classifying attacks and reducing false alerts for the purpose of deploying them in a real-world environment. In the third phase, the proposed integrated SIEM systems based on combining Wazuh SIEM with hybrid intelligent models are employed for centralized monitoring of network devices to improve the accuracy and efficiency of attack detection and classification while reducing false alerts in real time. Therefore, these integrated systems are deployed in a virtual network environment to evaluate their performance, where real attacks are conducted against network devices. Consequently, new data coming into the system from network devices in real-time is classified within an operational environment.

Experimental results show that Wazuh performs well in detecting attacks based on the modified rules with a total accuracy of 84.4 percent over (200) running times of (5) types of attacks. Also, the performance of the suggested hybrid intelligent models has been evaluated based on the new dataset collected by Wazuh. The experimental outcomes show that the MI-PCA-LSTM model with a robust scaler and a batch size of 32 slightly outperformed the rest of the models for multi-class classification with an accuracy of 99.68 and a loss of 0.016. In addition, the performance of the suggested hybrid intelligent models is evaluated using the public dataset, namely UNSW-NB15. The experimental outcomes show that the suggested models (MI-LSTM, MI-PCA-LSTM, MI-ICA-LSTM) achieved better performance compared to other methods in the literature in terms of accuracy by 99.64, 99.61, and 99.60 for multiclass classification. In addition, the experimental results of the proposed integrated SIEM systems show that the performance is efficient and has the ability to detect and classify attacks in a real environment with an average detection accuracy of 97.8%. for the system based on the MI-PCA-LSTM model with a robust scaler and a batch size of 32, which slightly outperforms the rest of the models in multi-classification.

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

تصنيف تهديدات الشبكات بناءً على منصة SIEM المعدلة وتقنيات التعلم الالي

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

زياد صفاء يونس طليع

بإشراف
أ.د. مفاز محسن خليل حسن

١٤٤٧هـ ٢٠٢٥م

المستخلص

ادى التطور السريع للشبكات والاعتماد الكبير عليها من قِبل الدول والمنظمات إلى زيادة تعرضها للهجمات السيبرانية. تتوفر العديد من أنظمة الأمان التقليدية لغرض حماية موارد الشبكة من الهجمات السيبرانية، ولكنها تعمل بشكل منفصل، مما يؤدي إلى تقييم مجزأ وغير دقيق للوضع الأمني لموارد الشبكة. وبالتالي، تبرز الحاجة الى حلول دفاعية أكثر فعالية. تُعد إدارة المعلومات الأمنية والأحداث، المعروفة باسم SIEM، إحدى الحلول الأمنية الرائدة التي تستخدم لزيادة مستوى أمن المعلومات وحماية البيانات من خلال إدارة أجهزة الشبكة بصورة مركزية. تعتمد المؤسسات الصغيرة والمتوسطة على حلول SIEM مفتوحة المصدر بسبب قلة التمويل والموارد. ومع ذلك، تواجه حلول الـ SIEM مفتوحة المصدر بعض الصعوبات، مثل تعقيد تحليل السجلات واحتمالية التحميل الزائد للتنبيهات مما يقلل من معدل الدقة ويزيد من عدد التنبيهات الكاذبة. بالإضافة إلى ذلك، لا توفر حلول SIEM من المنصات مفتوحة المصدر قدرات التعلم الآلي كوظائف اساسية مضمنة داخلها، حيث يمكن لتقنيات التعلم الآلي المساعدة في ربط الأحداث وتحديد الهجمات لتحسين أي نظام دفاع أمني. تزيد هذه الصعوبات من عبء العمل على فرق الأمن السيبراني، والتي تحتاج إلى حلول أمنية أكثر دقة وموثوقية.

في هذه الاطروحة، تم اقتراح انظمة أمنية هجينة جديدة تسمى انظمة SIEM المتكامل يعتمد على الدمج ما بين نظام الـ SIEM مفتوح المصدر والنماذج الذكية الهجينة المقترحة لغرض تحسين دقة الكشف عن الهجمات السيبرانية. يتضمن تنفيذ الانظمة المقترحة ثلاث مراحل. تتضمن المرحلة الأولى بناء شبكة افتراضية لغرض محاكاة الشبكة الحقيقية باستخدام برنامج VMware Workstation. بعد ذلك، يتم استخدام Wazuh، وهو نظام SIEM مفتوح المصدر، كخادم مركزي لمراقبة أجهزة الشبكة وجمع وتحليل السجلات للكشف عن التهديدات ومراقبة سلامة الملفات (FIM) للكشف عن التعديلات غير المصرح بها في الملفات. بعد ذلك، يتم دمج Wazuh مع أداة IDS وهي Suricata لتعزيز قدرته على اكتشاف التهديدات. بالإضافة إلى ذلك، تم تعديل قواعد Wazuh لتحسين قدرة Wazuh على تحليل السجلات وتصنيف التهديدات المكتشفة. بعد ذلك، يتم تنفيذ العديد من الهجمات (هجوم الحرمان من الخدمة، وهجوم حقن SQL، وهجوم القوة الغاشمة، وهجمات مسح الشبكة وعمليات التعديل على الملفات) ضد أجهزة الشبكة لتقييم قدرة Wazuh على اكتشاف التهديدات بناءً على القواعد المعدلة. في المرحلة الثانية، يتم انشاء مجموعة بيانات جديدة من البيانات

التاريخية التي جمعها من قبل Wazuh ومعالجتها مسبقًا للتصنيف الثنائي ومتعدد الفئات. بعد ذلك، تم اقتراح نماذج ذكية تعتمد على طريقة اختيار الميزات باستخدام طريقة المعلومات المتبادلة (MI)، مع أو بدون طرق تقليل الميزات من خلال استخدام طرق تحليل المكونات الرئيسية (PCA) أو تحليل المكونات المستقلة (ICA)، والتدريب والتصنيف باستخدام طريقة الذاكرة طويلة قصيرة المدى (LSTM)، والتي تسمى (MI–LSTM، MI–PCA–LSTM، MI–ICA–LSTM) مع تكوينات متنوعة، ويتم تدريبها على مجموعة البيانات الجديدة، والتحقق من كفاءتها في تصنيف الهجمات بدقة وتقليل التنبيهات الكاذبة، لغرض نشرها في بيئة حقيقية. في المرحلة الثالثة، يتم استخدام أنظمة الـ SIEM المتكاملة المقترحة بناءً على الدمج ما بين Wazuh مع النماذج الذكية الهجينة لغرض المراقبة المركزية لأجهزة الشبكة لتحسين دقة وكفاءة الكشف عن الهجمات وتصنيفها مع تقليل الإنذارات الكاذبة. لذلك، يتم نشر هذه الانظمة في بيئة شبكة افتراضية لتقييم أدائها من خلال شن هجمات حقيقية ضد اجهزة الشبكة. وبالتالي، يتم تصنيف البيانات الجديدة الواردة إلى النظام من أجهزة الشبكة في الوقت الحقيقي ضمن بيئة تشغيلية.

تظهر النتائج التجريبية أن Wazuh يعمل بشكل جيد في اكتشاف الهجمات بالاعتماد على القواعد المعدلة بدقة إجمالية تبلغ 84.4 بالمائة من خلال تنفيذها (200) مرة لـ (5) أنواع من الهجمات. كما تم تقييم أداء النماذج الذكية الهجينة المقترحة بناءً على مجموعة البيانات الجديدة التي تم جمعها بواسطة Wazuh. أظهرت النتائج التجريبية أن أداء نموذج MI–PCA–LSTM باستخدام أداة (robust scaler) و(batch size=32) تفوق قليلاً على باقي النماذج الذكية للتصنيف المتعدد بدقة 99.68 وخسارة 0.016. بالإضافة إلى ذلك، تم تقييم أداء النماذج الذكية الهجينة المقترحة باستخدام مجموعة البيانات العامة، وهي UNSW–NB15. وقد أظهرت النتائج التجريبية أن النماذج المقترحة (MI–LSTM، MI–PCA–LSTM، MI–ICA–LSTM) حققت أداء أفضل بالمقارنة مع الطرق السابقة من حيث الدقة بنسبة 99.64 و99.61 و99.60 للتصنيف متعدد الفئات. بالإضافة إلى ذلك، أظهرت النتائج التجريبية لأنظمة SIEM المتكاملة المقترحة أن الأداء فعال ولديه القدرة على اكتشاف وتصنيف الهجمات في بيئة حقيقية بمتوسط دقة كشف 97.8٪ للنظام القائم على نموذج MI–PCA–LSTM باستخدام أداة (robust scaler) و(batch size=32)، والذي يتفوق قليلاً على بقية النماذج في التصنيف المتعدد.