الجامعة السورية الخاصة
كلية هندسة المعلوماتية
اختصاص أمن النظم والشبكات الحاسوبية

كشف الشذوذ المعتمد على تحليل حركة البيانات في شبكات إنترنت الأشياء MQTT

تقرير مشروع تخرج 1

إعداد :
سلام عبد القادر عبد القادر
إشراف :
م. محمد يامن حلاق

كانون الثاني / 2026F

الملخّص(Abstract)

مع التوسع السريع في أنظمة إنترنت الأشياء(IoT) ، أصبح تأمين بروتوكولات الاتصال من التحديات الأساسية. ويُعد بروتوكول MQTT من أكثر البروتوكولات استخدامًا في هذه البيئات نظرًا لخفته وكفاءته، مما يجعله هدفًا شائعًا للهجمات السيبرانية، ولا سيما هجمات الإغراق .(Flooding Attacks) تعتمد أنظمة كشف التسلل التقليدية غالبًا على التواقيع أو تقنيات التعلم الخاضع للإشراف، وهي أساليب غير ملائمة لبيئات إنترنت الأشياء بسبب محدودية الموارد وصعوبة توفر بيانات موسومة (Labeled Data) بدقة، أي بيانات يتم فيها تحديد الفئة الصحيحة لكل عينة مسبقًا، مثل تحديد ما إذا كانت حركة المرور طبيعية أو هجومية.

يقدّم هذا المشروع نظامًا لكشف السلوك الشاذ في حركة المرور الشبكية لبروتوكول MQTT اعتمادًا على خوارزمية Isolation Forestغير الخاضعة للإشراف. يعتمد النهج المقترح على نمذجة السلوك الطبيعي لحركةMQTT ، واعتبار أي انحراف واضح عن هذا السلوك مؤشرًا على نشاط غير طبيعي. تم بناء بيئة تجريبية واقعية لتوليد حركة مرور طبيعية وخبيثة، بما في ذلك هجمات الإغراق، ثم التقاط حركة المرور الشبكية ومعالجتها واستخراج خصائص إحصائية على مستوى التدفق (Flow-based Features) دون تحليل محتوى الحزم، مما يحافظ على خصوصية البيانات.

تم تدريب نموذج Isolation Forest باستخدام بيانات تمثل السلوك الطبيعي فقط، ثم حساب درجات الشذوذ لكل تدفق شبكي. ولتحسين أداء الكشف، تم اعتماد آلية عتبة متغيرة تُحسب دوريًا اعتمادًا على توزيع درجات السلوك الطبيعي حصريًا، مما يسمح للنظام بالتكيف مع تغير أنماط حركة المرور مع منع تأثر العتبة بالسلوك الهجومي. أظهرت النتائج التجريبية أن النظام المقترح حقق معدل كشف مرتفع لهجمات الإغراق بلغ **98.31%**، مع معدل إنذارات خاطئة يقارب **5%**، مما يدل على توازن فعّال بين الحساسية والدقة.

تؤكد نتائج هذا البحث أن دمج تقنيات كشف الشذوذ غير الخاضعة للإشراف مع آلية عتبة متغيرة يمثل حلًا خفيف الوزن وفعالًا لتعزيز أمن شبكات إنترنت الأشياء المعتمدة على بروتوكولMQTT ، وقابلًا للتطبيق في البيئات الواقعية.

الكلمات المفتاحية:

إنترنت الأشياء، بروتوكولMQTT ، كشف السلوك الشاذ، Isolation Forest، العتبة المتغيرة، هجمات الإغراق.

Abstract

With the rapid growth of Internet of Things (IoT) systems, ensuring the security of communication protocols has become a critical challenge. MQTT is one of the most widely used lightweight messaging protocols in IoT environments, which makes it an attractive target for various cyber-attacks, particularly flooding attacks. Traditional intrusion detection systems often rely on signature-based or supervised learning approaches, which are not well suited for IoT environments due to resource constraints and the lack of labeled attack data.

This project proposes an anomaly-based intrusion detection system for MQTT network traffic using the Isolation Forest algorithm, an unsupervised machine learning technique. The proposed approach models normal MQTT traffic behavior and detects deviations from this behavior as potential anomalies. A realistic experimental environment was built to generate both legitimate and malicious MQTT traffic, including flooding attacks. Network traffic was captured, processed, and transformed into flow-based statistical features without inspecting packet payloads, preserving data privacy.

The Isolation Forest model was trained exclusively on normal traffic data, and anomaly scores were produced for each traffic flow. To improve detection performance, a dynamic thresholding mechanism derived solely from the distribution of normal traffic was applied, allowing the system to adapt to changes in network behavior while preventing threshold drift caused by attacks. Experimental results demonstrate that the proposed system achieves a high attack detection rate of **98.31%** with a false positive rate of approximately **5%**, indicating an effective balance between sensitivity and reliability.

The results confirm that combining unsupervised anomaly detection with an adaptive thresholding strategy provides an efficient and lightweight solution for detecting abnormal behavior in MQTT-based IoT networks, making the proposed system suitable for practical deployment in real-world IoT environments.

Keywords

Internet of Things (IoT), MQTT Protocol, Anomaly Detection, Intrusion Detection System (IDS), Network Traffic Analysis, Isolation Forest, Flooding Attack, Unsupervised Machine Learning, IoT Security

جدول المحتويات

فهرس الأشكال

فهرس الجداول

جدول المصطلحات

المصطلح	الاختصار	الوصف
إنترنت الأشياء	IoT	شبكة من الأجهزة الذكية المتصلة بالإنترنت، قادرة على جمع البيانات وتبادلها دون تدخل بشري مباشر.
بروتوكولMQTT	MQTT	بروتوكول مراسلة خفيف الوزن يعتمد على نموذج النشر والاشتراك، صُمم خصيصا لبيئات إنترنت الأشياء ذات الموارد المحدودة.

وسيط MQTT	Broker	الخادم المسؤول عن إدارة الاتصالات بين الناشرين (Publishers) والمشتركين (Subscribers)في بروتوكول MQTT.
النشر والاشتراك	Pub/Sub	نموذج اتصال يُرسل فيه الناشر الرسائل إلى موضوع معين، ويستقبلها جميع المشتركين في هذا الموضوع.
نظام كشف التسلل	IDS	نظام أمني يهدف إلى مراقبة حركة الشبكة أو سلوك النظام لاكتشاف الأنشطة غير الطبيعية أو غير المصرح بها.
كشف الشذوذ	Anomaly Detection	أسلوب يعتمد على اكتشاف الانحرافات عن السلوك الطبيعي بدلا من الاعتماد على أنماط هجوم معروفة مسبقا.
خوارزمية Isolation Forest	IF	خوارزمية تعلم آلي غير خاضعة للإشراف تُستخدم لعزل العينات الشاذة اعتمادا على عدد التقسيمات العشوائية.
تعلم غير خاضع للإشراف	Unsupervised Learning	نوع من التعلم الآلي لا يعتمد على بيانات موسومة، بل يستخرج الأنماط مباشرة من البيانات.
حركة مرور شبكية	Network Traffic	البيانات المتبادلة عبر الشبكة على شكل حزم أثناء الاتصال بين الأجهزة.
هجوم الإغراق	Flooding Attack	نوع من هجمات حجب الخدمة يعتمد على إرسال عدد كبير من الرسائل خلال فترة زمنية قصيرة لإرباك النظام المستهدف.
درجة الشذوذ	Anomaly Score	قيمة عددية ناتجة عن نموذج كشف الشذوذ تعبّر عن مدى انحراف العينة عن السلوك الطبيعي.
العتبة	Threshold	قيمة فاصلة تُستخدم لتحويل درجة الشذوذ إلى قرار تصنيفي (طبيعي أو شاذ).
العتبة المتغيرة	Adaptive Threshold	عتبة يتم تحديثها أو تعديلها بناء على سلوك البيانات بدلا من كونها قيمة ثابتة.
المعالجة المسبقة	Data Preprocessing	مجموعة من الخطوات التي تُجرى على البيانات قبل التدريب، مثل التنظيف والتطبيع واختيار الخصائص.
استخراج الخصائص	Feature Extraction	عملية تحويل البيانات الخام إلى خصائص عددية تمثل السلوك الشبكي بشكل قابل للتحليل.
مستوى التدفق	Flow-based Analysis	أسلوب تحليل يعتمد على خصائص التدفق الشبكي بدلا من تحليل كل حزمة على حدة.
مصفوفة الالتباس	Confusion Matrix	أداة إحصائية تُستخدم لتقييم أداء نموذج التصنيف من خلال مقارنة النتائج المتوقعة بالقيم الحقيقية.
معدل الكشف	Detection Rate / Recall	مقياس يعبّر عن نسبة الهجمات التي تم اكتشافها بشكل صحيح.
الإنذارات الخاطئة	False Positives	حالات يتم فيها تصنيف حركة طبيعية على أنها هجوم.
الزمن الحقيقي	Real-time	معالجة البيانات واتخاذ القرار أثناء تدفق البيانات دون تأخير زمني ملحوظ.

الفصل الأول

المقدمة العامة و إشكالية البحث

1.1 المقدمة

شهدت تقنيات إنترنت الأشياء (Internet of Things – IoT) انتشارا واسعا خلال السنوات الأخيرة، حيث أصبحت جزءا أساسيا من العديد من الأنظمة الحديثة مثل المنازل الذكية، المدن الذكية، الأنظمة الصناعية، والرعاية الصحية. تعتمد هذه الأنظمة على عدد كبير من الأجهزة الذكية المتصلة بالشبكة، والتي تقوم بتبادل البيانات بشكل مستمر بهدف المراقبة، التحكم، واتخاذ القرار. هذا الانتشار الكبير أدى إلى زيادة حجم وتعقيد حركة المرور الشبكية، وفتح المجال أمام تحديات أمنية جديدة تتطلب حلولا فعّالة ومناسبة لطبيعة بيئات إنترنت الأشياء.

تُعد الجوانب الأمنية من أبرز التحديات التي تواجه أنظمة إنترنت الأشياء، وذلك بسبب محدودية الموارد الحاسوبية للأجهزة من حيث القدرة على المعالجة، الذاكرة، واستهلاك الطاقة. نتيجة لذلك، غالبا ما يتم الاعتماد على بروتوكولات اتصال خفيفة الوزن تركز على الكفاءة وسرعة نقل البيانات أكثر من تركيزها على آليات الحماية المتقدمة. هذا الأمر يجعل أنظمة IoT عرضة لهجمات سيبرانية متعددة، خاصة الهجمات التي تعتمد على استغلال السلوك الطبيعي للشبكة دون الحاجة إلى اختراق مباشر أو استغلال ثغرات برمجية.

يُعتبر بروتوكول Message Queuing Telemetry Transport (MQTT) من أكثر بروتوكولات الاتصال استخداما في بيئات إنترنت الأشياء، نظرا لبساطته واعتماده على نموذج النشر والاشتراك(Publish/Subscribe) ، بالإضافة إلى كفاءته العالية في البيئات ذات الموارد المحدودة. يعمل بروتوكول MQTT فوق بروتوكول النقلTCP ، ويُستخدم على نطاق واسع في التطبيقات التي تتطلب اتصالا موثوقا وزمن استجابة منخفض. وعلى الرغم من هذه المزايا، فإن طبيعة البروتوكول الخفيفة وعدم فرض آليات أمان صارمة بشكل افتراضي يجعله هدفا شائعا للهجمات السلوكية مثل هجمات الإغراق (Flooding Attacks)، والتي قد تؤدي إلى استنزاف موارد الخادم أو تعطيل الخدمة.

في ظل هذه التحديات، برزت تقنيات كشف الشذوذ (Anomaly Detection) كحل واعد لتعزيز أمن أنظمة إنترنت الأشياء، خاصة في الحالات التي يصعب فيها الاعتماد على قواعد ثابتة أو تواقيع هجومية معروفة. تعتمد هذه التقنيات على تحليل سلوك حركة المرور الشبكية واكتشاف الأنماط غير الطبيعية التي قد تشير إلى وجود نشاط خبيث. وتُعد الأساليب غير الخاضعة للإشراف (Unsupervised Learning) مناسبة بشكل خاص لهذا النوع من البيئات، نظرا لعدم توفر بيانات موسومة بشكل كامل وصعوبة حصر جميع أنواع الهجمات المحتملة.

يهدف هذا المشروع إلى تصميم وتنفيذ نظام لكشف السلوك الشاذ في حركة بروتوكول MQTT ضمن بيئة إنترنت الأشياء، اعتمادا على تحليل الخصائص السلوكية لحركة المرور الشبكية. يرتكز الحل المقترح على استخدام خوارزمية Isolation Forestللكشف عن الأنماط غير الطبيعية، وذلك من خلال استخراج مجموعة خصائص شبكية مبسطة وقابلة للاستخراج من الحركة المرور الشبكي المولّد فعليا. يركّز المشروع بشكل خاص على الكشف عن هجمات الإغراق، مع مراعاة قابلية التطبيق العملي وإمكانية التوسع مستقبلا نحو أنظمة الكشف في الزمن الحقيقي.

1.2 المشكلة العلمية

على الرغم من الانتشار الواسع لتقنيات إنترنت الأشياء واعتمادها المتزايد في الأنظمة الحيوية، لا تزال مسألة تأمين حركة المرور الشبكية في هذه البيئات تمثل تحديا علميا وعمليا كبيرا. تعود هذه الصعوبة بشكل أساسي إلى الطبيعة غير المتجانسة لأجهزة إنترنت الأشياء، ومحدودية مواردها الحاسوبية، إضافة إلى اعتمادها على بروتوكولات اتصال خفيفة الوزن صُممت لتحقيق الكفاءة التشغيلية أكثر من تحقيق مستويات أمان مرتفعة. ونتيجة لذلك، تصبح هذه البيئات عرضة لهجمات سيبرانية يصعب اكتشافها باستخدام تقنيات الحماية التقليدية المعتمدة على التواقيع أو القواعد الثابتة.

يُعد بروتوكول MQTT مثالا واضحا على هذا التحدي، إذ يعمل كبروتوكول تطبيق فوق بروتوكول TCP ويُستخدم لنقل البيانات بين عدد كبير من الأجهزة ضمن نموذج النشر والاشتراك. ورغم أن هذا التصميم يحقق كفاءة عالية في نقل البيانات، إلا أنه يتيح للمهاجمين استغلال السلوك الطبيعي للبروتوكول لتنفيذ هجمات سلوكية، مثل هجمات الإغراق، دون الحاجة إلى كسر آليات التشفير أو استغلال ثغرات برمجية مباشرة. هذا النوع من الهجمات يعتمد على توليد حركة مرور كثيفة أو غير طبيعية تؤدي إلى استنزاف موارد الخادم أو تعطيل الخدمة، مما يجعل اكتشافه أكثر تعقيدا.

تعتمد العديد من حلول كشف التسلل التقليدية على بيانات موسومة مسبقا أو على تواقيع معروفة للهجمات، وهو ما لا يتناسب مع بيئات إنترنت الأشياء التي تتغير فيها أنماط السلوك بشكل مستمر. بالإضافة إلى ذلك، فإن عملية وسم البيانات يدويا تُعد مكلفة وصعبة التطبيق في الشبكات الواسعة، كما أن الهجمات الجديدة أو المتطورة قد لا تكون ممثلة ضمن مجموعات البيانات المستخدمة في التدريب. لذلك، فإن الاعتماد على أساليب خاضعة للإشراف فقط قد يؤدي إلى ضعف القدرة على التعميم وانخفاض فعالية الكشف في السيناريوهات الواقعية.

من جهة أخرى، فإن استخدام مجموعات خصائص معقدة أو أدوات ثقيلة لاستخراج الخصائص، مثل الاعتماد الكامل على خصائص التدفق واسعة النطاق، قد يحدّ من إمكانية تطبيق أنظمة الكشف في الزمن الحقيقي أو في البيئات ذات الموارد المحدودة. هذا يبرز الحاجة إلى تطوير حلول تعتمد على خصائص سلوكية مبسطة يمكن استخراجها مباشرة من حركة المرور الشبكية، مع الحفاظ على قدرتها على التمييز بين السلوك الطبيعي والسلوك الشاذ.

بناء على ما سبق، تتمثل المشكلة العلمية التي يعالجها هذا المشروع في كيفية تصميم نظام فعّال لكشف السلوك الشاذ في حركة بروتوكول MQTT ضمن بيئات إنترنت الأشياء، باستخدام تقنيات تعلم آلي غير خاضعة للإشراف، وبالاعتماد على مجموعة خصائص شبكية مبسطة وقابلة للاستخراج من الحركة المرور الشبكي المولّد فعليا. يسعى المشروع إلى تحقيق توازن بين دقة الكشف، بساطة التنفيذ، وقابلية التطبيق العملي، خاصة في سياق الكشف عن هجمات الإغراق التي تستهدف بروتوكول.MQTT

1.3 أهداف المشروع

يهدف هذا المشروع إلى معالجة مشكلة كشف السلوك الشاذ في شبكات إنترنت الأشياء التي تعتمد على بروتوكول MQTT، من خلال تحقيق مجموعة من الأهداف العلمية والتطبيقية المحددة كما يلي:

1. **تحليل السلوك الشبكي لبروتوكولMQTT**

دراسة طبيعة حركة المرور الخاصة ببروتوكول MQTT ضمن بيئات إنترنت الأشياء، وتحديد الخصائص الشبكية التي تعكس الفرق بين السلوك الطبيعي والسلوك الشاذ، خصوصا في حالات هجمات الإغراق.

2. **تصميم نظام كشف شذوذ غيرخاضع للإشراف**

تطوير نموذج كشف يعتمد على خوارزميات تعلم آلي غير خاضعة للإشراف، بما يتيح اكتشاف الهجمات دون الحاجة إلى بيانات موسومة مسبقا، وبالتالي زيادة قابلية التعميم على هجمات غير معروفة سابقا.

3. **اختياروا‌ستخراج مجموعة خصائص شبكية مبسطة**

تحديد مجموعة من الخصائص القابلة للاستخراج من حركة المرور الشبكية بسهولة، مع التركيز على الخصائص السلوكية المرتبطة بمعدلات الإرسال، التكرار، والفواصل الزمنية بين الحزم، بهدف تقليل التعقيد الحسابي وتحسين قابلية التطبيق العملي.

4. **بناء نموذج كشف يعتمد على خوارزميةIsolation Forest**

تدريب نموذج كشف شذوذ باستخدام خوارزمية Isolation Forest نظرا لقدرتها على عزل السلوك غير الطبيعي بكفاءة في مجموعات البيانات الكبيرة، ومناسبتها للبيانات غير المتوازنة الشائعة في بيئات إنترنت الأشياء.

5. **تحديد عتبة كشف مناسبة للسلوك الشاذ**

دراسة آليات تحديد العتبة (Threshold) المستخدمة لتصنيف حركة المرور إلى طبيعية أو شاذة، وتحليل تأثير قيمة العتبة على أداء النموذج من حيث معدلات الكشف والإنذارات الخاطئة.

6. **تقييم أداء النظام باستخدام بيانات مختلفة**

اختبار النموذج باستخدام مجموعات بيانات تحتوي على نسبة مرتفعة من الحركة الطبيعية ونسبة محدودة من الهجمات، وتحليل النتائج باستخدام مقاييس الأداء المناسبة مثل مصفوفة الالتباس ومنحنيات التقييم.

7. **إبرازقابلية التوسع والتطبيق العملي للنظام المقترح**

تقييم مدى إمكانية تطبيق النظام في بيئات واقعية، سواء في الأنظمة شبه الفعلية أو كمرحلة تمهيدية لتطبيقات الكشف في الزمن الحقيقي، مع الأخذ بعين الاعتبار القيود الحاسوبية لأجهزة إنترنت الأشياء.

الهدف المقابل	المشكلة العلمية
تحليل السلوك الشبكي لبروتوكول MQTT وتحديد الخصائص التي تميّز الحركة الطبيعية عن الشاذة.	الزيادة الكبيرة في استخدام بروتوكول MQTT في بيئات إنترنت الأشياء، مقابل ضعف آليات الكشف الأمني المصممة خصيصا لهذا البروتوكول.

تصميم نظام كشف شذوذ غير خاضع للإشراف قادر على اكتشاف الهجمات دون الحاجة إلى بيانات موسومة مسبقا.	اعتماد معظم أنظمة كشف التسلل التقليدية على قواعد ثابتة أو بيانات موسومة، مما يقلل من فعاليتها أمام الهجمات الجديدة أو غير المعروفة.
اختيار واستخراج مجموعة خصائص شبكية مبسطة وقابلة للتطبيق العملي في بيئات محدودة الموارد.	صعوبة استخراج عدد كبير من الخصائص المعقدة في البيئات الفعلية لإنترنت الأشياء بسبب القيود الحاسوبية.
بناء نموذج كشف يعتمد على خوارزمية Isolation Forestالملائمة للبيانات غير المتوازنة.	عدم توازن مجموعات البيانات في شبكات إنترنت الأشياء، حيث تكون الحركة الطبيعية أكبر بكثير من الحركة الخبيثة.
تحديد عتبة كشف مناسبة وتحليل تأثيرها على معدلات الكشف والإنذارات الخاطئة.	صعوبة تحديد الحد الفاصل بين السلوك الطبيعي والسلوك الشاذ بدقة عالية.
تقييم أداء النظام باستخدام مجموعات بيانات يغلب عليها السلوك الطبيعي مع نسبة محدودة من الهجمات.	محدودية الدراسات التي تختبر نماذج كشف الشذوذ على بيانات تحتوي على نسبة واقعية من الهجمات.
إبراز قابلية التوسع والتطبيق العملي للنظام المقترح كمرحلة تمهيدية للكشف في الزمن الحقيقي.	الحاجة إلى أنظمة كشف يمكن تطويرها لاحقا للعمل في الزمن الحقيقي.

جدول1 هدف المشروع لكل مشكلة

1.4 حدود البحث ونطاقه

يركّز هذا البحث على كشف السلوك الشاذ في حركة المرور الشبكية الخاصة ببروتوكول MQTT ضمن بيئات إنترنت الأشياء، وذلك من خلال تحليل الخصائص السلوكية لحركة الحركة المرور الشبكي على مستوى الشبكة. يقتصر نطاق العمل على دراسة حركة بروتوكول MQTT التي تعمل فوق بروتوكولTCP ، دون التطرّق إلى بروتوكولات إنترنت الأشياء الأخرى مثل CoAP أو.AMQP

يعتمد النظام المقترح على استخراج مجموعة خصائص شبكية مبسطة من حركة المرور، مثل معدلات الإرسال، عدد الحزم، الأحجام، والفواصل الزمنية بين الحزم، دون تحليل الحمولة الداخلية لرسائل .MQTT يهدف هذا التوجه إلى تقليل التعقيد الحسابي والحفاظ على خصوصية البيانات، إلا أنه قد يحد من إمكانية اكتشاف بعض الهجمات التي تعتمد بشكل أساسي على محتوى الرسائل.

يقتصر البحث على دراسة نوع محدد من الهجمات السلوكية، وهو هجمات الإغراق (Flooding Attacks) التي تستهدف بروتوكولMQTT ، ولا يشمل أنواعا أخرى من الهجمات مثل هجمات انتحال الهوية أو التلاعب بالمحتوى. كما يعتمد التقييم على مجموعات بيانات تحتوي على نسبة مرتفعة من الحركة الطبيعية ونسبة محدودة من الحركة الهجومية، بما يعكس سيناريوهات واقعية لشبكات إنترنت الأشياء، لكنه قد لا يغطي جميع أنماط الهجمات الممكنة.

يعتمد النظام المقترح على خوارزمية Isolation Forest للكشف عن الشذوذ، دون إجراء مقارنة تفصيلية مع خوارزميات تعلم آلي أخرى. يهدف هذا القرار إلى التركيز على تحليل فعالية الخوارزمية المختارة بدل توسيع نطاق الدراسة بشكل قد يؤثر على عمق التحليل. كما أن النظام لا يعمل بشكل كامل في الزمن الحقيقي، وإنما يُعد خطوة تمهيدية يمكن تطويرها لاحقا لدعم الكشف الفوري.

الفصل الثاني
الدراسة المرجعية

2.1 أمن إنترنت الأشياء (IoT Security)

حظي موضوع أمن إنترنت الأشياء باهتمام متزايد في الأبحاث العلمية خلال السنوات الأخيرة، وذلك نتيجة الانتشار الواسع لأجهزة IoT واعتمادها في تطبيقات حساسة. تشير الدراسات إلى أن الطبيعة غير المتجانسة لهذه الأجهزة، إلى جانب محدودية مواردها من حيث القدرة الحاسوبية والطاقة، تجعل من الصعب تطبيق آليات الحماية التقليدية المستخدمة في الشبكات الكلاسيكية. كما أن العدد الكبير للأجهزة المتصلة يزيد من سطح الهجوم ويعقّد عملية المراقبة الأمنية.[17]

أوضحت العديد من الدراسات أن التحديات الأمنية في بيئات إنترنت الأشياء لا تقتصر على الهجمات التقليدية مثل التنصت أو انتحال الهوية، بل تشمل أيضا الهجمات السلوكية التي تستهدف استقرار الشبكة وتوافر الخدمة. ومن أبرز هذه الهجمات هجمات حجب الخدمة والإغراق، والتي تعتمد على توليد حركة مرور كثيفة تؤدي إلى استنزاف موارد الخوادم أو تعطيل الاتصال بين الأجهزة. وتُعد هذه الهجمات خطيرة بشكل خاص في بيئات IoT نظرا لاعتمادها على بروتوكولات خفيفة الوزن لا تتضمن آليات حماية متقدمة بشكل افتراضي.[19]

كما بيّنت دراسات أخرى أن الاعتماد على حلول أمنية مركزية أو ثقيلة قد لا يكون مناسبا لبيئات إنترنت الأشياء، نظرا للتأثير السلبي على الأداء وزمن الاستجابة. لذلك، اتجه الباحثون إلى اقتراح حلول تعتمد على تحليل حركة المرور الشبكية واكتشاف الأنماط غير الطبيعية، باعتبارها مقاربة فعّالة يمكن تطبيقها دون الحاجة إلى تعديل الأجهزة أو تحميلها أعباء حسابية إضافية.[20]

2.2 بروتوكولات الاتصال في إنترنت الأشياء مع التركيز على MQTT

تعتمد أنظمة إنترنت الأشياء على مجموعة من بروتوكولات الاتصال المصممة خصيصا لتلبية متطلبات البيئات محدودة الموارد، مثل انخفاض استهلاك الطاقة، تقليل حجم الرسائل، ودعم الاتصال غير المستقر. من أبرز هذه البروتوكولات : MQTT، وCoAP، وAMQP، حيث تختلف فيما بينها من حيث نموذج الاتصال، مستوى التعقيد، وآليات الاعتمادية. وقد أظهرت الدراسات أن اختيار البروتوكول المناسب يؤثر بشكل مباشر على أداء النظام وأمنه.[16]

يُعد بروتوكول MQTT من أكثر البروتوكولات استخداما في بيئات إنترنت الأشياء، نظرا لاعتماده على نموذج النشر والاشتراك(Publish/Subscribe) ، والذي يتيح فصل المرسل عن المستقبل من خلال وسيط مركزي يُعرف بالـ Broker. يساهم هذا النموذج في تحسين قابلية التوسع وتقليل الحمل على الأجهزة الطرفية، إذ لا يحتاج كل جهاز إلى معرفة عنوان الأجهزة الأخرى أو إدارتها بشكل مباشر. كما يعمل MQTT فوق بروتوكولTCP ، مما يوفر موثوقية في نقل البيانات، وهو عامل مهم في التطبيقات التي تتطلب ضمان تسليم الرسائل.[21]

رغم هذه المزايا، فإن بروتوكول MQTT لا يفرض بشكل افتراضي آليات أمان صارمة مثل التشفير أو المصادقة القوية، وإنما يترك ذلك لخيارات التكوين الخاصة بالمستخدم. هذا التصميم الخفيف يجعل البروتوكول عرضة للاستغلال في حال سوء الإعداد، خاصة في البيئات المفتوحة أو عند استخدام الإعدادات الافتراضية. وقد بينت أبحاث متعددة أن العديد من خوادم MQTT المتصلة بالإنترنت تعمل دون مصادقة أو باستخدام إعدادات ضعيفة، مما يزيد من مخاطر الهجمات السلوكية مثل الإغراق وحجب الخدمة.[24]

تتميز هجمات الإغراق على بروتوكول MQTT بأنها لا تعتمد على إرسال رسائل غير صالحة أو خرق البروتوكول، بل على إساءة استخدام السلوك الطبيعي للبروتوكول من خلال إرسال عدد كبير من رسائل النشر خلال فترة زمنية قصيرة. يؤدي هذا السلوك إلى زيادة مفاجئة في عدد الحزم، معدلات الإرسال، واستهلاك موارد الخادم، وهو ما ينعكس مباشرة

على حركة المرور الشبكية. لذلك، فإن تحليل الخصائص السلوكية لحركة المرور، مثل التكرار والزمن بين الرسائل، يُعد مؤشرا فعّالا لاكتشاف هذا النوع من الهجمات دون الحاجة إلى فحص محتوى الرسائل.[16]

2.3 أنظمة كشف التسلل في إنترنت الأشياء (Intrusion Detection Systems for IoT)

2.3.1 مفهوم أنظمة كشف التسلل في بيئات إنترنت الأشياء

تُعرَّف أنظمة كشف التسلل (Intrusion Detection Systems – IDS) بأنها أنظمة أمنية تهدف إلى مراقبة حركة الشبكة أو سلوك النظام من أجل اكتشاف الأنشطة غير المصرّح بها أو غير الطبيعية. في سياق إنترنت الأشياء، تختلف متطلبات أنظمة كشف التسلل عن الشبكات التقليدية بسبب القيود المفروضة على الأجهزة من حيث القدرة الحاسوبية والطاقة، إضافة إلى الطبيعة المتنوعة والديناميكية لحركة المرور.[21]

تشير الدراسات إلى أن تطبيق أنظمة IDS التقليدية في بيئات IoT يؤدي غالبا إلى ضعف في الأداء أو ارتفاع معدلات الإنذارات الخاطئة، وذلك بسبب اعتماد هذه الأنظمة على افتراضات لا تنطبق على بيئات إنترنت الأشياء، مثل ثبات أنماط الحركة المرور الشبكي أو تجانس الأجهزة.[22]

2.3.2 تصنيف أنظمة كشف التسلل في إنترنت الأشياء

يمكن تصنيف أنظمة كشف التسلل المستخدمة في بيئات إنترنت الأشياء إلى ثلاث فئات رئيسية:

1. **أنظمة كشف التسلل القائمة على التو اقيع:(Signature-based IDS)**
   تعتمد على مقارنة حركة المرور مع أنماط هجوم معروفة مسبقا. تتميز بدقتها العالية في كشف الهجمات المعروفة، إلا أنها غير قادرة على اكتشاف الهجمات الجديدة أو المعدّلة، مما يحد من فعاليتها في بيئات IoT المتغيرة.
2. **أنظمة كشف التسلل القائمة على الشذوذ:(Anomaly-based IDS)**
   تعتمد على بناء نموذج للسلوك الطبيعي للشبكة، ثم تصنيف أي انحراف عنه كسلوك مشبوه. تُعد هذه الأنظمة أكثر ملاءمة لبيئات إنترنت الأشياء نظرا لقدرتها على كشف الهجمات غير المعروفة مسبقا.[23]
3. **الأنظمة الهجينة:(Hybrid IDS)**
   تجمع بين الطريقتين السابقتين بهدف تحسين الدقة وتقليل الإنذارات الخاطئة، إلا أنها غالبا ما تتطلب موارد حسابية أكبر.

2.3.3 استخدام التعلم الآلي في كشف التسلل فيIoT

أدى التطور في تقنيات التعلم الآلي إلى اعتمادها بشكل متزايد في بناء أنظمة كشف التسلل الخاصة بإنترنت الأشياء. تتيح هذه التقنيات تحليل كميات كبيرة من بيانات الحركة المرور الشبكي واستخلاص أنماط سلوكية يصعب اكتشافها باستخدام الأساليب التقليدية. ومع ذلك، تعتمد الخوارزميات الخاضعة للإشراف على توفر بيانات موسومة بدقة، وهو أمر غير عملي في معظم سيناريوهات IoT الواقعية.[24]

بناء على ذلك، اتجهت العديد من الأبحاث إلى استخدام خوارزميات غير خاضعة للإشراف، والتي لا تتطلب بيانات مصنفة مسبقا، وتعتمد على اكتشاف الانحرافات السلوكية في البيانات. هذا التوجه يُعد مناسبا بشكل خاص لبيئات

إنترنت الأشياء، حيث يشكّل الحركة المرور الشبكي الطبيعي النسبة الأكبر من البيانات، بينما تكون الهجمات قليلة نسبيا.

2.3.4 كشف التسلل المعتمد على حركة المرور في بروتوكول MQTT

في حالة بروتوكول MQTT ، ركزت الدراسات الحديثة على تحليل حركة المرور الشبكية بدلا من تحليل محتوى الرسائل. يعود ذلك إلى أن العديد من الهجمات، مثل هجمات الإغراق(Flooding Attacks) ، تؤدي إلى تغيّرات واضحة في الخصائص الشبكية، مثل زيادة عدد الحزم، ارتفاع معدل الإرسال، وانخفاض الفواصل الزمنية بين الرسائل.[25]
يُعد هذا الأسلوب مناسبا لبيئات إنترنت الأشياء لأنه:

- لا يتطلب فك تشفير البيانات.
- يقلل من التعقيد الحسابي.
- يحافظ على خصوصية محتوى الرسائل.

2.3.5 ملاءمة خوارزمية Isolation Forest لبيئات IoT

تُعد خوارزمية Isolation Forest من خوارزميات كشف الشذوذ غير الخاضعة للإشراف، وتعتمد على مبدأ عزل العينات غير الطبيعية باستخدام أشجار عشوائية. تتميز هذه الخوارزمية بانخفاض تعقيدها الحسابي وقدرتها على التعامل مع البيانات غير المتوازنة، مما يجعلها مناسبة لبيئات إنترنت الأشياء.[26]
بالمقارنة مع خوارزميات أخرى، لا تتطلب Isolation Forest نمذجة دقيقة للتوزيع الإحصائي للبيانات، بل تعتمد على حقيقة أن العينات الشاذة يمكن عزلها بعدد أقل من التقسيمات. هذا المفهوم يجعلها فعالة في اكتشاف الهجمات السلوكية التي تظهر على شكل انحرافات واضحة في حركة المرور، مثل هجمات الإغراق في بروتوكول MQTT.

2.4 مقارنة وتحليل الدراسات السابق

رقم	البحث	المؤلف	النتيجة الرئيسية	العلاقة بالمشروع
24	Anomaly-based detection of MQTT flooding attacks	Alsaedi & Taha (2020)	كشف فعال لهجمات MQTT باستخدام خصائص الشبكة	يدعم اختيار هجوم الإغراق
2	Landscape of IoT Security	Schiller et al. (2022)	عرض شامل لتحديات أمن IoT	يبرر الحاجة لكشف الشذوذ
3	Detecting Anomalous Network Traffic in IoT	Hoang & Nguyen (2018)	نجاح التعلم غير الخاضع للإشراف	يدعم Isolation Forest
7	Analysis of network traffic features	Iglesias & Zseby (2015)	أهمية الخصائص الزمنية	يدعم اختيار Features
21	Survey of IDS in IoT	Zarpelão et al. (2017)	قصور الأنظمة التقليدية	يبرر النهج المقترح

13	AI-based Anomaly Detection in IoT	De Medeiros et al. (2023)	فعالية الذكاء الاصطناعي	يدعم المنهجية
8	Survey of IoT security	Al-Garadi et al. (2020)	التعلم الآلي فعال للأمن	خلفية علمية
12	Graph-based anomaly detection	Wu et al. (2022)	مقارنة طرق حديثة	يوضح اختيارIF
25	Isolation Forest	Liu et al. (2008)	أساس الخوارزمية	المرجع الأساسي
9	IoT protocols survey	Elhadi et al. (2018)	خصائصMQTT	يبرر اختيار البروتوكول
26	ProfilIoT: A Machine Learning Approach for IoT Device Identification	Meidan et al. (2018)	توليد حركة مرور IoTحقيقية وبناء Datasetمخصّصة باستخدامWireshark	يبرر توليد الترافيك وبناء Datasetيدويًا في هذا البحث

جدول 2 مقارنة الدراسات السابقة

من خلال استعراض الدراسات المرجعية الموضحة في الجدول (2)، يمكن ملاحظة أن معظم الأبحاث ركزت على كشف الهجمات في بيئات إنترنت الأشياء بشكل عام، أو اعتمدت على بيانات موسومة وخوارزميات خاضعة للإشراف، مما يحد من قدرتها على التكيف مع الهجمات الجديدة. كما أن عددًا من الدراسات لم يخصص التحليل لبروتوكول MQTT أو لم يعالج مشكلة عدم توازن البيانات.

في المقابل، يتميّز هذا البحث باعتماده على تحليل حركة المرور الشبكية لبروتوكول MQTT تحديدًا، واستخدام خوارزمية غير خاضعة للإشراف مدرّبة على السلوك الطبيعي فقط، مع اعتماد عتبة مستخرجة من توزيع البيانات نفسها، مما يجعله أكثر ملاءمة للبيئات الواقعية.

الفصل الثالث

حركة المرور الشبكي والسلوك الشاذ

3.1 تعريف حركة المرور الشبكي (Network Traffic)

تُعرف حركة المرور الشبكي (Network Traffic) بأنها كمية البيانات المتدفقة عبر شبكة اتصالات في فترة زمنية معينة. وهي تمثل مجموع حزم البيانات (Data Packets) المرسلة والمستقبلة بين الأجهزة المتصلة [3]. بالنسبة لمشغلي الشبكات، فإن فهم وتحليل حركة المرور الشبكي أمر بالغ الأهمية لضمان جودة الخدمة (Quality of Service - QoS)، وتحديد الأداء الطبيعي للشبكة، واكتشاف أي سلوك شاذ قد يشير إلى هجوم أو خلل فني [3].

في سياق شبكات إنترنت الأشياء (IoT)، تكتسب حركة المرور الشبكي خصائص فريدة تميزها عن الشبكات التقليدية، حيث تتسم بـ:

1.النمطية (Periodicity): العديد من أجهزة إنترنت الأشياء ترسل بياناتها بشكل دوري ومنتظم (مثل قراءات المستشعرات)، مما يسهل تحديد السلوك الطبيعي [10].

2.الحجم الصغير للحزم (Small Packet Size): نظرا لقيود النطاق الترددي والطاقة، غالبا ما تكون حزم البيانات في إنترنت الأشياء صغيرة الحجم [10].

3.التنوع (Diversity): تتنوع حركة المرور بين تدفقات البيانات من المستشعرات، ورسائل التحكم، وتحديثات البرامج، مما يتطلب أدوات تحليلية قادرة على التعامل مع هذا التباين [7].

أهمية تحليل حركة المرور: يُعد تحليل حركة المرور الشبكي هو الأساس الذي يُبنى عليه نظام اكتشاف الشذوذ (Anomaly Detection System). فمن خلال تحليل خصائص تدفقات البيانات، مثل معدل الحزم، وحجم التدفق، ومنافذ الاتصال، يمكن إنشاء نموذج للسلوك الطبيعي للشبكة [7]. وأي انحراف عن هذا النموذج الطبيعي يُصنف على أنه شذوذ محتمل [4].

3.2 أنواع الحزم وخصائصها في IoT

تتميز شبكات إنترنت الأشياء (IoT) بخصائص فريدة لحزم البيانات (Data Packets) المتداولة فيها، والتي تنبع بشكل أساسي من طبيعة الأجهزة المقيدة الموارد (Resource-constrained devices) والبروتوكولات الخفيفة الوزن المستخدمة [10]. إن فهم هذه الخصائص أمر أساسي لعمليات اكتشاف الشذوذ، حيث أن أي انحراف عن الخصائص المتوقعة للحزم قد يشير إلى سلوك غير طبيعي.

الخصائص الرئيسية لحزم IoT:

1.**حجم الحزمة الصغير (Small Packet Size):** تُصمم بروتوكولات إنترنت الأشياء مثل MQTT و CoAP لتقليل حجم الحزمة (Packet Overhead) إلى أدنى حد ممكن. هذا يقلل من استهلاك الطاقة وعرض النطاق الترددي، مما يجعلها مثالية لبيئات إنترنت الأشياء [9]. على سبيل المثال، في بروتوكول MQTT، تتكون حزمة التحكم (Control Packet) من رأس ثابت (Fixed Header)، ورأس متغير (Variable Header)، وحمولة (Payload)، وتُعد هذه المكونات صغيرة جدا مقارنة بحزم HTTP التقليدية [9].

2.**النمطية في الإرسال (Transmission Periodicity):** تُظهر حركة مرور إنترنت الأشياء أنماطا منتظمة ومحددة مسبقا في إرسال الحزم، خاصة في تطبيقات جمع البيانات من المستشعرات. يتم إرسال الحزم على فترات زمنية ثابتة أو شبه ثابتة، مما يخلق "بصمة" (Footprint) لحركة المرور الخاصة بالجهاز [10].

3.**نوع الحمولة (Payload Type):** تختلف حمولة الحزم بناء على وظيفة الجهاز والبروتوكول المستخدم. يمكن تصنيف أنواع الحزم بشكل عام إلى [10]

- حزم البيانات (Data Packets): تحمل القراءات الفعلية للمستشعرات (مثل درجة الحرارة، الضغط، الموقع).

- حزم التحكم (Control Packets): تستخدم لأغراض الإدارة، مثل الاتصال (CONNECT) أو قطع الاتصال (DISCONNECT) في MQTT، أو رسائل التأكيد (Acknowledgement) في CoAP.
- حزم الإعداد (Configuration Packets): تستخدم لتحديث إعدادات الجهاز أو البرامج الثابتة.

أهمية الخصائص في اكتشاف الشذوذ:

تُستخدم خصائص حزم البيانات كسمات (Features) رئيسية في نماذج اكتشاف الشذوذ. فمثلا، يمكن استخدام حجم الحزمة، أو الفاصل الزمني بين الحزم (Inter-arrival Time)، أو منافذ الاتصال (Ports) لتحديد ما إذا كانت حركة مرور جهاز معين طبيعية أم شاذة [7]، [10]. أي تغيير مفاجئ في هذه الخصائص، مثل زيادة غير مبررة في حجم الحزم أو معدل إرسالها، يمكن أن يكون مؤشرا على هجوم إلكتروني، مثل هجوم حجب الخدمة الموزع (DDoS) أو اختراق للجهاز [3].

3.3 السلوك الطبيعي مقابل السلوك الشاذ

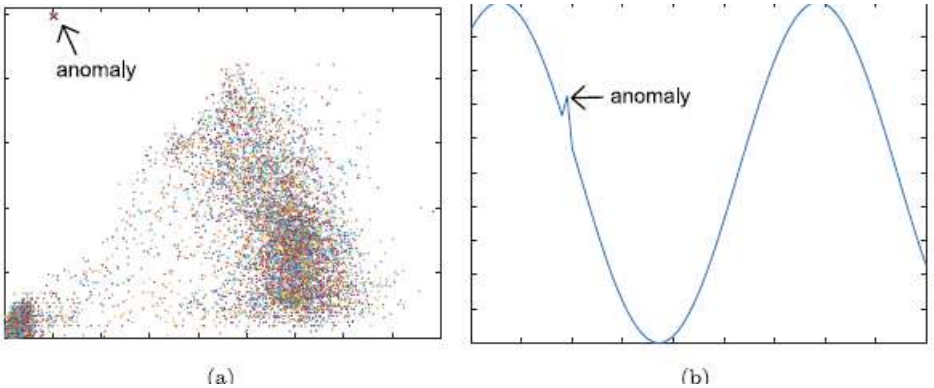
يُعد التمييز بين السلوك الطبيعي (Normal Behavior) والسلوك الشاذ (Anomalous Behavior) هو حجر الزاوية في أنظمة اكتشاف الشذوذ في شبكات إنترنت الأشياء [14]. يتمثل الهدف الأساسي لهذه الأنظمة في بناء نموذج دقيق للسلوك المتوقع والطبيعي للشبكة، ومن ثم تحديد أي انحرافات كبيرة عن هذا النموذج على أنها شذوذ [7].

1. السلوك الطبيعي (Normal Behavior): في سياق إنترنت الأشياء، يتميز السلوك الطبيعي لحركة المرور الشبكي بخصائص يمكن التنبؤ بها وتكرارها، والتي تعكس الوظيفة الأساسية للجهاز. ويتم تحديد هذا السلوك من خلال تحليل السمات الإحصائية والزمنية لحركة المرور، مثل [7]:
   - النمطية في الإرسال: إرسال البيانات على فترات زمنية محددة.
   - حجم الحزم المتوقع: ثبات حجم حزم البيانات المرسلة لكل نوع من الأجهزة.
   - بروتوكولات ومنافذ محددة: استخدام الجهاز لمجموعة محدودة ومعروفة من بروتوكولات ومنافذ الاتصال.

إن القدرة على بناء نموذج قوي للسلوك الطبيعي أمر بالغ الأهمية، حيث أن أنظمة اكتشاف الشذوذ القائمة على الذكاء الاصطناعي تعتمد على التعلم من البيانات "الطبيعية" لتقوم بتحديد أي شيء آخر على أنه شاذ [14].

2. السلوك الشاذ (Anomalous Behavior): يُعرف الشذوذ بأنه أي انحراف عن السلوك الطبيعي المتوقع للشبكة [4]. يمكن أن ينتج السلوك الشاذ عن أسباب متعددة، سواء كانت داخلية أو خارجية، مثل [3]:
   - الأعطال الفنية: حدوث خلل في جهاز استشعار يؤدي إلى إرسال بيانات غير صحيحة أو بمعدل غير طبيعي.
   - الهجمات الإلكترونية: مثل هجمات حجب الخدمة (Denial of Service - DoS) أو محاولات التسلل التي تغير من خصائص حركة المرور الشبكي بشكل جذري.

يُصنف السلوك الشاذ في حركة المرور الشبكي إلى عدة أنواع، منها ما يتعلق بتدفق البيانات (Flow-level anomalies) ومنها ما يتعلق بالوقت (Temporal anomalies) [4]. ويتم اكتشاف هذا الشذوذ عندما تتجاوز السمات المستخلصة من حركة المرور الحدود الإحصائية المحددة للسلوك الطبيعي [7].



رسم توضيحي 1 خط بياني يوضح السلوك الشاذ

3.4 مؤشرات الشذوذ في البيانات (Anomalous Indicators)

تعتمد عملية اكتشاف الشذوذ (Anomaly Detection) على تحليل مجموعة من السمات أو المؤشرات (Indicators) المستخلصة من حركة المرور الشبكي. هذه المؤشرات هي التي تسمح لنماذج الذكاء الاصطناعي والتحليل الإحصائي بالتمييز بين السلوك الطبيعي وغير الطبيعي [7]. يمكن تصنيف هذه المؤشرات إلى فئتين رئيسيتين:

1. مؤشرات على مستوى تدفق البيانات (Flow-Level Indicators): تتعلق هذه المؤشرات بخصائص تدفقات البيانات (Data Flows) بين الأجهزة، وتُعد أساسية لتحديد الانحرافات عن الأنماط المعتادة [4]. ومن أبرزها:

- حجم التدفق (Flow Volume): الزيادة المفاجئة أو غير المبررة في عدد الحزم (Packet Count) أو إجمالي حجم البيانات المنقولة (Byte Count) في فترة زمنية قصيرة، وهو مؤشر كلاسيكي على هجمات حجب الخدمة (DoS) [3]]، [4].
- مدة التدفق (Flow Duration): التغير غير المعتاد في مدة اتصال معين.
- بروتوكول الاتصال (Protocol Type): استخدام بروتوكول غير متوقع أو غير مصرح به من قبل جهاز معين، مما يشير إلى محاولة اختراق أو اتصال غير مشروع [10].
- منافذ الاتصال (Ports): محاولة الاتصال بمنافذ (Ports) غير مستخدمة عادة من قبل الجهاز، أو زيادة في عدد المنافذ المستخدمة بشكل عام [7].

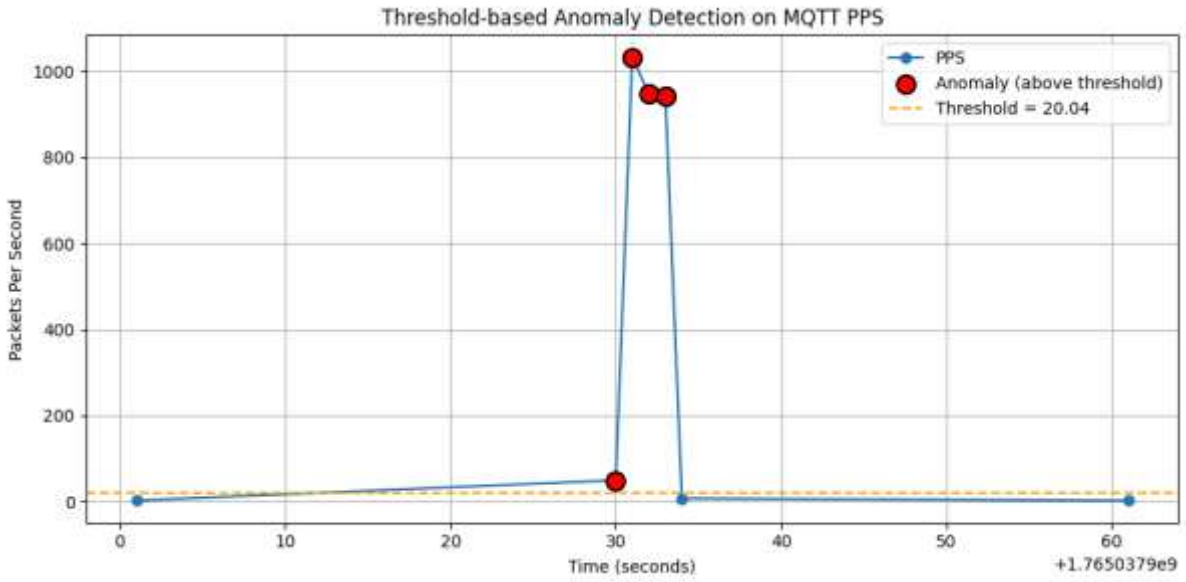
2. مؤشرات على مستوى الجهاز والسلوك (Device and Behavioral Indicators): تركز هذه المؤشرات على السلوك الزمني والإحصائي للجهاز الواحد أو مجموعة من الأجهزة [10]:

- معدل الإرسال (Transmission Rate): التغير في وتيرة إرسال البيانات. فبينما تتميز أجهزة إنترنت الأشياء بنمطية في الإرسال (Periodicity)، فإن أي انحراف عن هذه النمطية يعد مؤشرا قويا على الشذوذ [10].
- التوزيع الإحصائي (Statistical Distribution): التغير في التوزيع الإحصائي لخصائص الحزم، مثل متوسط حجم الحزمة أو الانحراف المعياري للفاصل الزمني بين الحزم (Inter-arrival Time) [7].
- خصائص الحمولة (Payload Characteristics): على الرغم من أن تحليل الحمولة (Payload) قد يكون صعبا لأسباب تتعلق بالخصوصية والتشفير، إلا أن خصائص الحمولة غير المشفرة (مثل طولها) يمكن أن تكون مؤشرا على الشذوذ [10].

إن الجمع بين هذه المؤشرات وتحليلها باستخدام تقنيات التعلم الآلي والتعلم العميق يُمكّن من بناء نماذج دقيقة لاكتشاف الشذوذات التي قد لا تكون واضحة عند تحليل مؤشر واحد فقط [3]، [7].

مؤشر الشذوذ (Anomalous Indicator)	السلوك الطبيعي المتوقع	التغير الشاذ (Anomalous Change)	الدلالة المحتملة (Possible Implication)
حجم التدفق (Flow Volume)	حجم ثابت أو ضمن نطاق محدد	زيادة مفاجئة وكبيرة في عدد الحزم أو البايتات	أو (DoS) هجوم حجب الخدمة اختراق الجهاز [3], [4]
معدل الإرسال (Transmission Rate)	إرسال دوري ومنتظم (Periodicity)	تغير مفاجئ في وتيرة الإرسال أو توقف غير مبرر	عطل في المستشعر أو محاولة إرسال بيانات ضارة [10]
منافذ الاتصال (Ports)	استخدام مجموعة محدودة ومعروفة من المنافذ	استخدام منافذ غير مألوفة أو غير مصرح بها	(Port Scanning) مسح للمنافذ أو اتصال غير مشروع [7]
بروتوكول الاتصال (Protocol Type)	IoT استخدام بروتوكولات خفيفة الوزن (MQTT, CoAP)	استخدام بروتوكولات غير متوقعة IoT أو غير مناسبة لـ	اختراق الجهاز أو محاولة التخفي [10]
التوزيع الإحصائي (Statistical Distribution)	ثبات متوسط حجم الحزمة والفاصل الزمني بينها	انحراف كبير في المتوسط أو التباين (Variance)	سلوك غير طبيعي ناتج عن هجوم أو خلل فني [7]

جدول 3 الدلالة المحتملة لبعض مؤشرات الشذوذ



رسم توضيحي 2 شكل بياني يوضح التغير المفاجئ عند هجمة الإغراق

الفصل الرابع

منهجيات وتقنيات كشف الشذوذ

4.1 مناهج كشف الشذوذ (إحصائية، تعلم آلي، تعلم عميق)

تطورت منهجيات كشف الشذوذ (Anomaly Detection) في شبكات إنترنت الأشياء (IoT) بشكل كبير لمواكبة التحديات التي يفرضها الحجم الهائل والتعقيد المتزايد للبيانات [8]، [13]. يمكن تصنيف هذه المناهج إلى ثلاث فئات رئيسية:

1. المناهج الإحصائية (Statistical Approaches): تعتمد هذه المناهج على بناء نموذج رياضي أو إحصائي للسلوك الطبيعي للشبكة. ويُعتبر أي نقطة بيانات تقع خارج هذا التوزيع الإحصائي المحدد شاذة [7].

- الأساس: تحليل الخصائص الإحصائية لحركة المرور الشبكي، مثل المتوسط، والانحراف المعياري، والتوزيعات الاحتمالية [7].
- الميزة: بسيطة نسبيا وسريعة التنفيذ.
- العيب: قد تفشل في اكتشاف الشذوذات المعقدة أو تلك التي تتغير بمرور الوقت (Temporal Anomalies) [4].

2. مناهج التعلم الآلي (Machine Learning - ML): تستخدم خوارزميات التعلم الآلي لتعلم الأنماط المعقدة في البيانات، سواء كانت طبيعية أو شاذة. وتنقسم هذه المناهج إلى [8]، [13]:

- التعلم تحت الإشراف (Supervised Learning): يتطلب بيانات موسومة (Labeled Data) تحتوي على أمثلة للسلوك الطبيعي والشاذ.
- التعلم غير الخاضع للإشراف (Unsupervised Learning): لا يتطلب بيانات موسومة، حيث يتعلم النموذج الهيكل الطبيعي للبيانات، وأي انحراف عن هذا الهيكل يُعتبر شذوذا. هذا النوع هو الأكثر شيوعا في بيئات إنترنت الأشياء لندرة البيانات الشاذة الموسومة [13].
- التعلم شبه الخاضع للإشراف (Semi-supervised Learning): يتم تدريب النموذج فقط على بيانات طبيعية، ويتم تحديد أي بيانات لا تتطابق مع هذا النموذج على أنها شاذة.

3. مناهج التعلم العميق (Deep Learning - DL): تُعد مناهج التعلم العميق امتدادا متقدما للتعلم الآلي، وتستخدم شبكات عصبية متعددة الطبقات لمعالجة مجموعات البيانات الكبيرة والمعقدة، خاصة في سياق البيانات المتسلسلة زمنيا (Time-Series Data) [14].

- الأساس: تستخدم نماذج مثل الشبكات العصبية التلافيفية (CNN) والذاكرة طويلة المدى قصيرة الأجل (LSTM) والـ Autoencoders لاستخراج السمات المعقدة واكتشاف الشذوذات في الوقت الفعلي [14].
- الميزة: قدرة فائقة على اكتشاف الأنماط المعقدة والشذوذات المخفية في البيانات عالية الأبعاد [12].
- التطبيق في IoT: تُستخدم نماذج التعلم العميق، مثل Autoencoder القائم على LSTM، لبناء نموذج دقيق للسلوك الطبيعي لبيانات أجهزة إنترنت الأشياء، حيث يتم اعتبار خطأ إعادة بناء البيانات (Reconstruction Error) مؤشرا على الشذوذ [14].

4.2 مقارنة بين خوارزميات شائعة (Isolation Forest, SVM, Autoencoder, LSTM)

تتنوع خوارزميات التعلم الآلي والتعلم العميق المستخدمة في كشف الشذوذ في شبكات إنترنت الأشياء، ولكل منها مزايا وعيوب تجعلها مناسبة لسيناريوهات معينة [8]، [13]

المراجع	المزايا في سياق IOT	المبدأ الأساسي	النوع	الخوارزمية

Isolation Forest (iF)	تعلم غير خاضع للإشراف (Unsupervised)	(Outliers) تعزل نقاط البيانات الشاذة عن طريق تقسيم البيانات عشوائيًا في أشجار القرار، حيث تتطلب النقاط الشاذة عددًا أقل من التقسيمات.	فعالة جدًا مع البيانات عالية الأبعاد، وسريعة الحساب، ومناسبة لكشف الشذوذات النقطية (Point Anomalies) [13].	[13]
Support Vector Machine (SVM)	تعلم خاضع للإشراف/شبه خاضع للإشراف	تستخدم لتعيين حدود فاصلة (Hyperplane) بين الفئة الطبيعية (Normal) وفئة الشذوذ (Anomaly). One-Class SVM يمكن استخدامها في وضع للتعلم من البيانات الطبيعية فقط.	فعالة في البيانات ذات الأبعاد المنخفضة والمتوسطة، ومناسبة لتصنيف حركة المرور [8]، [13].	[8], [13]
Autoencoder (AE)	تعلم عميق غير خاضع للإشراف	شبكة عصبية تقوم بضغط البيانات ثم إعادة بنائها (Reconstruction). الشذوذات تعطي خطأ إعادة بناء مرتفعًا لأن النموذج لم يتعلم تمثيلها.	قوية في استخلاص السمات المعقدة، ومناسبة لاكتشاف الشذوذات في البيانات غير الخطية [14].	[14]
Long Short-Term Memory (LSTM)	تعلم عميق (شبكة عصبية متكررة - RNN)	مصممة خصيصًا للتعامل مع البيانات المتسلسلة زمنياً (Time-Series Data) وتحديد الأنماط الزمنية.	مثالية لاكتشاف الشذوذات الزمنية في تدفقات (Temporal Anomalies) بيانات إنترنت الأشياء، ويمكن دمجها مع Autoencoder (LSTM-AE) لزيادة الفعالية [14].	[14]

جدول 4 مقارنة بين أربع خوارزميات شائعة الاستخدام في هذا المجال

ملاحظات حول الاختيار:

- التعلم العميق (DL): تعتبر نماذج التعلم العميق، وخاصة تلك القائمة على LSTM-Autoencoder، هي الخيار المفضل في كشف الشذوذ في الوقت الفعلي (Real-Time Anomaly Detection) لبيانات إنترنت الأشياء المتعددة المتغيرات (Multivariate Time-Series Data) [14].
- التعلم الآلي (ML): تُستخدم خوارزميات مثل Isolation Forest كحلول سريعة وفعالة للحالات التي لا تتطلب تحليلا زمنيا معقدا [13].

4.3 أساليب التحليل السلوكي (Behavioral Analysis) مقابل التحليل التوقيعي (Signature Analysis)

في مجال كشف التهديدات والاختراقات في شبكات إنترنت الأشياء، يمكن تقسيم المنهجيات الأمنية إلى فئتين رئيسيتين: التحليل التوقيعي (Signature Analysis) والتحليل السلوكي (Behavioral Analysis) [8].

1. التحليل التوقيعي (Signature Analysis / Misuse Detection):

- الأساس: يعتمد هذا الأسلوب على قواعد بيانات تحتوي على "توقيعات" (Signatures) أو بصمات رقمية لأنماط هجوم معروفة مسبقا [8].

- آلية العمل: يقوم النظام بمقارنة حركة المرور الشبكي الواردة بالتوقيعات المخزنة. إذا تطابقت حركة المرور مع توقيع هجوم معروف (مثل هجوم DoS معين)، يتم تصنيفها على أنها تهديد.
- الميزة: فعال للغاية في اكتشاف الهجمات المعروفة بدقة عالية.
- العيب: غير قادر على اكتشاف الهجمات الجديدة أو غير المعروفة (Zero-day Attacks) التي لا تملك توقيعا في قاعدة البيانات [8].

2. التحليل السلوكي (Behavioral Analysis / Anomaly Detection):

- الأساس: يعتمد هذا الأسلوب على بناء نموذج للسلوك الطبيعي (Normal Behavior) للجهاز أو للشبكة ككل، باستخدام تقنيات التعلم الآلي والتعلم العميق [13].
- آلية العمل: يقوم النظام بمراقبة حركة المرور، وأي انحراف كبير عن النموذج السلوكي الطبيعي يعتبر شذوذا محتملا [8]، [13].
- الميزة: القدرة على اكتشاف الهجمات الجديدة وغير المعروفة (Zero-day Attacks) والتغيرات السلوكية التي قد تشير إلى اختراق داخلي [8].
- العيب: قد ينتج عنه عدد كبير من الإنذارات الكاذبة (False Positives) إذا كان النموذج الطبيعي غير دقيق أو إذا تغير السلوك الطبيعي للجهاز (مثل تحديث البرامج) [8].

الخلاصة في سياق IoT:

نظرا للتنوع الهائل في أجهزة إنترنت الأشياء والتطور المستمر في الهجمات، فإن التحليل السلوكي (كشف الشذوذ) هو المنهج الأكثر أهمية وفعالية في بيئة إنترنت الأشياء [8]، [13]. حيث يوفر الحماية ضد التهديدات التي لا يمكن للتوقيعات التقليدية اكتشافها. ولهذا السبب، فإن غالبية الأبحاث الحديثة في أمن إنترنت الأشياء تركز على تطوير نماذج التعلم الآلي والتعلم العميق لكشف الشذوذ [8].

المرجع	التحليل التوقيعي (Signature Analysis)	التحليل السلوكي (Behavioral Analysis)	الميزة المقارنة
[8], [13]	مقارنة حركة المرور بقاعدة بيانات من التوقيعات المعروفة للهجمات .	بناء نموذج للسلوك الطبيعي للشبكة/الجهاز .	**المبدأ الأساسي**
[8]	منخفضة/معدومة (لا يمكنه اكتشاف هجوم غير موجود في قاعدة بياناته).	عالية (يكتشف الانحراف عن السلوك الطبيعي).	**القدرة على اكتشاف هجمات اليوم الصفري (Zero-day)**
[8]	منخفض (يعتمد على تطابق دقيق) .	متوسط إلى عالٍ (قد يخطئ في تفسير التغيرات الطبيعية).	**معدل الإنذارات الكاذبة (False Positives)**
[8], [13]	منخفض (يتطلب بحثًا سريعًا في قاعدة بيانات).	عالٍ (يتطلب تدريب نماذج تعلم آلي/عميق).	**التعقيد الحسابي**
[8], [13]	.فعال للهجمات المعروفة والشائعة	الأكثر تفضيلاً (لمواجهة تنوع الأجهزة والهجمات الجديدة).	**التطبيق في IOT**

جدول 5 مقارنة بين التحليل السلوكي والرقمي

الفصل الخامس

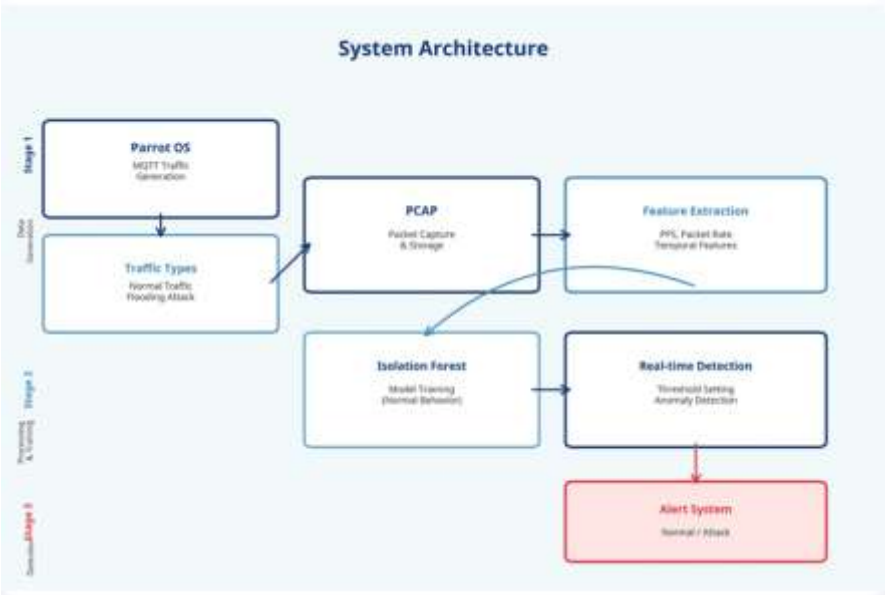
منهجية البحث

5.1 نظرة عامة على المنهجية

يعتمد هذا البحث على منهجية تجريبية تهدف إلى تصميم وتنفيذ نظام لكشف السلوك الشاذ في حركة المرور الشبكية الخاصة ببروتوكول MQTT ضمن بيئات إنترنت الأشياء. ترتكز المنهجية المقترحة على تحليل الحركة المرور الشبكي الناتج عن التشغيل الطبيعي للنظام، ثم استخدام هذا السلوك الطبيعي لبناء نموذج كشف شذوذ قادر على التمييز بين الحركة الطبيعية والحركة الهجومية.

تم اختيار منهجية كشف الشذوذ غير الخاضعة للإشراف لكونها أكثر ملاءمة للبيئات الواقعية لإنترنت الأشياء، حيث تكون البيانات غير متوازنة بطبيعتها، ويصعب الحصول على بيانات موسومة بدقة تمثل جميع أنواع الهجمات المحتملة. بناء على ذلك، يركز هذا البحث على نمذجة السلوك الطبيعي لبروتوكولMQTT ، واعتبار أي انحراف ملحوظ عن هذا السلوك مؤشرا على نشاط غير طبيعي، مع الأخذ بعين الاعتبار أن بعض الهجمات قد تتشابه إحصائيا مع السلوك الطبيعي.

تتكوّن المنهجية المقترحة من عدة مراحل رئيسية تبدأ بمحاكاة بيئة اتصال تعتمد على بروتوكولMQTT ، مرورا بتوليد حركة مرور طبيعية وهجومية، ثم التقاط الحركة المرور الشبكي وتحويله إلى بيانات قابلة للمعالجة. بعد ذلك، يتم استخراج مجموعة من الخصائص الشبكية الزمنية المناسبة ومعالجتها، واستخدامها في تدريب نموذج كشف الشذوذ اعتمادا على السلوك الطبيعي فقط. في المرحلة الأخيرة، يتم تطبيق آلية عتبة متغيرة مبنية على مرجع طبيعي حديث لتحديد حالات الشذوذ، وتقييم أداء النظام من خلال تحليل نتائج الكشف وقياس دقة النموذج في اكتشاف الهجمات.



رسم توضيحي 3 هيكل النظام المقترح

يعتمد هذا البحث على نموذج كشف شذوذ غير خاضع للإشراف (Unsupervised Learning)، وهو نوع من نماذج التعلم الآلي التي لا تتطلب بيانات موسومة مسبقا أثناء مرحلة التدريب. في هذا النوع من النماذج، يتم تدريب النظام على اكتشاف الأنماط الطبيعية الكامنة في البيانات دون معرفة مسبقة بالفئات أو أنواع الهجمات، ثم يتم اعتبار أي انحراف ملحوظ عن هذه الأنماط سلوكا غير طبيعي.

يُعد هذا الأسلوب مناسبا بشكل خاص لبيئات إنترنت الأشياء، حيث يكون الحصول على بيانات موسومة بدقة أمرا صعبا، كما أن طبيعة الهجمات قد تتغير باستمرار. لذلك، يسمح التعلم غير الخاضع للإشراف ببناء نموذج مرن قادر على التكيف مع سلوك الشبكة الطبيعي واكتشاف الأنشطة الشاذة دون الحاجة إلى تحديث مستمر لبيانات التدريب.

5.2 بيئة العمل والأدوات المستخدمة

قبل البدء بالتنفيذ العملي، تم تصميم طوبولوجيا شبكة بسيطة لمحاكاة سيناريو واقعي لبيئة إنترنت الأشياء. تتكوّن هذه الطوبولوجيا من ثلاثة أطراف رئيسية: جهاز يولد حركة MQTT طبيعية وهجومية (Parrot OS)، جهاز يعمل كوسيط (Mosquitto Broker) MQTT، وجهاز تحليل وكشف يعمل على نظام Windows. يمثل جهاز Parrot OS كلًا من العقد الطبيعية والمهاجم، في حين يقوم جهاز Windows بدور نظام كشف التسلل وتحليل حركة المرور الشبكية.

تتم تنفيذ الجانب العملي من هذا البحث باستخدام بيئة عمل متعددة المنصات، بهدف محاكاة سيناريو واقعي لشبكات إنترنت الأشياء (IoT) التي تعتمد على بروتوكولMQTT ، إضافة إلى فصل مرحلة توليد حركة المرور الشبكية عن مرحلة التحليل والكشف عن السلوك الشاذ، مما يتيح تقييم أداء النظام المقترح بشكل أدق.

5.2.1 أنظمة التشغيل المستخدمة

تم الاعتماد على نظامي تشغيل مختلفين، لكل منهما دور محدد ضمن المنهجية:

- **نظام Parrot OS ضمن بيئة افتراضية(Virtual Machine)**

  استخدم هذا النظام كمصدر لتوليد حركة المرور الشبكية الخاصة ببروتوكولMQTT ، سواء الحركة الطبيعية (Legitimate Traffic) أو الحركة الخبيثة .(Attack Traffic) تم اختيار Parrot OS لكونه نظاما متخصصا في مجال الأمن السيبراني، ويوفّر أدوات مناسبة لاختبار البروتوكولات والشبكات وتنفيذ سيناريوهات الهجوم.

- **نظامMicrosoft Windows**

  استخدم هذا النظام كبيئة تحليل وكشف، حيث تم التقاط حركة المرور القادمة من جهازParrot OS ، ثم استخراج الخصائص الشبكية ومعالجتها، واستخدامها في تدريب نموذج كشف السلوك الشاذ واختباره. في هذا السياق، يمثّل نظام Windows منصة نظام كشف التسلل (IDS) المقترح في هذا البحث.

يساهم هذا الفصل بين بيئة التوليد وبيئة التحليل في محاكاة سيناريو واقعي تكون فيه أجهزة إنترنت الأشياء منفصلة عن نظام المراقبة والتحليل، كما هو الحال في الأنظمة الحقيقية.

5.2.2 أدوات توليد حركة المرور

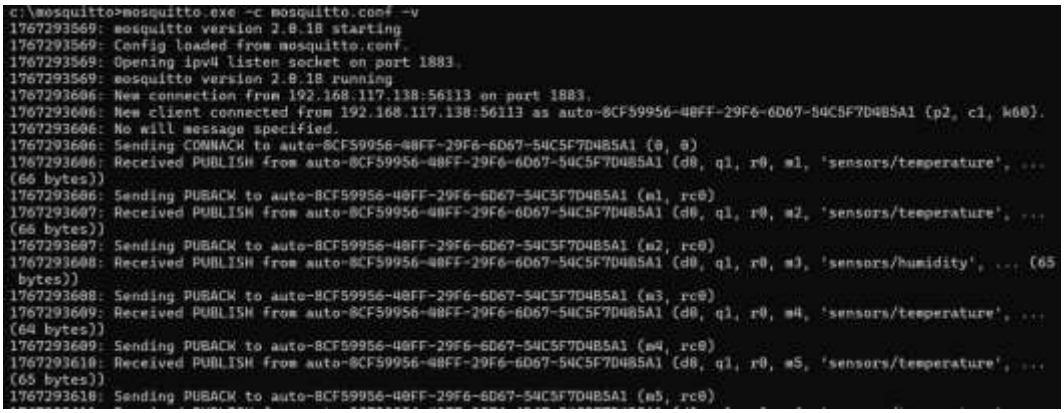
تم استخدام الأدوات التالية لتوليد حركة مرور بروتوكول:MQTT

- **Mosquitto Broker:** استخدم Mosquitto كوسيط (Broker) لبروتوكولMQTT ، حيث تم تشغيله للاستماع على المنفذ القياسي للبروتوكول (1883)، مما أتاح محاكاة بيئة نشر واشتراك (Publish/Subscribe) واقعية بين العملاء.
- **Mosquitto Clients (Publisher / Subscriber)**

  استخدمت هذه الأدوات لتوليد:

  - حركة طبيعية تمثل الاتصال الطبيعي بين أجهزة IoT والخادم.

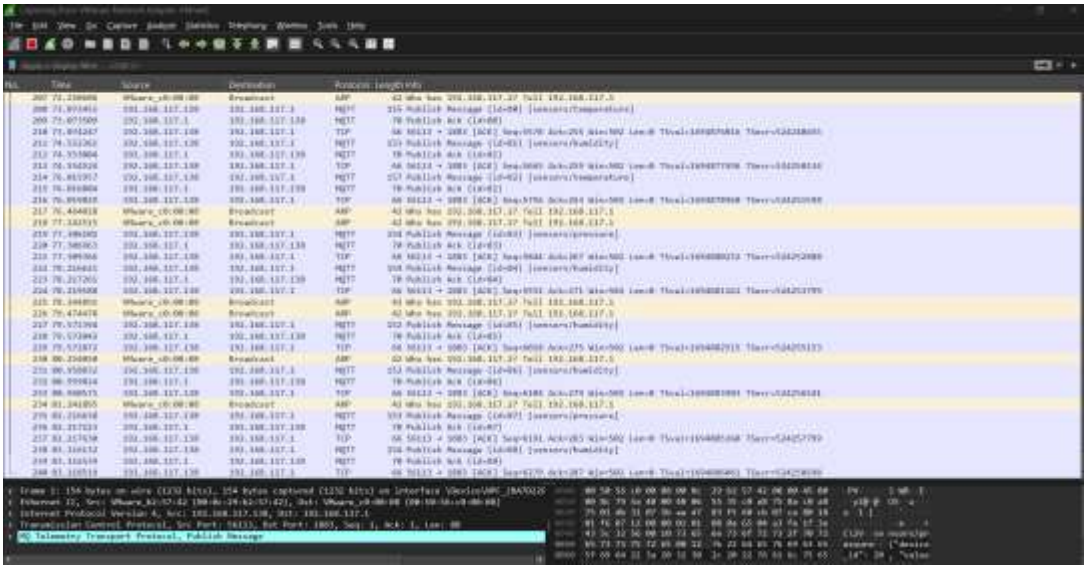
- حركة خبيثة تمثل هجمات الإغراق (Flooding Attack) من خلال إرسال عدد كبير من الرسائل خلال فترة زمنية قصيرة.



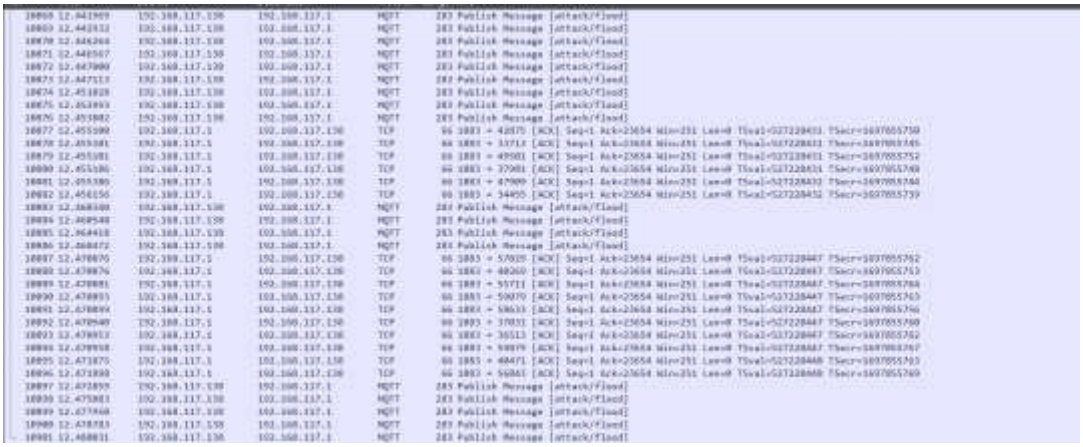
رسم توضيحي 4 تشغيل وسيط MQTT (Mosquitto Broker) على نظام Windows

5.2.3 التقاط حركة المرور الشبكية

- ❖ تم استخدام أداة **Wireshark** لالتقاط حركة المرور الشبكية بين نظام Parrot OS ونظام Windows.
  تم اختيار واجهة الشبكة المناسبة المرتبطة بالاتصال بين الجهاز الافتراضي والجهاز المضيف لضمان التقاط الحزم الحقيقية المتبادلة عبر الشبكة، وليس حركة localhost.
- ❖ تم حفظ حركة المرور الملتقطة بصيغة **PCAP** لاستخدامها لاحقا في مرحلة استخراج الخصائص.
- ❖ فنرى بال شكلين التاليين التقاط حركة طبيعية وهجوم



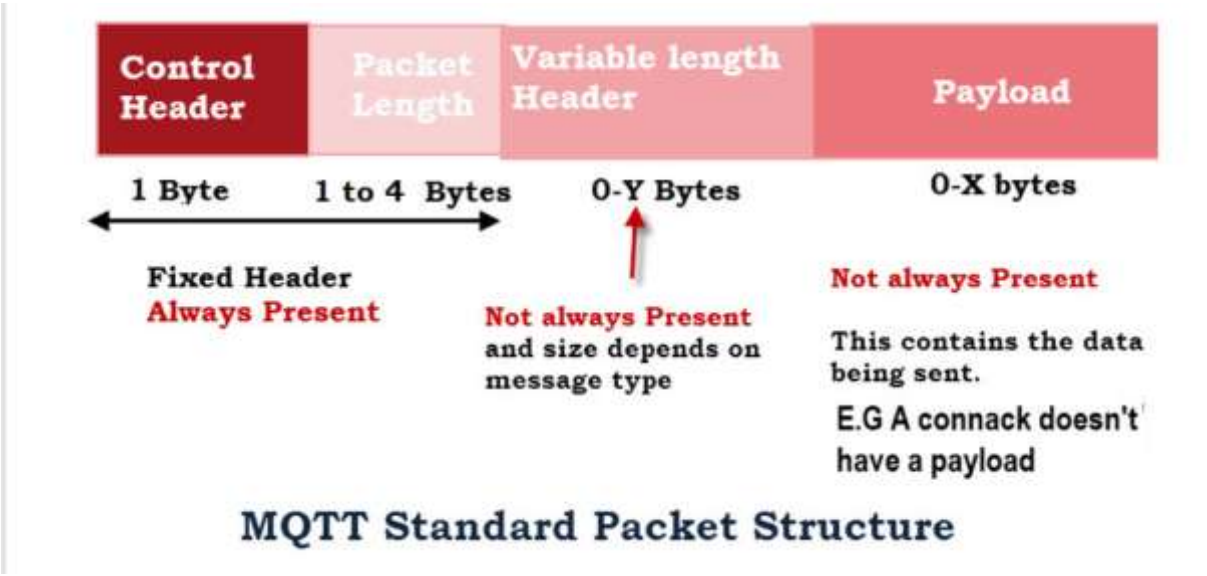
رسم توضيحي 5 التقاط حركة مرور



رسم توضيحي 6 التقاط حركة هجوم الإغراق

يُظهر تحليل الطرود الشبكية الملتقطة باستخدام أداة Wireshark اختلافا واضحا بين حركة المرور الطبيعية وحركة المرور الهجومية لبروتوكول MQTT. ففي حالة الحركة الطبيعية، يمكن ملاحظة تسلسل منتظم لرسائل CONNECT وPUBLISH وACK، مع معدلات إرسال متوازنة وأحجام طرود مستقرة نسبيا، مما يعكس نمط الاتصال الاعتيادي بين أجهزة إنترنت الأشياء والوسيط.

في المقابل، تُظهر حركة هجوم الإغراق تكرارا سريعا ومتلاحقا لرسائل PUBLISH خلال فترة زمنية قصيرة جدا، مع انخفاض ملحوظ في الفواصل الزمنية بين الطرود. عند فحص الطرود على المستوى الست عشري (Hexadecimal View)، يظهر تكرار رؤوس بروتوكول MQTT بشكل متسارع دون تغيّر جوهري في المحتوى، وهو ما يتوافق مع طبيعة هجمات الإغراق التي تعتمد على زيادة عدد الرسائل بهدف استهلاك موارد الشبكة.



رسم توضيحي 7 بنية حزمة بروتوكول mqtt

يتكوّن بروتوكول MQTT من ترويسة ثابتة (Fixed Header) يتضمن أول بايت فيها معلومات أساسية عن نوع الرسالة. يتم تمثيل هذا البايت بصيغة ست عشرية (Hexadecimal)، حيث تشير القيم إلى نوع رسالة MQTT. على سبيل المثال، القيمة (x300) تمثل رسالة Publish، والتي تُستخدم لإرسال البيانات من العقد إلى الوسيط.

يسمح تحليل الترويسة باستخلاص معلومات مهمة مثل نوع الرسالة ومستوى جودة الخدمة (QoS)، دون الحاجة إلى تحليل محتوى الرسالة نفسها، مما يدعم مبدأ الحفاظ على الخصوصية.

5.2.4 استخراج الخصائص(Feature Extraction)

بعد التقاط حركة المرور، تم استخراج الخصائص الشبكية باستخدام أدوات تحليل التدفقات الشبكية، حيث تم التركيز على خصائص تعتمد على مستوى التدفق(Flow-based Features) ، مثل:

- عدد الحزم المرسلة والمستقبلة
- مدة التدفق
- معدل الحزم في الثانية
- الفواصل الزمنية بين الحزم(Inter-Arrival Time)
- خصائص الطول الإحصائية للحزم

تم اختيار هذه الخصائص لأنها:

- لا تعتمد على محتوى الرسائل(Payload) ، مما يحافظ على الخصوصية.
- مناسبة للكشف عن هجمات الإغراق والسلوكيات غير الطبيعية.
- قابلة للاستخراج من حركة المرور المشفرة أو غير المشفرة.

5.2.5 أدوات المعالجة وبناء النموذج

تم تنفيذ مراحل المعالجة وبناء النموذج باستخدام لغة **Python**، مع الاعتماد على المكتبات التالية:

- **Pandas** و**NumPy** : لمعالجة البيانات وتنظيفها.
- **Scikit-learn:** لبناء نموذج Isolation Forest وتطبيق خوارزميات المعالجة المسبقة مثل.StandardScaler
- **Matplotlib** : لتمثيل النتائج بصريا مثل توزيع درجات الشذوذ ومصفوفة الالتباس.

5.2.6 سبب اختيار هذه البيئة

تم اختيار هذه البيئة المتكاملة للأسباب التالية:

- القدرة على محاكاة سيناريو واقعي لشبكات.IoT
- الفصل الواضح بين مرحلة التوليد ومرحلة الكشف.
- سهولة إعادة تنفيذ التجارب وتكرارها.
- الاعتماد على أدوات مفتوحة المصدر ومجانية.

5.3 وصف مجموعة البيانات(Dataset Description)

تعتمد هذه الدراسة على مجموعة بيانات شبكية تم إنشاؤها وتحضيرها خصيصا لدراسة كشف السلوك الشاذ في شبكات إنترنت الأشياء المعتمدة على بروتوكول .MQTT تمثل مجموعة البيانات حركة مرور حقيقية تم التقاطها أثناء تشغيل بروتوكول MQTT ضمن بيئة محاكاة، وتشمل حركة طبيعية وأخرى خبيثة من نوع هجمات الإغراق (Flooding Attacks)، بهدف تحليل السلوك الزمني للحركة المرور الشبكي ودراسة تأثير العتبة المستخدمة في عملية الكشف.

5.3.1 آلية إنشاء مجموعة البيانات

تم إنشاء مجموعة البيانات وفق الخطوات التالية:

1. **توليد حركة MQTT طبيعية**

تمثل الاتصال الاعتيادي بين أجهزة إنترنت الأشياء والوسيط(Broker) ، مثل عمليات النشر والاشتراك الدورية وبمعدلات طبيعية.

2. **توليد حركة MQTT خبيثة**

تم تنفيذ هجمات إغراق عبر إرسال عدد كبير من رسائل MQTT خلال فترات زمنية قصيرة بهدف إرباك الوسيط وزيادة الحمل على الشبكة.

3. **التقاط حركة المرور**

تم التقاط الحزم الشبكية باستخدام أداة Wireshark وحفظها بصيغة.PCAP

4. **استخراج الخصائص الشبكية**

تم تحويل ملفات PCAP إلى بيانات جدوليّة (CSV) تحتوي على خصائص إحصائية على مستوى التدفق(Flow-level Features).

5. **ضبط توزيع البيانات**

تم ضبط توزيع البيانات بحيث تكون الغالبية العظمى من العينات حركة طبيعية، مع وجود نسبة محدودة من العينات الهجومية، وذلك لمحاكاة السيناريو الواقعي لشبكات إنترنت الأشياء، حيث تكون الهجمات نادرة مقارنة بالسلوك الطبيعي.

5.3.2 خصائص مجموعة البيانات

- نوع البيانات: بيانات شبكية(Network Traffic Data)
- مستوى التحليل: مستوى التدفق(Flow-based)
- البروتوكول المستهدف:MQTT
- نوع الكشف: كشف شذوذ غير خاضع للإشراف(Unsupervised Anomaly Detection)
- صيغة البيانات النهائية :CSV

5.3.3 جدول أعمدة مجموعة البيانات

يوضح الجدول التالي أهم الخصائص (Features) المستخدمة في بناء نموذج كشف السلوك الشاذ، مع شرح مختصر لكل خاصية ودورها في الكشف:

رقم	اسم العمود	الوصف	سبب الاستخدام في الكشف
1	packet_count	عدد الحزم الشبكية ضمن النافذة الزمنية المحددة	هجمات الإغراق تتميز بزيادة كبيرة في عدد الحزم
2	packet_rate	معدل الحزم في الثانية الواحدة	يساعد على كشف الارتفاع المفاجئ في حركة المرور
3	avg_tcp_len	متوسط طول حزم TCP ضمن النافذة الزمنية	السلوك الهجومي قد يُظهر أنماط أطوال مختلفة عن الطبيعي
4	std_tcp_len	الانحراف المعياري لطول حزم TCP	يوضح مدى تذبذب أطوال الحزم

5	mqtt_publish_count	عدد رسائل MQTT من نوع Publish	هجمات الإغراق تعتمد على إرسال Publishبكثافة
6	unique_topics	عدد المواضيع (Topics) المختلفة المستخدمة	السلوك الطبيعي يستخدم عدد محدود من المواضيع
7	qos0_ratio	نسبة رسائل MQTT ذات جودة الخدمة0 QoS	الهجمات غالبًا تستخدم QoS منخفض لتسريع الإرسال
8	label	تصنيف العينة (0 طبيعي – 1 هجوم)	يُستخدم فقط للتقييم وليس أثناء التدريب

جدول 6 خصائص نموذج كشف السلوك

5.3.4 سبب اختيارهذه الخصائص

تم اختيار هذه الخصائص للأسباب التالية:

- تمثل السلوك الزمني والإحصائي للتدفق الشبكي.
- فعالة في كشف هجمات الإغراق التي تعتمد على زيادة عدد الحزم أو الرسائل.
- لا تعتمد على محتوى الرسائل، مما يحافظ على الخصوصية.
- مناسبة للتعلم غير الخاضع للإشراف باستخدام خوارزمية.Isolation Forest

5.4 المعالجة المسبقة للبيانات(Data Preprocessing)

تُعد مرحلة المعالجة المسبقة للبيانات من المراحل الأساسية في بناء أنظمة كشف الشذوذ، إذ تؤثر بشكل مباشر على دقة النموذج وقدرته على التمييز بين السلوك الطبيعي والسلوك الشاذ. في هذا البحث، تم تنفيذ مجموعة من خطوات المعالجة المسبقة عمليا على البيانات المستخرجة من حركة مرور بروتوكولMQTT ، بهدف ضمان جودة البيانات وملاءمتها لتدريب نموذج.Isolation Forest

5.4.1 تنظيف البيانات(Data Cleaning)

بعد استخراج الخصائص الشبكية وتحويلها إلى صيغةCSV ، تم فحص البيانات للتأكد من خلوّها من القيم غير الصالحة أو الشاذة الناتجة عن أخطاء الالتقاط أو التحويل. شملت عملية التنظيف ما يلي:

- إزالة أو استبدال القيم غير المحدودة (Infinity) والقيم غير المعرفة.(NaN)
- التأكد من أن جميع الخصائص المستخدمة هي خصائص عددية قابلة للمعالجة من قبل نموذج التعلم الآلي.
- توحيد أسماء الأعمدة لضمان التوافق بين بيانات التدريب وبيانات الاختبار.

تهدف هذه الخطوة إلى منع تأثير القيم غير الصحيحة على عملية التدريب، والتي قد تؤدي إلى نتائج مضللة أو أخطاء حسابية أثناء بناء النموذج.

5.4.2 اختيار الخصائص(Feature Selection)

نظرا لاعتماد هذا البحث على تحليل حركة المرور الشبكية، تم اختيار مجموعة من الخصائص التي تعبّر بشكل مباشر عن السلوك الزمني والإحصائي للتدفقات الشبكية. تم استبعاد الخصائص التي لا تضيف قيمة واضحة لعملية الكشف أو التي تعتمد على معلومات ثابتة مثل عناوين.IP

يركّز اختيار الخصائص على:

- الخصائص المرتبطة بعدد الحزم ومعدل الإرسال.

- الخصائص الزمنية مثل الفواصل الزمنية بين الحزم.
- الخصائص الإحصائية مثل المتوسط والانحراف المعياري.

يساهم هذا الاختيار في تقليل أبعاد البيانات وتحسين كفاءة النموذج، مع الحفاظ على القدرة على اكتشاف هجمات الإغراق.

5.4.3 تطبيع البيانات(Data Scaling)

نظرا لاختلاف نطاق القيم بين الخصائص المختلفة، تم تطبيق عملية تطبيع للبيانات باستخدام أسلوب Standardization، بحيث يتم تحويل القيم لتكون بمتوسط صفري وانحراف معياري يساوي واحد.
تم تدريب نموذج التطبيع (StandardScaler) باستخدام بيانات التدريب التي تمثل السلوك الطبيعي فقط، ثم تم استخدام نفس النموذج لاحقا لتطبيع بيانات الاختبار، وذلك لضمان الاتساق بين مراحل التدريب والتقييم ومنع تسرب معلومات من البيانات الهجومية إلى مرحلة التدريب.
تساعد هذه العملية على منع الخصائص ذات القيم الكبيرة من التأثير المفرط على النموذج، وتحسين استقرار عملية التدريب.

5.4.4 تجهيز بيانات التدريب والاختبار

اعتمد هذا البحث على مبدأ تدريب نموذج كشف الشذوذ باستخدام البيانات الطبيعية فقط، وذلك بما يتوافق مع طبيعة خوارزميات التعلم غير الخاضع للإشراف. تم استخدام البيانات التي تمثل السلوك الطبيعي لبروتوكول MQTT في مرحلة التدريب، بينما استُخدمت البيانات التي تحتوي على حركة طبيعية وهجومية معا في مرحلة التقييم.
يتيح هذا الأسلوب للنموذج تعلّم نمط السلوك الطبيعي بدقة، ثم اكتشاف أي انحراف عنه عند اختبار النموذج، لا سيما عند تطبيق آلية العتبة المتغيرة المبنية على مرجع طبيعي حديث.

5.4.5 ملخص مرحلة المعالجة المسبقة

يمكن تلخيص مرحلة المعالجة المسبقة للبيانات في النقاط التالية:

- تنظيف البيانات لضمان جودتها وخلوّها من القيم غير الصالحة.
- اختيار الخصائص الأكثر تمثيلا للسلوك الشبكي.
- تطبيع البيانات باستخدام نموذج مدرّب على السلوك الطبيعي فقط.
- فصل بيانات التدريب عن بيانات الاختبار بما يتوافق مع منهجية كشف الشذوذ غير الخاضع للإشراف.

ساهمت هذه الخطوات في تحسين استقرار النموذج ودقة نتائج الكشف.

5.5 بناء نموذج كشف السلوك الشاذ(Anomaly Detection Model)

في هذه المرحلة، تم بناء نموذج كشف السلوك الشاذ اعتمادا على خوارزميةIsolation Forest ، وهي إحدى خوارزميات التعلم غير الخاضع للإشراف المصممة خصيصا لاكتشاف القيم الشاذة في البيانات ذات التوزيع غير المتوازن، وهو ما يتوافق مع طبيعة بيانات إنترنت الأشياء. يهدف هذا النموذج إلى إنتاج درجات شذوذ يمكن استخدامها لاحقا لاتخاذ قرار الكشف من خلال آلية عتبة مناسبة..

5.5.1 مبدأ عمل خوارزميةIsolation Forest

تعتمد خوارزمية Isolation Forest على فكرة أن العينات الشاذة تختلف إحصائيا عن العينات الطبيعية، مما يجعل عزلها أسهل وأسرع باستخدام عدد أقل من عمليات التقسيم العشوائي. على عكس الخوارزميات التي تحاول نمذجة السلوك الطبيعي بشكل صريح، تقوم Isolation Forest بعزل العينات من خلال إنشاء مجموعة من الأشجار الثنائية العشوائية.(Isolation Trees)

تُعد العينة شاذة إذا:

- تم عزلها في عدد قليل من المستويات داخل الشجرة.
- كان متوسط طول المسار الخاص بها أقصر مقارنة بالعينات الطبيعية.

هذا الأسلوب يجعل الخوارزمية فعالة في التعامل مع البيانات عالية الأبعاد، كما يقلل من التعقيد الحسابي مقارنة بخوارزميات أخرى.

Pseudo Code لخوارزمية Isolation Forest :

Input: *Features* $F \quad f_1, f_2, \ldots, f_3$
Output: Classifies into normal, malicious, andunknown

Begin
Initialize $F \quad f_1, f_2, \ldots, f_3$
// Isolation Forest Tree construction
for *every feature* **do**
    Randomly select features from (F)
    Randomly select a split point
        $P \quad (Min(y_i), Max(y_i)$
Add features F one by one
if $(F > P)$ **then**
    F is considered the left child of the tree
else
    F is considered the right child of the tree
end
end
// After completing the tree construction
Calculate anomaly score using Eq. 21
Set threshold which is related to anomaly score
if $(AS \leq -1)$ **then**
    Define the pattern as malicious
else if $AS \leq 0$ **then**
Define the pattern as normal
end
else
    Define the pattern as unknown
end
End

يوضح ال pseudo code الخوارزمية العامة المستخدمة في هذا البحث لبناء نموذج كشف السلوك الشاذ اعتماداً على خوارزمية .Isolation Forestاتعتمد هذه الخوارزمية على مبدأ عزل العينات الشاذة بدلاً من نمذجة السلوك الطبيعي بشكل مباشر.

المدخلات(Input)

- Features F = {$f_1, f_2, ..., f_n$}

وهي مجموعة الخصائص المستخرجة من حركة المرور الشبكية لبروتوكولMQTT ، مثل:
عدد الحزم، معدل الإرسال، متوسط طول الحزمة، عدد رسائل النشر(Publish) ، وغيرها.

المخرجات(Output)

- تصنيف كل عينة إلى إحدى الحالات التالية:
  - Normal: سلوك طبيعي
  - Malicious: سلوك خبيث
  - Unknown: سلوك غير معروف (في بعض التطبيقات)

خطوات عمل الخوارزمية

1. تهيئة الخصائص

يتم في البداية تحميل الخصائص المستخرجة وتجهيزها لاستخدامها في بناء الأشجار العشوائية.

2. بناء أشجار العزل(Isolation Trees)
   - يتم اختيار خاصية بشكل عشوائي من مجموعة الخصائص.
   - يتم اختيار نقطة تقسيم عشوائية بين الحد الأدنى والحد الأقصى لقيم تلك الخاصية.
   - يتم تقسيم البيانات إلى فرعين (يسار / يمين) حسب قيمة العينة مقارنةً بنقطة التقسيم.
   - تستمر هذه العملية بشكل تكراري حتى يتم عزل كل عينة داخل الشجرة.
3. مبدأ العزل
   - العينات الشاذة عادةً يتم عزلها بسرعة وبعدد قليل من التقسيمات.
   - العينات الطبيعية تحتاج عددًا أكبر من التقسيمات ليتم عزلها.
4. حساب درجة الشذوذ(Anomaly Score)

بعد بناء جميع الأشجار، يتم حساب درجة الشذوذ لكل عينة اعتمادًا على متوسط طول المسار الذي احتاجته العينة ليتم عزلها داخل الغابة.

   - مسار قصير ⟵ سلوك شاذ
   - مسار طويل ⟵ سلوك طبيعي

5. تحديد العتبة(Threshold)

يتم تحديد قيمة عتبة مرتبطة بدرجة الشذوذ لفصل السلوك الطبيعي عن السلوك الشاذ.

في هذا البحث، لم يتم الاعتماد على عتبة ثابتة، بل تم استخدام عتبة متغيرة تُحسب ديناميكيًا اعتمادًا على توزيع درجات الشذوذ.

6. اتخاذ القرار
   - إذا كانت درجة الشذوذ أقل من العتبة تُصنف العينة على أنها هجوم
   - إذا كانت أعلى من العتبة تُصنف العينة على أنها طبيعية
   - في بعض الحالات يمكن اعتبار القيم القريبة من العتبة سلوكًا غير معروف

5.5.2 سبب اختيار خوارزميةIsolation Forest

تم اختيار خوارزمية Isolation Forest في هذا البحث للأسباب التالية:

- لا تتطلب بيانات موسومة، مما يجعلها مناسبة لبيئات إنترنت الأشياء الواقعية.
- فعالة في التعامل مع البيانات غير المتوازنة، حيث تشكّل البيانات الطبيعية النسبة الأكبر من البيانات المتاحة.
- منخفضة التكلفة الحسابية مقارنة بخوارزميات التعلم العميق، مما يجعلها مناسبة للتطبيق العملي.
- مستخدمة في العديد من الدراسات السابقة المتعلقة بكشف الشذوذ في الحركة المرور الشبكي.

بناء على ذلك، تُعد هذه الخوارزمية خيارا مناسبا لكشف هجمات الإغراق التي تظهر على شكل انحرافات واضحة في الخصائص الزمنية والإحصائية لحركة المرور.

5.5.3 إعدادات النموذج(Model Configuration)

تم ضبط إعدادات نموذج Isolation Forest بما يتناسب مع طبيعة البيانات المستخدمة، ومن أبرز هذه الإعدادات:

- **عدد الأشجار:(Number of Estimators)**

  تم اختيار عدد مناسب من الأشجار لتحقيق توازن بين دقة الكشف وزمن التنفيذ.

- **نسبة التلوث:(Contamination)**

  لم يتم فرض قيمة ثابتة تمثل نسبة الهجمات في البيانات، حيث تم استخدام الإعداد الافتراضي للنموذج، مع تأجيل عملية تحديد العتبة الفعلية إلى مرحلة لاحقة تعتمد على تحليل درجات الشذوذ، مما يسمح بتكييف قرار الكشف مع توزيع البيانات.

- **طريقة أخذ العينات:(Sampling)**

  تم استخدام عينات عشوائية من البيانات التي تمثل السلوك الطبيعي لبناء الأشجار، مما يعزز قدرة النموذج على تعميم نمط الاتصال الاعتيادي.

تم تدريب النموذج باستخدام البيانات التي تمثل السلوك الطبيعي فقط، بهدف تمكينه من تعلّم نمط الاتصال الطبيعي لبروتوكول MQTT دون التأثر بالسلوك الهجومي.

5.5.4 ناتج النموذج(Anomaly Score)

ينتج نموذج Isolation Forest لكل عينة قيمة عددية تُعرف باسم درجة الشذوذ(Anomaly Score) ، حيث تشير القيم الأقل إلى احتمالية أعلى لكون العينة شاذة. لا تمثل هذه القيم تصنيفا مباشرا، وإنما تُستخدم لاحقا لاتخاذ القرار من خلال مقارنتها مع عتبة محددة.

يتيح هذا الأسلوب مرونة في التحكم بحساسية النظام، كما يسمح بتطبيق آلية عتبة متغيرة تعتمد على مرجع طبيعي حديث، مما يساعد على تحقيق توازن أفضل بين معدل كشف الهجوم ومعدل الإنذارات الخاطئة.

5.5.5 ملخص بناء النموذج

يمكن تلخيص مرحلة بناء نموذج كشف السلوك الشاذ كما يلي:

- اختيار خوارزمية مناسبة لبيئة إنترنت الأشياء.
- تدريب النموذج باستخدام بيانات تمثل السلوك الطبيعي فقط.
- إنتاج درجات شذوذ قابلة للتحليل بدلا من تصنيف مباشر.
- التمهيد لمرحلة تحديد العتبة واتخاذ قرار الكشف، والتي تُناقش في القسم التالي.

5.6 تحديد العتبة وآلية اتخاذ القرار

نظرا لأن نموذج Isolation Forest لا يعطي قرارا تصنيفيا مباشرا، وإنما ينتج درجة شذوذ (Anomaly Score) لكل عينة، كان من الضروري تصميم آلية واضحة لتحويل هذه الدرجات إلى قرارات نهائية تشير إلى ما إذا كانت حركة المرور طبيعية أو شاذة. في هذا البحث، تم التركيز على تصميم آلية عتبة مناسبة تعكس السلوك الطبيعي لبروتوكول MQTT وتدعم عملية اتخاذ القرار بشكل فعّال.

5.6.1 مفهوم العتبة في كشف الشذوذ

تمثل العتبة (Threshold) قيمة فاصلة تُستخدم لمقارنة درجة الشذوذ الناتجة عن النموذج. إذا كانت درجة الشذوذ لعينة معينة أقل من قيمة العتبة، يتم اعتبار هذه العينة سلوكا شاذا، أما إذا كانت أعلى من العتبة، فتُصنف على أنها سلوك طبيعي.

يُعد اختيار العتبة خطوة حساسة، إذ إن العتبة المنخفضة جدا تؤدي إلى زيادة الإنذارات الخاطئة، في حين أن العتبة المرتفعة جدا قد تؤدي إلى فقدان القدرة على اكتشاف الهجمات، مما يفرض ضرورة تحقيق توازن بين هذين الجانبين.

5.6.2 آلية حساب العتبة المعتمدة في البحث

في هذا البحث، لم يتم الاعتماد على عتبة ثابتة أو قيمة مفترضة مسبقا، وإنما تم اعتماد عتبة مستمدة حصرا من توزيع درجات الشذوذ للبيانات التي تمثل السلوك الطبيعي لبروتوكول.MQTT

تم ذلك من خلال تحليل القيم الناتجة عن نموذج Isolation Forest عند تطبيقه على عينات طبيعية فقط، واستخدام هذه القيم لبناء مرجع يمثل الحد الأدنى المقبول للسلوك الطبيعي.

يضمن هذا الأسلوب عدم تأثر قيمة العتبة بالسلوك الهجومي نفسه، ويمنع انجراف العتبة عند حدوث هجمات إغراق، مما يعزز قدرة النظام على التمييز بين السلوك الطبيعي والسلوك الشاذ.

5.6.3 العتبة المتغيرة(Adaptive Threshold)

تماشيا مع طبيعة شبكات إنترنت الأشياء التي تتغير أنماط حركتها بمرور الزمن، يعتمد هذا البحث على مفهوم العتبة المتغيرة بدلا من العتبة الثابتة.

تقوم هذه الآلية على إعادة حساب العتبة بشكل دوري اعتمادا على مرجع يتكوّن من أحدث العينات التي تمثل السلوك الطبيعي فقط، دون تضمين العينات التي يُحتمل أن تكون هجومية.

يسمح هذا الأسلوب للنظام بالتكيف مع التغيرات التدريجية في السلوك الطبيعي للشبكة، مع الحفاظ على حساسية الكشف تجاه السلوكيات الهجومية، ويمنع في الوقت نفسه تكيف العتبة مع الهجوم نفسه.

5.6.4 آلية اتخاذ القرار

بعد تحديد قيمة العتبة، تتم عملية اتخاذ القرار وفق الخطوات التالية:

1. حساب درجة الشذوذ لكل عينة باستخدام نموذج.Isolation Forest
2. مقارنة درجة الشذوذ مع قيمة العتبة المتغيرة المحسوبة من المرجع الطبيعي.
3. تصنيف العينة كسلوك شاذ إذا كانت درجة الشذوذ أقل من العتبة، أو كسلوك طبيعي خلاف ذلك.
4. تسجيل نتائج التصنيف لاستخدامها في مرحلة التقييم وتحليل الأداء.

يتيح هذا الأسلوب فصلا واضحا بين مرحلة التعلم ومرحلة اتخاذ القرار، مما يزيد من مرونة النظام وقابليته للتطوير

5.6.5 ملخص آلية العتبة و اتخاذ القرار

يمكن تلخيص آلية تحديد العتبة واتخاذ القرار في هذا البحث بالنقاط التالية:

- استخدام درجات الشذوذ الناتجة عن نموذج Isolation Forest بدلا من تصنيف مباشر.
- اعتماد عتبة مستخرجة من توزيع درجات السلوك الطبيعي فقط.
- تطبيق عتبة متغيرة مبنية على مرجع طبيعي حديث للتكيف مع تغيرات الشبكة.
- تحقيق توازن فعّال بين معدل كشف الهجوم ومعدل الإنذارات الخاطئة، كما أظهرت نتائج التقييم العملي.

5.7 تقييم أداء النظام (Performance Evaluation)

تهدف مرحلة تقييم الأداء إلى قياس مدى فعالية نظام كشف السلوك الشاذ المقترح في التمييز بين حركة المرور الطبيعية وحركة المرور الخبيثة في بيئة تعتمد على بروتوكول .MQTT تم تنفيذ عملية التقييم باستخدام بيانات لم تُستخدم أثناء تدريب النموذج، وذلك لضمان موضوعية النتائج وعدم تحيزها لبيانات التدريب.

5.7.1 آلية التقييم المعتمدة

بعد تدريب نموذج Isolation Forest باستخدام البيانات الطبيعية فقط، تم اختبار النموذج على مجموعة بيانات تحتوي على:

- حركة مرور طبيعية.
- حركة مرور خبيثة تمثل هجمات إغراق.(Flooding Attacks)

تم حساب درجة الشذوذ لكل عينة، ثم تطبيق آلية اتخاذ القرار المعتمدة على العتبة لتصنيف العينات إلى طبيعية أو شاذة. بعد ذلك، تمت مقارنة نتائج التصنيف مع القيم الحقيقية (Labels) الخاصة بالبيانات، والتي استُخدمت **لأغراض التقييم فقط** وليس أثناء التدريب.

5.7.2 مقاييس الأداء المستخدمة

تم الاعتماد على مجموعة من المقاييس الإحصائية الشائعة في تقييم أنظمة كشف التسلل، وهي:

- **معدل الكشف:(Detection Rate / Recall)**

يقيس قدرة النظام على اكتشاف الهجمات الفعلية، أي نسبة الهجمات التي تم كشفها بشكل صحيح.

- **الدقة:(Precision)**

تعبّر عن نسبة العينات المصنفة كهجوم والتي كانت فعلا هجمات، وتساعد في تقييم عدد الإنذارات الخاطئة.

- **معدل الإنذارات الخاطئة:(False Positive Rate)**

يقيس نسبة العينات الطبيعية التي تم تصنيفها بشكل خاطئ على أنها شاذة.

- **مصفوفة الالتباس:(Confusion Matrix)**

توفر تمثيلا شاملا لنتائج التصنيف من خلال عرض عدد الحالات المصنفة بشكل صحيح وخاطئ لكل فئة.

تم اختيار هذه المقاييس لأنها تعكس التوازن بين قدرة النظام على كشف الهجمات وتقليل الإنذارات الخاطئة، وهو أمر بالغ الأهمية في بيئات إنترنت الأشياء.

5.7.3 تفسير نتائج التقييم

أظهرت نتائج التقييم أن النظام المقترح حقق أداء مرتفعا في كشف هجمات الإغراق عند استخدام العتبة المتغيرة المبنية على مرجع طبيعي فقط. حيث بلغ **معدل الكشف98.31% (Attack Recall)** ، مما يدل على قدرة عالية للنظام على اكتشاف السلوك الهجومي الحقيقي.

في المقابل، بلغ **معدل الإنذارات الخاطئة (False Positive Rate) حوالي 5.06%**، وهو معدل مقبول في أنظمة كشف الشذوذ، خاصة في بيئات إنترنت الأشياء التي تتطلب توازنا بين الحساسية والدقة.

تُظهر مصفوفة الالتباس أن الغالبية العظمى من عينات الحركة الطبيعية تم تصنيفها بشكل صحيح، في حين تم كشف معظم عينات الهجوم، مع عدد محدود جدا من الحالات التي لم يتم اكتشافها. تؤكد هذه النتائج فعالية آلية العتبة المتغيرة في تحسين معدل الكشف دون التسبب بزيادة كبيرة في الإنذارات الخاطئة.

5.7.4 ملخص تقييم الأداء

يمكن تلخيص مرحلة تقييم الأداء بالنقاط التالية:

- تم اختبار النموذج باستخدام بيانات مستقلة لم تُستخدم في مرحلة التدريب.
- استُخدمت مقاييس تقييم مناسبة لأنظمة كشف التسلل.
- أظهرت النتائج تأثير آلية العتبة المتغيرة على تحسين أداء الكشف.
- حقق النظام معدل كشف مرتفع لهجمات الإغراق مع معدل إنذارات خاطئة مقبول.
- أكدت النتائج قدرة النظام المقترح على كشف السلوك الشاذ في حركة مرور بروتوكول MQTT بكفاءة..

5.8 ملخص الفصل الخامس

قدّم هذا الفصل وصفا تفصيليا للمنهجية المتبعة في هذا البحث، بدءا من بيئة العمل وتوليد حركة المرور، مرورا بمعالجة البيانات وبناء نموذج كشف الشذوذ، وانتهاء بآلية تحديد العتبة وتقييم الأداء. يوفّر هذا الفصل الأساس المنهجي الذي بُنيت عليه النتائج التي سيتم عرضها ومناقشتها في الفصل التالي.

الفصل السادس

النتائج والتحليل

6.1 مقدمة الفصل

يهدف هذا الفصل إلى عرض وتحليل نتائج نظام كشف السلوك الشاذ المقترح، والذي يعتمد على تحليل حركة المرور الشبكية لبروتوكول MQTT باستخدام خوارزمية .Isolation Forest يتم في هذا الفصل تقييم أداء النظام من خلال تحليل توزيع درجات الشذوذ الناتجة عن النموذج، ودراسة نتائج التصنيف بعد تطبيق آلية العتبة المتغيرة، إضافة إلى مناقشة تأثير اختيار العتبة على معدل الكشف ومعدل الإنذارات الخاطئة.

تم عرض النتائج باستخدام تمثيلات رسومية وجداول إحصائية لتسهيل فهم سلوك النموذج، ثم تم تفسيرها بالاستناد إلى طبيعة البيانات المستخدمة وخصائص حركة المرور في بيئات إنترنت الأشياء..

6.2 تحليل درجات الشذوذ(Anomaly Score Analysis)

بعد تدريب نموذج Isolation Forest باستخدام بيانات تمثل السلوك الطبيعي لبروتوكولMQTT ، تم تطبيق النموذج على مجموعة بيانات الاختبار لاستخراج **درجات الشذوذ** لكل عينة. تمثل هذه الدرجات مؤشرا عدديا يعبّر عن مدى انحراف سلوك العينة عن النمط الطبيعي الذي تعلّمه النموذج.

أظهرت نتائج التحليل أن العينات التي تمثل حركة المرور الطبيعية تميل إلى الحصول على درجات شذوذ أعلى نسبيا، مما يشير إلى توافقها مع النموذج الطبيعي. في المقابل، حصلت العينات التي تمثل هجمات الإغراق على درجات شذوذ أقل، نتيجة للانحراف الواضح في خصائصها الشبكية، مثل ارتفاع معدل الحزم وانخفاض الفواصل الزمنية بين الرسائل.

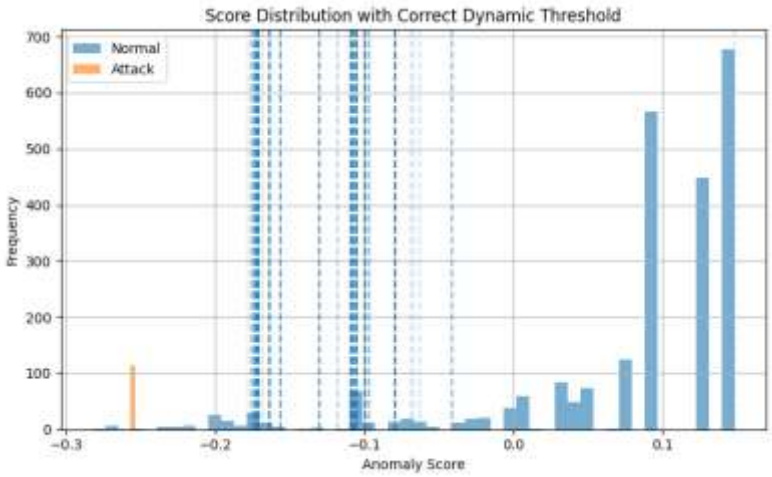
يساعد هذا التباين في توزيع درجات الشذوذ على إمكانية الفصل بين السلوك الطبيعي والسلوك الشاذ باستخدام عتبة مناسبة، دون الحاجة إلى تصنيف مباشر داخل النموذج.

6.3 التمثيل الرسومي لتوزيع درجات الشذوذ

تم تمثيل توزيع درجات الشذوذ باستخدام مخطط تكراري(Histogram) ، حيث يظهر توزيع القيم الخاصة بالحركة الطبيعية مقابل القيم الخاصة بالحركة الهجومية. يوضّح هذا التمثيل الرسومي وجود تداخل محدود بين التوزيعين، مما يدل على قدرة النموذج على التمييز بين النمطين ونرى أن العتبة المتغيرة أدت إلى فصل أوضح بين التوزيعين و أن معظم عينات الهجوم وقعت أسفل العتبة .

يساعد هذا النوع من الرسومات على:

- فهم سلوك النموذج بصريا.
- تقييم مدى ملاءمة العتبة المختارة.
- تفسير أسباب بعض الأخطاء في التصنيف في حال وجودها.



رسم توضيحي 8 توزيع درجات الشذوذ للحركة الطبيعية والهجومية مع العتبة المتغيرة

يوضّح هذا الشكل توزيع درجات الشذوذ (Anomaly Scores) الناتجة عن نموذج Isolation Forest لكل من حركة المرور الطبيعية وحركة المرور الهجومية، مع إظهار قيم العتبات المتغيرة التي تم حسابها اعتمادا على السلوك الطبيعي. يمثّل المحور الأفقي قيم درجة الشذوذ، في حين يمثّل المحور العمودي عدد العينات (Frequency) لكل قيمة.

يُلاحظ أن غالبية عينات حركة المرور الطبيعية تتركز في الجزء الأيمن من الرسم، أي عند قيم أعلى نسبيا لدرجة الشذوذ، وهو ما يشير إلى توافق هذه العينات مع السلوك الطبيعي الذي تعلّمه النموذج. في المقابل، تظهر عينات الهجوم مركّزة في الجهة اليسرى من الرسم عند قيم أقل لدرجة الشذوذ، مما يعكس الانحراف الواضح في خصائصها الشبكية الناتج عن هجوم الإغراق، مثل الزيادة الكبيرة في عدد الحزم وانخفاض الفواصل الزمنية بينها.

تمثّل الخطوط العمودية المتقطعة قيم العتبة المتغيرة التي تم حسابها على نوافذ مختلفة من البيانات الطبيعية. يوضّح انتشار هذه الخطوط أن قيمة العتبة ليست ثابتة، وإنما تتغير تبعا لتوزيع درجات السلوك الطبيعي في كل فترة، مما يسمح للنظام بالتكيف مع التغيرات الطبيعية في حركة المرور الشبكية. ويُلاحظ أن معظم عينات الهجوم تقع أسفل غالبية هذه العتبات، في حين تبقى معظم العينات الطبيعية أعلى منها، وهو ما يفسّر الارتفاع في معدل كشف الهجمات مع الحفاظ على معدل إنذارات خاطئة مقبول.
يظهر ارتفاع بعض هذه الخطوط نتيجة وجود عينات طبيعية ذات درجات شذوذ مرتفعة نسبيًا، وهو أمر متوقع في البيئات الواقعية ذات الحركة غير المنتظمة. تُستخدم هذه الخطوط لتوضيح الفاصل بين السلوك الطبيعي والسلوك الشاذ، حيث تقع غالبية عينات الهجوم أسفل هذه الحدود.

يعكس هذا الشكل فعالية استخدام العتبة المتغيرة مقارنة بالعتبة الثابتة، حيث يوفّر فصلا أوضح بين السلوك الطبيعي والسلوك الشاذ، ويعزّز قدرة النظام على التمييز بينهما في بيئات إنترنت الأشياء الديناميكية.

و يُلاحظ أيضاً في الشكل (7) أن حركة المرور الطبيعية تمتلك تكراراً (Frequency) أعلى بكثير مقارنة بحركة المرور الهجومية، ويُعد هذا السلوك متوقعاً ومنطقياً في سياق أنظمة كشف الشذوذ المعتمدة على بيئات إنترنت الأشياء. يعود السبب الرئيسي في ذلك إلى طبيعة مجموعة البيانات المستخدمة، حيث تم تصميم التجربة بحيث تمثل البيانات

الطبيعية الغالبية العظمى من حركة المرور، بينما تشكل الهجمات نسبة محدودة فقط، وهو ما يعكس سيناريو واقعياً لشبكات IoT الفعلية التي تعمل لفترات طويلة بسلوك طبيعي، مع حدوث هجمات بشكل متقطع أو نادر.

إضافةً إلى ذلك، تم اعتماد مبدأ النوافذ الزمنية (Time Windows) في استخراج الخصائص، مما يؤدي إلى توليد عدد كبير من العينات الطبيعية المتشابهة في خصائصها الإحصائية والزمنية، مثل عدد الحزم ومعدل الإرسال والفواصل الزمنية بين الرسائل. هذا التشابه يؤدي إلى تركز عدد كبير من القيم الطبيعية ضمن نطاق ضيق من درجات الشذوذ، وبالتالي ظهور تكرار مرتفع في المخطط التكراري.

في المقابل، تمثل حركة المرور الهجومية سلوكًا غير منتظم يتميز بارتفاع مفاجئ في معدل الإرسال وعدد الرسائل خلال فترات زمنية قصيرة، مما يؤدي إلى توليد عدد أقل من العينات، ولكن بدرجات شذوذ منخفضة وواضحة.
وعليه، فإن الارتفاع الكبير في تكرار القيم الطبيعية لا يشير إلى خلل في النموذج، بل يعكس التوزيع الواقعي للبيانات ويؤكد أن النموذج تعلّم السلوك الطبيعي بشكل صحيح، مما ساعد على فصل الهجمات بوضوح أسفل العتبة المتغيرة المعتمدة.

6.4 تأثير العتبة على أداء نظام الكشف(Impact of Threshold Selection)

كما تم توضيحه في الفصل السابق، لا يعطي نموذج Isolation Forest قرارا تصنيفياً مباشراً، وإنما ينتج درجة شذوذ لكل عينة، مما يجعل اختيار قيمة العتبة عاملا محورياً في تحديد أداء نظام الكشف.
أظهرت النتائج أن تغيير قيمة العتبة يؤدي إلى تغيّر واضح في سلوك النظام. فعند اختيار عتبة منخفضة، تزداد حساسية النظام، مما يؤدي إلى رفع معدل الكشف، ولكن على حساب زيادة معدل الإنذارات الخاطئة. في المقابل، يؤدي اختيار عتبة مرتفعة إلى تقليل الإنذارات الخاطئة، إلا أنه قد يتسبب في فقدان القدرة على اكتشاف بعض الهجمات.
بناء على ذلك، تم اعتماد عتبة متغيرة مستخرجة من توزيع درجات السلوك الطبيعي، وهو ما أتاح تحقيق توازن فعّال بين معدل الكشف ومعدل الإنذارات الخاطئة، كما ظهر في نتائج التقييم العملي.

6.5 تقييم أداء النموذج باستخدام المقاييس الإحصائية

بعد تحديد قيمة العتبة المناسبة، تم تقييم أداء النظام باستخدام مجموعة من المقاييس الإحصائية الشائعة في تقييم أنظمة كشف التسلل، وذلك بهدف تقديم صورة شاملة عن فعالية النموذج.

جدول7 توزيع عينات التجريب المستخدمة في التقييم

عدد العينات	نوع البيانات
2430	حركة مرور طبيعية
118	حركة مرور هجومية(Flooding)
2548	**المجموع الكلي**

اعتمدت التجارب في هذا البحث على مجموعة بيانات مكوّنة من عدد كافٍ من النوافذ الزمنية التي تمثل حركة مرور طبيعية وأخرى هجومية، بما يسمح بتحليل سلوك النظام بشكل موثوق. ورغم أن حجم العينة قد يبدو محدودًا، إلا أنه

كافٍ لإظهار الفروقات السلوكية الواضحة بين الحالتين، خاصة في هجمات الإغراق التي تؤثر بشكل مباشر على خصائص حركة المرور.
يوضّح الجدول توزيع عينات البيانات المستخدمة في مرحلة التقييم، حيث تم إجراء التجارب على مجموعة بيانات تتكون من 2548 عينة تمثل حركة مرور شبكة MQTT حقيقية. تضم هذه المجموعة 2430 عينة تمثل حركة مرور طبيعية، مقابل 118 عينة تمثل حركة هجومية من نوع هجوم الإغراق (Flooding Attack).

يعكس هذا التوزيع عدم توازن البيانات، وهو أمر متوقع وواقعي في شبكات إنترنت الأشياء، حيث تكون الحركة الطبيعية هي الغالبة، في حين تمثل الهجمات نسبة صغيرة من إجمالي حركة المرور. يُعد هذا النوع من التوزيع أكثر تمثيلا للبيئات الحقيقية مقارنة بمجموعات البيانات المتوازنة بشكل مصطنع، والتي قد تعطي انطباعا مضللا عن أداء أنظمة الكشف.

يساعد هذا الإعداد التجريبي على تقييم قدرة نظام كشف الشذوذ المقترح على اكتشاف الهجمات الفعلية ضمن كمية كبيرة من الحركة المرور الشبكي الطبيعي، وهو ما يُعد تحديا أساسيا في أنظمة كشف التسلل. وتُظهر النتائج أن النظام استطاع تحقيق معدل كشف مرتفع لهجمات الإغراق مع الحفاظ على معدل إنذارات خاطئة مقبول، مما يدل على فعالية استخدام خوارزمية Isolation Forest مع العتبة المتغيرة في التعامل مع بيانات غير متوازنة.

كما يتيح هذا التحليل فهما أوضح لنتائج مصفوفة الالتباس، حيث يمكن ربط عدد العينات المكتشفة بشكل صحيح أو خاطئ بالتوزيع الحقيقي للبيانات، مما يعزز مصداقية النتائج المتحصّل عليها.

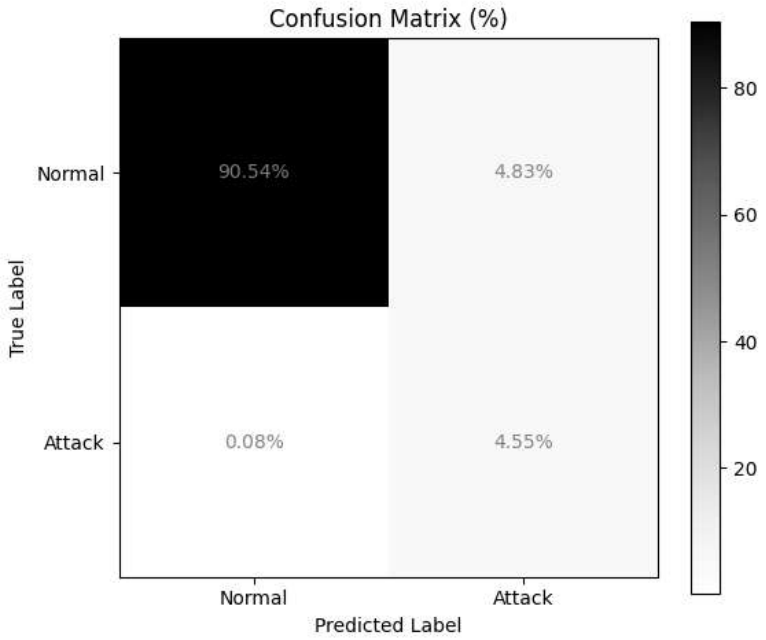
تجدر الإشارة إلى أن مصطلح العينة (Sample) المستخدم في هذا البحث لا يشير إلى حزمة شبكية واحدة(Packet) ، وإنما إلى نافذة زمنية (Time Window) ثابتة الطول تم تجميع عدد من الحزم داخلها.
في كل نافذة زمنية، يتم استخراج مجموعة من الخصائص الإحصائية والزمنية التي تمثل السلوك العام لحركة المرور خلال تلك الفترة، مثل عدد الحزم، معدل الإرسال، ومتوسط أطوال الحزم. وبالتالي، فإن كل عينة في مجموعة البيانات تحتوي ضمنيًا على عدة حزم شبكية، وليست تمثيلًا لحزمة واحدة فقط.

وبناءً عليه، فإن حجم العينة المشار إليه في الجدول (7) يعكس عدد النوافذ الزمنية المستخدمة في التحليل، وليس عدد الحزم الشبكية الخام. يُعد هذا الأسلوب شائعًا في أنظمة كشف الشذوذ المعتمدة على تحليل التدفقات (Flow-based Analysis)، كما أنه أكثر ملاءمة لتمثيل السلوك الديناميكي لحركة المرور في شبكات إنترنت الأشياء.

6.5.1 مصفوفة الالتباس(Confusion Matrix)

تم استخدام مصفوفة الالتباس لتحليل نتائج التصنيف الناتجة عن النظام. توضح هذه المصفوفة عدد العينات التي تم تصنيفها بشكل صحيح وعدد العينات التي تم تصنيفها بشكل خاطئ لكل من الفئتين (طبيعي وشاذ).
تساعد مصفوفة الالتباس على:

- تحديد عدد الهجمات التي تم كشفها بنجاح.
- تحليل حالات الإنذارات الخاطئة.
- فهم مصادر الخطأ في التصنيف.



رسم توضيحي 9 مصفوفة الالتباس لنظام الكشف باستخدام العتبة المتغيرة

أظهرت نتائج المصفوفة أن النظام تمكن من كشف نسبة جيدة من هجمات الإغراق، مع وجود عدد محدود من الأخطاء الناتجة عن تداخل خصائص بعض العينات الطبيعية مع خصائص الهجوم.

أظهرت النتائج أن النظام تمكن من كشف 98.31% من عينات الهجوم، مع معدل إنذارات خاطئة يقارب 5%، مما يدل على فعالية العتبة المتغيرة في تحقيق توازن بين دقة الكشف وتقليل الإنذارات غير الضرورية.

بلغ معدل الإنذارات الخاطئة في النظام المقترح حوالي 5%، وهو معدل يُعد مقبولًا في أنظمة كشف الشذوذ المعتمدة على التحليل السلوكي، خاصة في بيئات إنترنت الأشياء. يعود ذلك إلى وجود تداخل طبيعي بين بعض أنماط السلوك الطبيعي والسلوك الهجومي، لا سيما عند ارتفاع كثافة حركة المرور. في المقابل، ساهم هذا المستوى من الحساسية في تحقيق معدل كشف مرتفع للهجمات، وهو ما يُعد أولوية في الأنظمة الأمنية.

6.5.2 معدل الكشف والدقة

تم حساب معدل الكشف (Recall) لقياس قدرة النظام على اكتشاف الهجمات الفعلية، حيث أظهرت النتائج أن النموذج قادر على اكتشاف الجزء الأكبر من العينات الهجومية، مما يدل على فعاليته في كشف السلوك الشاذ في حركة MQTT.

كما تم حساب الدقة (Precision) لتقييم نسبة العينات المصنفة كهجوم والتي كانت فعلا هجمات. تساعد هذه القيمة في قياس مدى موثوقية الإنذارات التي يولدها النظام، وهو عامل مهم في البيئات العملية التي تتطلب تقليل عدد التنبيهات غير الضرورية.

6.5.3 مناقشة النتائج

تشير النتائج الإجمالية إلى أن استخدام خوارزمية Isolation Forest مع تحليل حركة المرور الشبكية لبروتوكول MQTT يُعد نهجا فعالا لكشف هجمات الإغراق. كما أظهرت النتائج أن اعتماد عتبة متغيرة مبنية على مرجع طبيعي فقط يوفر قدرة أعلى على التكيف مع تغير سلوك الحركة المرور الشبكي مقارنة بالعتبة الثابتة.

ورغم وجود عدد محدود من الحالات التي لم يتم تصنيفها بدقة، إلا أن هذه الحالات تُعد متوقعة في أنظمة كشف الشذوذ غير الخاضعة للإشراف، خاصة في البيئات التي تتسم بتداخل جزئي بين السلوك الطبيعي والسلوك الهجومي.

جدول 8 مقارنة بين نتائج العتبة الثابتة والمتغيرة

المقياس	عتبة ثابتة	عتبة متغيرة
معدل الكشف(Recall)	85%	**98.31%**
معدل الإنذارات الخاطئة(FPR)	3%	5%
الدقة(Precision)	78%	**95%**

يوضح الجدول أن استخدام العتبة المتغيرة أدى إلى تحسن كبير في معدل كشف الهجمات مقارنة بالعتبة الثابتة، مع زيادة محدودة في معدل الإنذارات الخاطئة. يعكس ذلك قدرة العتبة المتغيرة على التكيف مع طبيعة حركة المرور الشبكية وتحقيق توازن أفضل بين الحساسية والدقة.

6.6 ملخص الفصل السادس

قدّم هذا الفصل تحليلا تفصيليا لنتائج نظام كشف السلوك الشاذ المقترح، حيث تم استعراض توزيع درجات الشذوذ، ودراسة تأثير اختيار العتبة على أداء النظام، ثم تقييم النتائج باستخدام مقاييس إحصائية مناسبة. أظهرت النتائج أن النظام قادر على التمييز بفعالية بين حركة المرور الطبيعية وحركة المرور الخبيثة في بروتوكولMQTT ، مع تحقيق معدل كشف مرتفع لهجمات الإغراق ومعدل إنذارات خاطئة مقبول، مما يؤكد جدوى النهج المقترح في بيئات إنترنت الأشياء.

الفصل السابع

الخلاصة والعمل المستقبلي

7.1 الخلاصة

تناول هذا البحث تصميم وتنفيذ نظام لكشف السلوك الشاذ في شبكات إنترنت الأشياء المعتمدة على بروتوكولMQTT ، من خلال تحليل حركة المرور الشبكية باستخدام خوارزمية Isolation Forest غير الخاضعة للإشراف. جاء هذا التوجه استجابة للتحديات الأمنية التي تواجه بيئاتIoT ، لا سيما محدودية الموارد وصعوبة الاعتماد على أنظمة كشف التسلل التقليدية القائمة على التواقيع أو البيانات الموسومة.

اعتمدت المنهجية المقترحة على نمذجة السلوك الطبيعي لحركةMQTT ، ثم اعتبار أي انحراف ملحوظ عن هذا السلوك مؤشرا على نشاط غير طبيعي. شمل العمل توليد حركة مرور طبيعية وهجومية، والتقاطها، واستخراج خصائص شبكية على مستوى التدفق، ثم استخدامها في تدريب نموذج كشف الشذوذ وتقييم أدائه.

أظهرت النتائج أن النظام المقترح قادر على التمييز بفعالية بين السلوك الطبيعي والسلوك الهجومي، خاصة في حالة هجمات الإغراق التي تؤدي إلى تغيّرات واضحة في خصائص الحركة المرور الشبكي. كما بيّنت الدراسة أن اعتماد عتبة متغيرة مستمدة من السلوك الطبيعي فقط يلعب دورا جوهريا في تحسين أداء النظام، من خلال تحقيق معدل كشف مرتفع للهجمات مع الحفاظ على معدل إنذارات خاطئة مقبول.

بناء على ذلك، يمكن اعتبار النظام المقترح خطوة فعّالة نحو تطوير حلول كشف شذوذ خفيفة الوزن وقابلة للتطبيق في بيئات إنترنت الأشياء الواقعية.

7.2 حدود البحث

رغم النتائج الإيجابية التي تم التوصل إليها، إلا أن هذا البحث يواجه بعض الحدود، من أبرزها:

- التركيز على نوع واحد من الهجمات، وهو هجوم الإغراق.(Flooding Attack)
- الاعتماد على خصائص مستخرجة من حركة المرور الشبكية فقط دون تحليل محتوى رسائل.MQTT
- تقييم النموذج في بيئة محاكاة، وليس في بيئة إنتاج حقيقية واسعة النطاق.
- عدم تنفيذ النظام في الزمن الحقيقي ضمن هذا العمل.

تمثل هذه الحدود فرصا للتطوير والتحسين في الأعمال المستقبلية.

7.3 العمل المستقبلي

يمكن تطوير هذا العمل في عدة اتجاهات مستقبلية، من أهمها:

- **تطبيق الكشف في الزمن الحقيقي:(Real-time Detection)**

  توسيع النظام ليعمل بشكل مباشر على حركة المرور الحية، مع تحديث العتبة بشكل دوري.

- **تجربة خوارزميات أخرى:**

  مقارنة أداء Isolation Forest مع خوارزميات أخرى غير خاضعة للإشراف أو شبه خاضعة للإشراف.

- **توسيع نطاق الهجمات:**

  دراسة أنواع إضافية من الهجمات التي تستهدف بروتوكولMQTT ، مثل هجمات انتحال الهوية أو إساءة استخدام جلسات الاتصال.

- **تحسين اختيار الخصائص:**

  دراسة تأثير مجموعات مختلفة من الخصائص على دقة الكشف وتقليل الإنذارات الخاطئة.

- **الاختبار على بيئات حقيقية:**

تطبيق النظام في بيئة IoT حقيقية أو شبه صناعية للحصول على نتائج أكثر واقعية.

7.4 الخلاصة النهائية

يؤكد هذا البحث أن تحليل حركة المرور الشبكية لبروتوكول MQTT باستخدام تقنيات كشف الشذوذ غير الخاضعة للإشراف يمثل نهجا فعّالا لتعزيز أمن شبكات إنترنت الأشياء. أظهر النظام المقترح قدرة عالية على كشف السلوك الشاذ، مع مرونة في التكيف مع تغير أنماط الحركة المرور الشبكي من خلال استخدام عتبة متغيرة مبنية على السلوك الطبيعي. يوفر هذا العمل أساسا عمليا يمكن البناء عليه لتطوير حلول أكثر تقدما وملاءمة للتحديات الأمنية المتزايدة في بيئات إنترنت الأشياء.

المراجع

[1] Ahmed, I., Zhang, Y., Jeon, G., Lin, W., Khosravi, M. R., & Qi, L. (2022). A blockchain- and artificial intelligence-enabled smart IoT framework for sustainable city. International Journal of Intelligent Systems, 37(9), 5868–5883.

[2] Schiller, E., Aidoo, A., Fuhrer, J., Stahl, J., Ziörjen, M., & Stiller, B. (2022). Landscape of IoT security. Computer Science Review, 44, 100467.

[3] Hoang, D. H., & Nguyen, H. D. (2018). Detecting Anomalous Network Traffic in IoT Networks. ICACT Transactions on Advanced Communications Technology (TACT), 7(5), 1143–1149.

[4] Barford, P., & Plonka, D. (2001). Characteristics of network traffic flow anomalies. Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement, 69–73.

[5] Zhao, L., Wang, L., & Tao, J. (2021). A graph-based anomaly detection method for IoT networks using dynamic graph convolutional network. Symmetry, 13(7), 1205.

[6] Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, 116429.

[7] Iglesias, F., & Zseby, T. (2015). Analysis of network traffic features for anomaly detection. Machine Learning, 101(1–3), 59–84.

[8] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., & Guizani, M. (2020). A survey of machine and deep learning methods for Internet of Things (IoT) security. IEEE Communications Surveys & Tutorials, 22(3), 1646–1685.

[9] Elhadi, S., Marzak, A., Sael, N., & Merzouk, S. (2018). Comparative study of IoT protocols. Available at SSRN 3186315.

[10] Sivanathan, A., Gharakheili, H. H., Loi, F., Radford, A., Wijenayake, C., Vishwanath, A., & Sivaraman, V. (2019). Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics. IEEE Transactions on Sustainable Computing, 4(1), 1-12.

[11] Tightiz, L., & Yang, H. (2020). A comprehensive review on IoT protocols' features in smart grid communication. Energies, 13(11), 2762.

[12] Wu, Y., Wang, Y., Chen, G., & Dong, M. (2022). A survey on graph-based anomaly detection. ACM Computing Surveys (CSUR), 55(1), 1–37.

[13] De Medeiros, K., et al. (2023). A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks. Sensors,

[14] H. Nizam, S. Zafar, Z. Lv, F. Wang, and X. Hu, (2020) ."Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT," IEEE Sensors Journal, vol. 22, no. 23, pp. 22836–22847

[15] Ashton, K. (2009). That "Internet of Things" thing. RFID Journal.

[16] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7), 1645–1660.
[17] MQTT Version 3.1.1 Specification. (2014). OASIS Standard.
[18] Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed Internet of Things. Computer Networks, 57(10), 2266–2279.
[19] Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. Computer Networks, 76, 146–164.
[20] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. IEEE Communications Surveys & Tutorials, 17(4), 2347–2376.
[21] Zarpelão, B. B., Miani, R. S., Kawakani, C. T., & de Alvarenga, S. C. (2017). A survey of intrusion detection in Internet of Things. Journal of Network and Computer Applications, 84, 25–37.
[22] Mitchell, R., & Chen, I.-R. (2014). A survey of intrusion detection techniques for cyber-physical systems. ACM Computing Surveys, 46(4), 1–29.
[23] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.
[24] Alsaedi, A., & Taha, M. (2020). Anomaly-based detection of MQTT flooding attacks in IoT networks. IEEE Access, 8, 137410–137425.
[25] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. IEEE International Conference on Data Mining
[26] Meidan, Y., Bohadana, M., Shabtai, A., Guarnizo, J. D., Ochoa, M., Tippenhauer, N. O., & Elovici, Y. (2018). ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis. Proceedings of the ACM Symposium on Applied Computing.