مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (39) العدد (5) 2017
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (39) No. (5) 2017

Using Neural Networks to Build an Intrusion Detection System based on Standard Dataset (KDD99)

Dr. Hassan Alahmad•
Ruba Ibrahim••

(Received 15 / 2 / 2017. Accepted 12 / 10 / 2017)

□ ABSTRACT □

Network security has always been a critical issue when it comes to organizations confidentiality and individuals privacy, especially that delicate and important information is being transferred all the time through networks, also many systems have been increasingly relying on web services, such as e-government services, banking services, e-mail and e-commerce. That's why Intrusion Detection Systems (IDS) have become a very important component, which is widely used to protect information and reduce the damage caused by network attacks and violations.

The main purpose of this research is to build an intrusion detection system using neural networks, and KDDCup 99 data set since it is the mostly used comprehensive data set in intrusion detection domain, and it is shared by many Researchers which provide a great opportunity to compare results. And studying the influence of feature reduction on the process of intrusion detection. First, the preprocessing step was applied on the dataset, then few techniques have been applied on the dataset to decrease the number of the features used in the neural network classifier. The MATLAB software was used to train and test the dataset. The accuracy, detection rate and false rates were measured.

Key words : Intrusion Detection System, Neural Network, KDDCup99 dataset, Preprocessing, Feature Reduction, Matlab.

•Assistant Professor, Department of computer and automatic control, Faculty of mechanical and electrical engineering, Tishreen University, Lattakia, Syria.
••Postgraduate student, Department of computer and automatic control, Faculty of mechanical and electrical engineering, Tishreen University, Lattakia, Syria.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (39) العدد (5) 2017
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (39) No. (5) 2017

استخدام الشبكات العصبونية في بناء نظام كشف تسلل اعتماداً على مجموعة بيانات قياسية (KDD99)

د. حسن الأحمد•

ربا ابراهيم••

(تاريخ الإيداع 15 / 2 / 2017. قُبِل للنشر في 12/ 10 / 2017)

□ ملخّص □

تعد مسألة أمن الشبكات مسألة هامة ودقيقة عندما يتعلق الأمر بخصوصية المنظمات والأفراد، خاصة عند تناقل معلومات مهمة وحساسة عبر هذه الشبكات، من جهة أخرى ازداد اعتماد معظم الأنظمة مؤخرا على خدمات الويب المتطورة سواء كانت خدمات حكومية، أو خدمات مصرفية، أو بريد إلكتروني أو تسويق إلكتروني. كل ما سبق زاد من أهمية أنظمة كشف التسلل التي تعد مكون مهم جدا لحماية المعلومات والحد من الضرر الناتج عن الهجمات والاختراقات الشبكية.

الهدف الرئيسي لهذا البحث هو بناء نظام كشف تسلل شبكي باستخدام الشبكات العصبونية، بالاعتماد على مجموعة البيانات KDDCup99 نظرا لكونها حاليا أشمل مجموعة بيانات مستخدمة في مجال كشف التسلل، كما تمت مشاركتها من قبل العديد من الباحثين مما يتيح فرصة لمقارنة النتائج. بالإضافة إلى دراسة تأثير تخفيض السمات على دقة عملية الكشف. تم بداية معالجة مجموعة البيانات المختارة معالجة تحضيرية، ثم تطبيق عدة تقنيات بهدف تخفيض عدد السمات المستخدمة في مصنف الشبكة العصبونية. تم استخدام برنامج الماتلاب للتدريب ولاختبار مجموعة البيانات وقياس دقة المصنف، بالإضافة إلى قياس معدل الكشف ومعدلات الأخطاء.

الكلمات المفتاحية : نظام كشف التسلل، شبكات عصبونية، مجموعة بيانات KDDCup99، معالجة تحضيرية، تخفيض سمات، ماتلاب.

• مدرس، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية سورية.

•• طالبة دراسات عليا (ماجستير) ، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية سورية.

مقدمة:

يعد كشف التسلل من أهم وأوسع المواضيع في مجال حماية أمن الشبكات، المنظمات، والأفراد. وهناك اليوم العديد من النهج المستخدمة ضمن هذا المجال، كما يوجد العديد من أنظمة كشف التسلل المتاحة، لكن للأسف لا يوجد أي منها من دون عيوب حتى الآن. لذلك برزت الحاجة إلى استمرار إجراء الأبحاث على أنظمة كشف التسلل بهدف التوصل إلى هيكلية مثلى تحقق نسبة حماية عالية.

يقترح هذا البحث استخدام تقنية الشبكات العصبونية في بناء نظام كشف التسلل، وعرضَ المراحل المتّبعة لبناء هذا النظام، بداية من دراسة قواعد البيانات بهدف اختيار القاعدة المناسبة للبحث. ثم بعد ذلك البحث في عدد من التقنيات بهدف تحسين أداء نظام كشف التسلل المقترح، لتحقيق أفضل النتائج الممكنة من حيث دقة النظام، معدل الكشف، ومعدلات الأخطاء.

أهمية البحث وأهدافه:

إنَّ أهم العقبات التي تواجه نظم كشف التسلل هي معدلات الأخطاء وبشكل خاص الأخطاء الإيجابية False Positives (FP)، أو ما يعرف عنه الكشف الخاطئ للأصناف. تعد هذه الأخطاء مسألة حساسة بالنسبة لكشف التسلل، لأنه بعد فترة من الزمن سيتم تجاهل أي نظام كشف تسلل يستمر بإطلاق إنذارات كاذبة ولن يتم استخدامه بعد ذلك.

يتيح البحث استخدام الشبكات العصبونية في بناء نظام كشف تسلل ويهدف إلى تطبيق تقنيات وطرق مختلفة لتحسين أداء النظام بتحقيق معدل كشف صحيح عالي للأصناف التسللية وتخفيض معدلات الأخطاء الإيجابية أكثر ما يمكن.

طرائق البحث ومواده:

يرتكز البحث على ثلاث منظومات أساسية هي نظم كشف التسلل Intrusion Detection Systems (IDS)، والشبكات العصبونية Neural Networks (NN)، وقواعد البيانات. تم الاعتماد على بيانات حركة مرور شبكة قياسية لتقييم أنظمة كشف التسلل الشبكية هي مجموعة البيانات المعيارية KDDcup99 التي تضم بيانات طبيعية وتسللية، تم معالجتها معالجة تحضيرية قبل استخدامها. تم استخدام خوارزمية الانتشار الخلفي Back Propagation للشبكات العصبونية في بناء نظام كشف التسلل المقترح، حيث تعد الشبكات العصبونية مشارك جيد في أنظمة كشف التسلل وتحقق دقة عالية في كشف الشذوذ، كما تم اختيار تقنية الشبكات العصبونية بالإضافة إلى عدة تقنيات أخرى كطرق للبحث عن السمات المهمة والمؤثرة على عملية التصنيف، نظراً لكون التعرف على المدخلات المهمة يؤدي مباشرة إلى حجم أقل وتدريب أسرع وربما نتائج أكثر دقة. وبالتالي الوصول إلى أفضل هيكلية ممكنة لمصنف الشبكة العصبونية.

تم تقييم الأداء لنظام كشف التسلل المقترح والمقارنة بالاعتماد على مفهوم مصفوفة الاضطراب التي تعد إحدى أهم الوسائل المستخدمة في عملية تقييم أداء نظم كشف التسلل. تم إجراء التجارب وتدريب الشبكات العصبونية باستخدام برنامج الماتلاب Matlab.

1- أنظمة كشف التسلل Intrusion Detection Systems:

يمكن تعريف كشف التسلل أنه عملية مراقبة الأحداث وتحليلها ضمن نظام معين أو ضمن الشبكة كلها، لكشف السلوك غير الطبيعي ضمن النظام أو الشبكة. ويُعرّف نظام كشف التسلل أنه النظام المسؤول عن كشف التسلل من خلال مجموعة من البرامج والمعدات، وربما اتخاذ الفعل المناسب كإرسال تنبيه أو حذف الرزمة مثلا[4].

تصنف أنظمة كشف التسلل وفقا لمصدر المعلومات إلى أنظمة كشف تسلل معتمدة على الشبكة وأنظمة كشف تسلل معتمدة على المضيف[5]:

- نظام كشف تسلل معتمد على المضيف Host based IDS: يعمل هذا النظام على المعلومات التي يتم جمعها من نظام جهاز مضيف مفرد، حيث يتم توظيف عميل agent ضمن كل جهاز لمراقبته، وغالبا ما تكون مصادر المعلومات من أجل هذا النظام هي ملفات التسجيل التابعة لنظام التشغيل Audit Trials حيث يقوم العميل بفحص مصادر المعلومات من أجل كشف التغييرات غير الموثوقة أو الأنماط المشبوهة لنشاط أو فعل معين والتي يمكن أن تكون تسللا. وهذا يمكن نظام كشف التسلل من تحليل النشاطات والأفعال بدقة ووثوقية.
- نظام كشف التسلل المعتمد على الشبكة Network based IDS: يمثل الصيغة الأكثر شيوعا لأنظمة كشف التسلل المتداولة تجاريا، ويقوم بالتقاط وتحليل الرزم الواردة من أجل الكشف عن الهجمات والتهديدات من خلال مراقبة قطاع الشبكة أو مبدل switch يصل عدة قطاعات شبكية، حيث يقارن الرزم الملتقطة مع مجموعة تواقيع الهجمات المخزنة ضمن مجموعة بيانات، أو يقوم ببناء نماذج لكشف السلوك الشاذ ضمن الشبكة الحاسوبية، وذلك حسب طريقة الكشف المتبعة.
- نظام كشف التسلل الهجين Hybrid based IDS: يتم في هذه الأنظمة استخدام كلا تقنيات الكشف المعتمدة على المضيف والمعتمدة على الشبكة. هذا النوع من الأنظمة مفيد مع حركات مرور معينة وشروط ومتطلبات معينة لأنها توفر مرونة وأمن أكثر للنظام.

تصنف أنظمة كشف التسلل وفقا لطريقة الكشف إلى نظام كشف التسلل المعتمد على التواقيع أو القواعد المسبقة التعريف ونظام كشف التسلل المعتمد على الشذوذ[6]، كما يلي:

- أنظمة كشف التسلل المعتمدة على التواقيع Signature based Detection Systems: تدعى أيضا أنظمة كشف سوء الاستخدام Misuse IDS وأنظمة كشف التسلل المعتمدة على القواعد المسبقة التعريف Rule based IDS لأنها تعتمد على استخدام عينات هجمات معروفة مسبقا، أو نقاط ضعف ضمن النظام، تكون مخزنة ضمن مجموعة بيانات لمطابقتها مع الرزم الواردة إلى الشبكة، وبالتالي كشف الهجمات.
- أنظمة كشف التسلل المعتمدة على الشذوذ Anomaly based Detection Systems: تقوم ببناء ملف يتضمن عينات تصف السلوك العادي للنظام بالاعتماد على القياسات الإحصائية لمعايير النظام، مثل نشاطات المعالج، والدخل والخرج من قبل مستخدم أو برنامج محدد، ثم اعتبار أي انحراف عن هذا السلوك كدليل على وجود هجمات على النظام.

يعتمد هذا البحث على نظام كشف تسلل شبكي NIDS حيث تعد أنظمة محمية بشكل جيد بسبب طبيعة عملها في النمط الخفي مما يجعل المهاجم يواجه صعوبة في تحديد موقع نظام كشف التسلل، كما تكفي أعداد قليلة متمركزة في أماكن مناسبة من هذه الأنظمة لمراقبة شبكة كبيرة الحجم، ويعتمد تقنية الكشف المعتمد على الشذوذ كما هي الحال بالنسبة لمعظم أبحاث كشف التسلل [8][7]، وذلك لعدة أسباب:

- تكمن قوة أنظمة كشف التسلل هذه في كشف الشذوذ، أي لا يحتاج النظام أن يعتمد على التوقيع قبل أن يكشف الهجوم. وبالتالي قدرتها على كشف الهجمات الجديدة غير المعروفة وغير المضمّنة في بيانات التدريب.
- من جهة أخرى فيما يتعلق بنظم كشف سوء الاستخدام، من الصعب إنشاء مجموعة بيانات مفصلة لكل هجوم وبالتالي هناك بعض الهجمات التي لن يتم التعرف عليها.
- بالإضافة لذلك في نظم كشف سوء الاستخدام عند وجود أي اختلاف ولو كان بسيط جدا بين التوقيع الموجود في مجموعة البيانات والهجوم المعروف لن يستطيع نظام كشف التسلل المعتمد على سوء الاستخدام التعرف على هذا الهجوم تماما.

2- الشبكات العصبونية Neural Networks:

ازداد عدد الأبحاث التي تناولت تطبيق الشبكات العصبونية في أنظمة كشف التسلل المعتمدة على الشذوذ مؤخرا [10] [9]. تستطيع هذه الشبكات تجاوز العديد من المشاكل التي تعاني منها نظم كشف التسلل المعتمدة على القواعد. ويمكن استخدام الشبكات العصبونية في العديد من تطبيقات صنع القرار[11]، نظرا لقدرتها على التعلم من الأمثلة، وعلى أخذ القرار في الحالات التي لم تحدث في مجموعة التدريب بكفاءة.

تعتبر الشبكات العصبونية ذات التغذية الأمامية واحدة من أهم الطرق الحديثة التي لها كفاءة عالية في إعطاء نتائج مرضية وجيدة في التعرف، تم في هذا البحث استخدام الشبكة العصبونية ذات التغذية الأمامية والانتشار الخلفي للخطأ، وفيها يكون اتجاه الإشارات الداخلة في الشبكة دوماً إلى الأمام، وبذلك تكون الإشارة الخارجة من أي عصبون تعتمد على الإشارات الداخلة فقط، وتحتاج الشبكات العصبونية ذات التغذية الأمامية إلى وجود زوجين من المتجهات هما متجه الإدخال ومتجه الإخراج المطلوب، تبدأ عملية التدريب بمتجه الإدخال حيث يطبق على الشبكة فينتج الإخراج الحقيقي، ويقارن مع ما يقابله من متجه الإخراج المتوقع والفرق بينهما يمثل الخطأ الذي يستخدم لتعديل الأوزان طبقاً لخوارزمية التعليم، ويستمر التدريب إلى أن يصل الخطأ إلى أقل ما يمكن [12].

2-1 البنية المقترحة للشبكة العصبونية:

إن دخل مصنف الشبكة العصبونية هو شعاع السمات الخاص بسجل الاتصال، وخرجها هو نتيجة التصنيف لسجل الاتصال. تم في هذا البحث العمل على تخفيض عدد السمات المشاركة في عملية التصنيف على عدة مراحل ليتم على أساس ذلك بناء مصنف الشبكة العصبونية للنظام، حيث تمثل عصبونات الدخل عدد السمات النهائية المختارة. توفر طبقة الخرج جواب الشبكة، تم اعتماد تصنيف سجلات الاتصال سواء كانت طبيعية أم هجوم منتمية لواحد من أنواع الهجومات الرئيسية الأربعة، وبالتالي تم تصميم الشبكة العصبونية ليكون خرجها أحد قيم خمس مخارج (1,2,3,4,5) الذي يتوافق على التوالي مع التصنيفات الخمسة التالية (DoS,Probe,R2L,U2R, Normal). بالنسبة للطبقات المخفية وعدد العصبونات ضمنها تم اعتماداً على القيام بعدة تجارب واعتمادا على النتائج التي خلصت إليها الدراسات التي اعتمدت الشبكات العصبونية كمصنفات استخدام طبقة مخفية واحدة مع 1000 عصبون مخفي. تابع النقل هو تابع sigmoid في كلا الطبقة المخفية وطبقة الخرج، بينما تابع التدريب هو تابع خوارزمية الانتشار الخلفي 'trainscg'.

3- مجموعة البيانات Dataset:

تلعب مجموعة البيانات دورا هاما في عملية الاختبار والتحقق من صحة طرق كشف التسلل ضمن الشبكة أو الأنظمة. يمكن توليد البيانات الشبكية باستخدام طريقتين رئيسيتين، الأولى بالتقاط بيانات شبكية حقيقية والثانية عن

طريق توليد بيانات باستخدام المحاكاة. توفر مجموعات بيانات التسلل فرصة لتحليل متعمق للنماذج، لسلوكيات التسللات واختبار خوارزميات كشف التسلل. يمكن تصنيف مجموعات بيانات كشف التسلل هذه بشكل أساسي إلى [17]:

1. مجموعة بيانات شبكية حقيقية Real-life Dataset: تضم مجموعة البيانات الشبكية الحقيقية بيانات تم التقاطها من شبكات حقيقية خلال عدة أيام. تضم البيانات الملتقطة سلوكيات طبيعية وشاذة للشبكة. من مجموعات البيانات الشبكية الحقيقية المشهورة مجموعة بيانات UNIBS [20] ومجموعة بيانات ISCX-UNB [21].

2. مجموعة بيانات قياسية Benchmark Dataset: يتم توليد مجموعة البيانات القياسية عن طريق محاكاة بيئة ضمن شبكة ضخمة. يتم ضمن هذه البيئة توليد سيناريوهات تسللية مختلفة. من مجموعات البيانات القياسية الأكثر شهرة مجموعة بيانات KDDCup99 [2][1]، مجموعة بيانات NSL-KDD [22]، مجموعة بيانات DARPA [23]، مجموعة بيانات UNSW-NB15 [24]، مجموعة بيانات ICSI\LBNL [25]، ومجموعة بيانات CAIDA [26].

3. مجموعة بيانات تركيبية Synthetic Dataset: يتم توليد مجموعة البيانات التركيبية لتلبية متطلبات معينة. عادة ما يتم استخدامها للتقييم النظري للنموذج الأولي للنظام.

تمتاز مجموعة البيانات KDD بمايلي:

- توافريتها حيث يوجد العديد من المجموعات تكون مجهولة المصدر يعود ذلك لمخاطر أمنية محتملة للمنظمة. وبالتالي عدم إمكانية الوصول إليها من قبل الباحثين مثل DARPA و CAIDA وغيرها.
- سمات حركة المرور المفصلة التي تفتقر إليها مجموعات أخرى مثل LBNL. ومن الجدير بالذكر أن دراسة سمات حركة المرور هي إحدى الركائز التي يعتمد عليها هذا البحث.
- استخدامها الكبير من قبل الباحثين في مجال كشف التسلل مقارنة مع غيرها من المجموعات مما يتيح الفرصة لمتابعة نتائج الأبحاث ومقارنتها.

3-1- مجموعة بيانات KDDCup99:

تحت رعاية DARPA (Defense Advanced Research Project Agency, US) (وكالة مشروع البحث المتقدم والدفاع، الولايات المتحدة الأميركية) ومخبر أبحاث سلاح الجو، قام مخبر لينكولن في معهد ماساتشوستس للتكنولوجيا بتوليد بيانات حركة مرور شبكة قياسية لتقييم أنظمة كشف التسلل الشبكية [1]. تم إجراء هذه الجهود بين عامي 1998 و1999. كان الهدف استعراض وتقييم الأنشطة البحثية في مجال كشف التسلل.

تم تأمين مجموعة من البيانات القياسية للتدقيق مكونة من مجموعة متنوعة من التسللات التي تمت محاكاتها في بيئة شبكة عسكرية. تم الحصول على بيانات التدريب وهي عبارة عن 4 غيغابايت من بيانات tcpdump من حركة مرور الشبكة خلال سبع أسابيع. نتج عن العملية كاملة حوالي خمس ملايين سجل اتصال. وبالمثل، نتج عن أسبوعي بيانات الاختبار حوالي مليوني سجل اتصال. في عام 1999، اعترفت KDD (Knowledge Discovery and Data mining) (منظمة التنقيب عن البيانات والكشف المعرفي) والتي تعد المنظمة الاحترافية الأكثر شعبية لمنقبي البيانات ووافقت على بيانات DARPA بأن تكون كعلامة تقليدية لمجموعة بيانات من أجل أنظمة كشف التسلل IDS وتم تسميتها KDDCup99 أو KDD99 [14][2]. تضم KDD99 41 سمة وتضم بيانات طبيعية و22 نوع مختلف من الهجومات تقع تحت أربع تصنيفات من الهجومات DOS, PROBE, R2L,U2R [3].

3-2- سمات مجموعة البيانات KDD Features:

في بيانات KDDCUP99، تتضمن السمات البدائية المستخرجة من تسجيل الاتصال السمات الأساسية Basic Features لاتصال TCP فردي، مثل : مدته، نوع البروتوكول، عدد البايتات المنقولة والعلم الذي يشير إلى الحالة الطبيعية أو حالة الخطأ للاتصال. توفر هذه السمات معلومات لأغراض تحليل حركة مرور الشبكة العامة [2]. بالإضافة إلى سمات "نفس المضيف" "same host" التي تفحص فقط الاتصالات في الثانيتين السابقتين التي لها نفس المضيف الوجهة كاتصال حالي، وتحسب الاحصائيات المتعلقة بسلوك البروتوكول، الخدمة، إلخ. وسمات "نفس الخدمة" "same service" المشابهة التي تفحص فقط الاتصالات في الثانيتين السابقتين التي لها نفس الخدمة كاتصال حالي. تسمى سمات "نفس المضيف" و"نفس الحدمة" مع بعضها السمات المعتمدة على الزمن Time-based Features من سجلات الاتصال.

إن بعض هجمات التحقق تفحص مضيفين (أو منافذ) مستخدمة فترات زمنية أكبر من ثانيتين، على سبيل المثال خلال دقيقة. لذلك، تم تخزين سجلات الاتصال بواسطة المضيف الوجهة، وتم إنشاء السمات باستخدام نافذة لمئة اتصال لنفس المضيف بدلا من النافذة الزمنية. نتج عن هذا مجموعة سمات سميت السمات المعتمدة على المضيف Host-based Features.

من جهة أخرى وعلى عكس هجمات DOS والتحقق، لا يبدو هناك أي نماذج تسلسلية (متكررة) في سجلات هجمات R2L و U2R. هذا يعود لكون هجمات DoS والتحقق تتضمن العديد من الاتصالات لنفس المضيف (المضيفين) في فترة زمنية قصيرة، لكن هجمات U2R و R2L تكون مضمنة في أجزاء بيانات الحزم، وعادة ما تتضمن اتصال مفرد. لذلك تم استخدام معرفة المجال لإضافة سمات تبحث عن السلوك المشبوه في أجزاء البيانات، مثل عدد محاولات تسجيل الدخول الفاشلة. تدعى هذه السمات بسمات المحتوى Content Features.

بشكل عام، تتكون مجموعة بيانات التدريب KDD مما يقارب 4,900,000 سجل اتصال مفرد، وهناك 42 سمة (بما في ذلك نوع الهجوم) في كل سجل اتصال.

سجل طبيعي:

0,tcp,http,SF,239,486,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,0.00,19,19,1.00,0.00,0.05,0.00,0.00,0.00,0.00,0.00,normal.

سجل هجوم:

0,tcp,private,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,292,18,1.00,1.00,0.00,0.00,0.06,0.05,0.00,255,18,0.07,0.06,0.00,0.00,1.00,1.00,0.00,0.00,neptune.

تدل السمة الأخيرة لسجل الاتصال ذات الرقم 42 على نوع سجل الاتصال، كما في المثال السابق Normal سجل ذو بيانات طبيعية، وNeptune سجل ذو بيانات تسللية من النوع Neptune.

3-3- أنواع الهجومات في مجموعة البيانات KDD Attack Types:

تم تضمين 22 نوع هجوم في بيانات التدريب وتم إضافة 17 هجوم جديد لبيانات الاختبار [16]، لاختبار أداء IDS على الهجومات المعروفة وغير المعروفة. كل نوع من أنواع الهجومات يقع في واحدة من الفئات الرئيسية الأربعة التالية:

1. هجوم حجب الخدمة Denial of Service (DOS): هو محاولة جعل الجهاز أو مصدر الشبكة غير متاح لمستخدميه. على الرغم من تنوع أساليب تنفيذ هجوم حجب الخدمة، إلا أنها غالبا ما تتكون من الجهود التي تحاول تعليق أو التوقيف المؤقت أو المستمر لخدمات المضيف المتصل بالإنترنت. يمكن لهجمات حجب الخدمة التعطيل الكامل للحاسب أو الشبكة، مثل smurf، teardrop.

2. هجوم مستخدم إلى جذر User to Root (U2R): هذه الهجمات هي استغلال المهاجم للنظام الذي يبدأ به من حساب مستخدم طبيعي ويحاول من خلاله الاعتداء على نقاط ضعف في النظام بهدف الحصول على امتيازات مستخدم رئيسي مثل perl، xterm.

3. هجوم بعيد إلى محلي Remote to Local (R2L): هو هجوم يقوم به المستخدم بإرسال حزم عبر الانترنت إلى جهاز، لا يملك الوصول إليه بهدف كشف نقاط ضعف الأجهزة واستغلال الميزات التي يملكها مستخدم محلي على الجهاز مثل xlock، guest، xnsnoop، phf، sendmail.

4. هجوم التحقق Probe: هو الهجوم الذي يمسح فيه المهاجم جهاز أو جهاز شبكي بهدف تحديد نقاط الضعف التي يمكن استغلالها فيما بعد وبالتالي تعريض النظام للخطر. هذه التقنية مستخدمة كثيرا في التنقيب عن البيانات، مثل : saint، portsweep، mscan، nmap.

3-4- مجموعة بيانات التدريب Training Dataset:

تم استخدام مجموعة بيانات "10% KDD" من أجل تدريب أنظمة كشف التسلل المختلفة. وهي نسخة مختصرة من مجموعة البيانات KDD الكاملة، تضم 494021 سجل اتصال. تحتوي هذه القاعدة 22 نوع هجوم، وتحوي أمثلة من اتصالات الهجمات أكثر مما تحويه من الاتصالات الطبيعية، حيث أن أنواع الهجمات غير ممثلة على التساوي، كما يمثل هجوم حجب الخدمة غالبية سجلات الاتصال في مجموعة البيانات.

3-5- مجموعة بيانات الاختبار Testing Dataset:

يتم استخدام مجموعة بيانات "Corrected KDD" لغرض الاختبار. توفر مجموعة البيانات "Corrected KDD" بيانات ذات توزع إحصائي مختلف مقارنة مع مجموعة البيانات التدريب سواء كانت "10% KDD" الجزئية أو KDD الكاملة كما هو مبين في الجدول (1)، تضم 311029 سجل اتصال معنون بطبيعي أو أحد أنواع الهجمات الأربعة. يتم استخدام عناوين سجلات الاتصال للتحقق من توقعات التصنيف التي تم الإدلاء بها أثناء الاختبار.

الجدول (1) عدد العينات في كل من مجموعة البيانات الكاملة و الجزئية والاختبار.

Class	The whole KDD		10%KDD		Corrected KDD	
	عدد السجلات	النسبة المئوية	عدد السجلات	النسبة المئوية	عدد السجلات	النسبة المئوية
Normal	972781	19.8590%	97278	19.6911%	60593	19.4815%
Dos	3883366	79.2778%	391458	79.2391%	229853	73.9008%
Probe	41102	0.8391%	4107	0.8313%	4166	1.3394%
R2L	1126	0.0230%	1126	0.2279%	16347	5.2558%
U2R	52	0.0011%	52	0.0105%	70	0.0225%
مجموع الهجمات	3925646	80.1409%	396743	80.3089%	250436	80.5185%
مجموع السجلات	4898427	100%	494021	100%	311029	100%

يبين الجدول (1) مقارنة بين مجموعة البيانات "KDD99" الكاملة والجزئية "10%KDD" ومجموعة بيانات الاختبار "Corrected" المستخدمة في مرحلة الاختبارات العملية للتحقق من أداء النظام من حيث عدد سجلات كل صنف من الأصناف الخمسة الموجودة ضمن كل مجموعة بيانات، والنسبة المئوية لسجلات كل صنف.

4- المعالجة التحضيرية للبيانات Data Preprocessing:

يمكن النظر إلى الحاجة إلى المعالجة التحضيرية للبيانات من حقيقة كون البيانات المكررة والسمات غير الهامة غالبا ما تشوش خوارزمية التصنيف، مما يؤدي إلى كشف معرفي غير دقيق أو غير ناجح. بالإضافة لذلك، زمن المعالجة سيزداد عندما يتم استخدام جميع السمات. أخيرا، تفيد المعالجة التحضيرية في إزالة البيانات المكررة، والبيانات غير الكاملة، وتحويل البيانات إلى شكل موحد. علما أن المعالجة التحضيرية تتم قبل تدريب الشبكة العصبونية، أي لا تتم أثناء مرحلة التدريب أو الاختبار، وبالتالي لن يزداد زمن المعالجة لنظام كشف التسلل.

تضم مرحلة المعالجة التحضيرية كل من الخطوات التالية: (1) حذف السجلات المتكررة، (2) تحويل البيانات الرمزية إلى قيم رقمية، (3) وتخفيض السمات.

4-1- حذف السجلات المتكررة :

القيد الأساسي في مجموعة بيانات KDD99 هو وجود السجلات المكررة، لأن ذلك يسبب تحيز خوارزمية التعلم لهذه السجلات وبالتالي الحصول على نتائج غير صحيحة. لذلك من الضروري القيام بهذه الخطوة لضمان عدم تحيز خوارزمية التعليم أولا وبالتالي تحسين معدل الكشف، وبهدف التخفيض من متطلبات مساحة التخزين ثانيا، حيث إنّ حذف السجلات المكررة يفيد في التقليل من كمية البيانات التي تتم معالجتها من قبل الـIDS، وبالتالي الحد من زمن المعالجة.

4-2- تحويل البيانات الرمزية إلى بيانات رقيمة:

يتم تحويل السمات (41 سمة) في KDD99 إلى تمثيل رقمي موحد. طالما هناك سمات بقيم رمزية غير رقمية متوفرة في KDD99، فإنه يتم تحويل هذه القيم إلى واحدة رقمية، نظرا لكون الشبكة العصبونية لا تقبل إلا دخلا بقيم رقيمة.

4-3- تخفيض السمات Feature Reduction :

يمكن لبعض البيانات أن تعيق عملية التصنيف. يمكن للسمات أن تحوي على ارتباطات خاطئة، مما يعيق عملية كشف التسللات. علاوة على ذلك، يمكن لبعض البيانات أن تكون مكررة نظرا لكون المعلومات التي تضيفها موجودة مسبقا في غيرها من السمات. يمكن للسمات الإضافية أن تزيد من زمن الحساب، ويمكنها التأثير على دقة IDS. لا يوجد أي تصميم أو تابع يجسد العلاقة بين السمات المختلفة أو بين الهجومات المختلفة والسمات. لو كان مثل هذا التصميم موجود فعلا، لكانت عملية كشف التسلل عملية بسيطة ومباشرة. تم في هذا البحث الاعتماد على هندسة السمات وتقنية الشبكات العصبونية لهذا الغرض.

4-3-1- هندسة السمات Feature Engineering :

تفيد هندسة السمات في تخفيض كمية البيانات، عن طريق اختيار السمات المفيدة فقط، وبالتالي تجنب الحسابات غير المفيدة على السمات غير الهامة، وتحسين الدقة، حيث إزالة السمات التي تحوي معلومات مكررة يزيد التركيز على غيرها من السمات من ناحية أخرى. تم تطبيق تقنيتين على السمات:

- إيجاد السمات غير المفيدة Finding Useless Features: عن طريق حساب قيم التباين variance لقيم كل سمة من السمات على حدة، ثم دراسة هذه القيم بهدف الاستغناء عن السمات ذات التباينات المعدومة.
- إيجاد السمات المرتبطة Finding Correlated Features: بمعنى آخر إيجاد السمات ذات الارتباط الأعلى، بهدف الاستغناء عن سمة واحدة من كل زوج منها، نظرا لكون تأثير إحداها يعوض عن تأثير الاخرى. تتراوح قيم معامل الارتباط بين أي قيمتين بين [1−0] كللما كانت قيمة معامل الارتباط قريبة من الواحد تعد القيمتين مرتبطتين بشكل كبير، والعكس صحيح كلما كانت قريبة من الصفر تعد القيمتين غير مرتبطتين.

4−3−2− الشبكات العصبونية Neural Network:

تم في هذا البحث استخدام الشبكات العصبونية ليس فقط لتصميم مصنف نظام كشف التسلل، وإنما أيضا في عملية البحث عن السمات المهمة في عملية التصنيف. حيث تم إجراء تجارب بتصميم شبكات عصبونية بعدد السمات المتبقية، تم في كل شبكة حذف واحدة من السمات في كل وقت وإجراء التجارب على الشبكة العصبونية، ودراسة النتائج بحثا عن مدى تأثير حذف هذه السمة على عملية التصنيف وعلى دقة كشف التسلل، وعلى هذا الأساس يتم إما الاستغناء عن هذه السمة أو الإبقاء عليها.

5− مصفوفة الاضطراب Confusion Matrix:

يتم تقييم فعالية IDS من خلال قدرته على التصنيف الصحيح. أي القدرة على تحديد الصنف الذي ينتمي إليه سجل الاتصال طبيعي أم هجومي. عند مقارنة نتيجة تصنيف السجل مع الواقع الفعلي نجد أربع حالات مختلفة يبينها الجدول(2)، الذي يعبر عن مصفوفة الاضطراب التي تعد أهم الوسائل المستخدمة في عملية تقييم أداء IDS.

الجدول (2) مصفوفة الاضطراب[18] .

	تم التنبؤ أن الحدث إيجابي Predicted Positive	تم التنبؤ أن الحدث سلبي Predicted Negative
الحدث بالفعل إيجابي Actual positive	TP	FN
الحدث بالفعل سلبي Actual Negative	FP	TN

يبين الجدول (2) أربع قيم مهمة [19] لا بد من التطرق إليها عند تقييم أداء وفعالية نظم كشف التسلل، هي:

- الإيجابيات الصحيحة True Positives (TP): أو ما يعرف عنه الكشف الصحيح، عدد الحالات التي يتم فيها التنبؤ أن الحدث x و الحدث بالفعل x.
- السلبيات الخاطئة False Negatives (FN): أو ما يسمى بالأخطاء السلبية، هو عدد الحالات التي يتم فيها التنبؤ أن الحدث ليس x رغم أن الحدث x.
- الإيجابيات الخاطئة False Positives (FP): أو ما يسمى الأخطاء الإيجابية، أو ما يعرف عنه الكشف الخاطئ، هو عدد الحالات التي يتم فيها التنبؤ أن الحدث x لكنه ليس x.
- السلبيات الصحيحة True Negatives (TN): عدد الحالات التي يتم فيها التنبؤ أن الحدث ليس x والحدث فعلا ليس x.

من خلال مصفوفة الاضطراب والقيم التي تم شرحها في الأعلى، غالبا ما يتم تقييم أداء نظم كشف التسلل بالاعتماد على القيم التالية [18]:

- معدل الإيجابيات الصحيحة True Positive Rate (TPR): أو ما يطلق عليه معدل الكشف Detection Rate (DR)، أو الحساسية Sensitivity، عدد الإيجابيات الصحيحة على جميع المخرجات الإيجابية (دقة الحالات الإيجابية)، تعطى بالعلاقة التالية:

$$DR = Sensitivity = TPR = \frac{TP}{TP+FN} \quad \text{.....(1)}$$

- معدل السلبيات الصحيحة True Negative Rate (TNR): أو ما يطلق عليه التحديد Specificity، عدد الإيجابيات السلبية على جميع المخرجات السلبية (دقة الحالات السلبية) يعطى بالعلاقة:

$$TNR = Specificity = \frac{TN}{TN+FP} \quad \text{.....(2)}$$

- معدل الأخطاء الإيجابية False Positive Rate (FPR): عدد الإيجابيات الخاطئة على جميع المخرجات الإيجابية، يعطى بالعلاقة:

$$FPR = \frac{FP}{TN+FP} \quad \text{.....(3)}$$

- معدل الأخطاء السلبية False Negative Rate (FNR): عدد السلبيات الخاطئة على جميع المخرجات السلبية، ويعطى بالعلاقة:

$$FNR = \frac{FN}{TP+FN} \quad \text{.....(4)}$$

- الدقة Accuracy: تعد المقياس الأكثر شيوعا لتقييم المصنف، حيث تقيم الفعالية الكاملة للخوارزمية عن طريق تقدير احتمالية القيمة الصحيحة لعنوان الصنف. وتعطى بالعلاقة:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad \text{.....(5)}$$

يعد كل من معدل الكشف DR و معدل الأخطاء الإيجابية والسلبية والدقة Accuracy المقاييس الأكثر شيوعا عند تقييم أنظمة كشف التسلل.

النتائج والمناقشة:

قبل البدء بمرحلة التجارب المخبرية لا بد من القيام بحذف السجلات المكررة من مجموعة البيانات نظرا لأهمية هذه الخطوة و تأثيرها على نتائج الكشف، تم إجراء تجارب باستخدام مجموعة البيانات من دون حذف السجلات المتكررة وكانت نتائج كشف المصنف الناتج منخفضة بشكل واضح. تم معالجة مجموعة البيانات من خلال عملية بحث ومقارنة للسجلات تم على أساسها تحديد السجلات المتكررة وحذفها. يبين الجدول (3) عدد السجلات الموجودة في مجموعة بيانات التدريب KDD 10% وعدد السجلات التي تم الحصول عليها بعد إزالة التكرارية.

الجدول (3) عدد سجلات كل صنف قبل و بعد إزالة التكرارية.

After Preprocessing		Before Preprocessing		Class	
النسبة المئوية	عدد العينات	النسبة المئوية	عدد العينات		
46.6079%	95,233	19.6911%	97,278	Normal	
0.3029%	619	56.8377%	280,790	Smurf	DOS
50.2388%	102,652	21.6997%	107,201	Neptune	
0.6299%	1,287	0.4459%	2,203	Back	
0.4302%	879	0.1982%	979	Teardrop	
0.1135%	232	0.0534%	264	Pod	

0.0103%	21	0.0043%	21	Land	
51.7257%	105,690	79.2391%	391,458	المجموع	
0.4777%	976	0.3216%	1,589	Satan	Probe
0.3553%	726	0.2524%	1,247	Ipsweep	
0.208%	425	0.2105%	1,040	Portsweep	
0.0773%	158	0.0468%	231	Nmap	
1.1183%	2,285	0.8313%	4,107	المجموع	
0.4708%	962	0.2065%	1,020	Warezclient	R2L
0.0259%	53	0.0107%	53	Guess_passwd	
0.0098%	20	0.004%	20	Warezmaster	
0.0059%	12	0.0024%	12	Imap	
0.0039%	8	0.0016%	8	Ftp_write	
0.0034%	7	0.0014%	7	Multihop	
0.002%	4	0.0008%	4	Phf	
0.001%	2	0.0004%	2	Spy	
0.5227%	1,068	0.2279%	1,126	المجموع	
0.0147%	30	0.0061%	30	Buffer_overflow	U2R
0.0049%	10	0.002%	10	Rootkit	
0.0044%	9	0.0018%	9	Loadmodule	
0.0015%	3	0.0006%	3	Perl	
0.0254%	52	0.0105%	52	المجموع	
100%	204,328	100%	494,021	المجموع	

بعد قراءة الجدول (3) يمكن ملاحظة أن السجلات المكررة موجودة ضمن الصنف الشاذ بشكل أكبر مما هي عليه في الصنف الطبيعي، وبشكل خاص الصنف DoS حيث من الواضح وجود انخفاض كبير في عدد سجلات هجوم DoS بعد إزالة التكرارية عند مقارنتها مع الاصناف الأخرى.

تم بعد هذه الخطوة تحويل السمات (41 سمة) إلى تمثيل رقمي موحد، حيث تتنوع قيم السمات في KDDCup99 بين القيم الرقمية numeric والثنائية binary والمحرفية symbolic. طالما هناك سمات بقيم رمزية غير رقمية ضمن KDDCup99، فإنه يتم تحويل هذه القيم إلى واحدة رقمية، نظرا لكون الشبكة العصبونية لا تقبل إلا دخلا بقيم رقيمة. كل من السمات protocol_type، service، flag، معنونة بأنها سمات محرفية ولكل منها قيم مختلفة. توضح الجداول التالية (4)، (5)، (6) هذه القيم مع الترميز الرقمي المقابل لها الذي تم اتباعه في كلا مجموعتي بيانات التدريب والاختبار.

الجدول (4) الترميز الرقمي لقيم السمة (2)Protocol_type:

Protocol_type(2)	الترميز الرقمي	Protocol_type(2)	الترميز الرقمي	Protocol_type(2)	الترميز الرقمي
Icmp	1	Tcp	2	Udp	3

الجدول (5) الترميز الرقمي لقيم السمة (4)Flag :

Flag(4)	الترميز الرقمي	Flag(4)	الترميز الرقمي	Flag(4)	الترميز الرقمي
ROSTOS0	1	REJ	5	S3	9
RSTR	2	S0	6	SF	10
RSTO	3	S1	7	SH	11
OTH	4	S2	8		

الجدول (6) الترميز الرقمي لقيم السمة (3)Service:

Service(3)	الترميز الرقمي	Service(3)	الترميز الرقمي	Service(3)	الترميز الرقمي
Netbios_dgm	1	Gopher	23	Whois	45
Netbios_ssn	2	Domain	24	Time	46
Netbios_ns	3	Finger	25	Echo	47
Remote_job	4	Klogin	26	Idap	48
Hostnames	5	Kshell	27	Link	49
Uucp_path	6	Supdup	28	Http	50
Iso_tsap	7	Systat	29	Smtp	51
Csnet_ns	8	Telnet	30	Uucp	52
Domain_u	9	Shell	31	Auth	53
Ftp_data	10	Imap4	32	Nnsp	54
Http_443	11	Eco_i	33	Nntp	55
Daytime	12	Ecr_i	34	Name	56
Discard	13	Red_i	35	Exec	57
Netstat	14	Pop_2	36	Mtp	58
Courier	15	Pop_3	37	Rje	59
Pm_dump	16	Login	38	Ssh	60
Printer	17	Tim_i	39	Ftp	61
Private	18	Urh_i	40	Irc	62
Sql_net	19	Urp_i	41	X11	63
Tftp_u	20	Ntp_u	42	Bgp	64
Sunrpc	21	Vmnet	43	Ctf	65
Z39_50	22	Other	44	Efs	66

أيضا كمتابعة لعملية الترميز الرقمي، يضم العمود الأخير ذو الرقم 42 من كل سجل اتصال نوع سجل الاتصال سواء كان طبيعي أو هجوم تبعا لنوع الهجوم المحدد الذي ينتمي إليه، يبين الجدول (7) الترميز الرقمي لأنواع الهجومات كما تم الشرح في تصميم مصنف الشبكة العصبونية.

الجدول (7) الترميز الرقمي لأنواع الهجومات :

الصنف		الترميز الرقمي	الصنف		الترميز الرقمي
Dos	Smurf	1	R2L	Ftp_write	3
	Neptune	1		Multihop	3
	Back	1		Phf	3
	Teardrop	1		Spy	3
	Pod	1		httptunnel	3
	Land	1		Worm	3
	apache2	1		Xlock	3
	Mailbomb	1		Xsnoop	3
	Processtable	1		Named	3
	Udpstorm	1		Sendmail	3
Probe	Satan	2	U2R	Loadmodule	4
	Ipsweep	2		Buffer_overflow	4
	Portsweep	2		Perl	4
	Nmap	2		Rootkit	4
	Mscan	2		snmpgetattak	4
	Saint	2		snmpguess	4
R2L	Warzclient	3		Sqlattack	4
	Guess_passwd	3		Xterm	4
	Warzemaster	3		ps	4
	Imap	3	Normal		5

بعد القيام بالخطوتين السابقتين تم تنفيذ أربع تجارب أساسية ضمن مرحلة الاختبارات العملية، في كل تجربة هناك مصنف شبكة عصبونية، تختلف هذه المصنفات بعدد السمات المشاركة في بناء هيكلية المصنف وبالتالي المشاركة في عملية التصنيف.

- **التجربة الأولى :** تم تدريب مصنف الشبكة العصبونية على مجموعة السمات الكاملة (41 سمة)، وعرض النتائج من دقة المصنف ومعدل كشف التسلل للنظام ومعدلات الأخطاء ومعدل كشف كل صنف من الأصناف الخمسة على حدة كما هو مبين لاحقا في الجدول (11) بهدف المقارنة عند كل عملية تغيير في عدد سمات الدخل.
- **التجربة الثانية :** تم في هذه التجربة حساب قيم تباينات السمات الموضحة في الجدول (8)، باستخدام تعليمة var(i) ضمن الماتلاب حيث i هي العمود المعبر عن السمة المختارة في كل مرة، ودراسة هذه القيم بحثا عن السمات ذات التباينات الصفرية، عندما تكون قيمة التباين صفر يدل ذلك على عدم تغير قيمة السمة المقصودة بالنسبة لجميع سجلات الاتصال سواء كانت طبيعية أو تسللية، وبالتالي عدم مساهمتها في عملية التصنيف، لذلك من الممكن الاستغناء عنها.

الجدول (8) قيم التباينات لكل سمة من سمات مجموعة البيانات.

الرقم	السمة	التباين	الرقم	السمة	التباين
1	Duration	1188700	23	Count	11513
5	Src_bytes	2361100000000	24	Srv_count	667.119
6	Dst_bytes	2636600000	25	Serror_rate	0.24
7	Land	0.00010766	26	Srv_serror_rate	0.2405
8	Wrong_fragment	0.0393	27	Rerror_rate	0.1135
9	Urgent	0.000073411	28	Sev_rerror_rate	0.1135
10	Hot	1.4583	29	Same_srv_rate	0.2165
11	Num_failed_logins	0.00058227	30	Did_srv_rate	0.0115
12	Loggrd_in	0.2285	31	Srv_diff_host_rate	0.0443
13	Num_compromised	7.8144	32	Dst_host_count	8269.3
14	Root_shell	0.0002691	33	Dst_host_srv_count	12644
15	Su_attempted	0.00014682	34	Dst_host_same_srv_rate	0.2062
16	Num_root	9.7941	35	Dst_host_diff_srv_rate	0.0197
17	Num_file_creations	0.0225	36	Dst_host_sane_src_port_rate	0.0512
18	Num_shells	0.00029358	37	Dst_host_srv_diff_host_rate	0.0031
19	Num_access_files	0.0032	38	Dst_host_serror_rate	0.2397
20	Num_outbound_cmds	0	39	Dst_host_srv_serror_rate	0.4205
21	Is_hot_login	0	40	Dst_host_rerror_rate	0.1121
22	Is_guest_login	0.0033	41	Dst_host_srv_rerror_rate	0.1114

بعد دراسة الجدول (8) تمت ملاحظة القيم الصفرية لكل من السمتين (20)num_outband_cmds، (21)is_hot_logins، لذلك تم ضمن هذه التجربة حذفهما من مجموعة البيانات لافتراض عدم مساهمتهما في عملية كشف التسلل. على أساس ما سبق تم إجراء التجربة الثانية باستخدام 39 سمة كأشعة دخل عند بناء الشبكة العصبونية الثانية، يبين الجدول (11) نتائج المصنف الثاني من حيث دقة النظام و دقة كشف كل صنف بالإضافة للأخطاء الإيجابية لكل صنف.

- **التجربة الثالثة** : تم في هذه التجربة حساب قيم معامل الارتباط Correlation Coefficient بين كل زوج من السمات الناتجة عن التجربة السابقة (39 سمة)، باستخدام تعليمة الماتلاب corrcoef(A,B) حيث A و B هي السمات التي يتم حساب الترابط فيما بينها. وذلك بحثا عن السمات المرتبطة بقوة، أي ذات قيم معامل الارتباط القريبة من الواحد، 0.9 أو أكثر. بهدف الاستغناء عن واحدة من سمات كل زوج مرتبط بقوة، نظرا لكون تأثير إحداها يمثل تأثير الأخرى، وبالتالي تخفيض عدد السمات المشاركة في عملة التصنيف دون التأثير سلبا على دقة النظام. يبين الجدول (9) أزواج السمات ذات الارتباط القوي.

الجدول (9) أزواج السمات ذات الارتباط القوي.

معامل الارتباط	السمات المرتبطة			
0.9941	num_root	16	num_compromised	13
0.9977	sev_serror_rate	26	serror_rate	25
0.9981	dst_host_serror_rate	38	serror_rate	25
0.997	dst_host_srv_serror_rate	39	serror_rate	25
0.997	dst_host_serror_rate	38	sev_serror_rate	26
0,999	dst_host_srv_serror_rate	39	sev_serror_rate	26
0.9943	sev_rerror_rate	28	rerror_rate	27
0.9852	dst_host_rerror_rate	40	rerror_rate	27
0.9834	dst_host_srv_rerror_rate	41	rerror_rate	27
0.98	dst_host_rerror_rate	40	sev_rerror_rate	28
0.9847	dst_host_srv_rerror_rate	41	sev_rerror_rate	28
0.9491	dst_host_same_srv_rate	34	dst_host_srv_count	33
0.9974	dst_host_srv_serror_rate	39	dst_host_serror_rate	38
0.9828	dst_host_srv_rerror_rate	41	dst_host_rerror_rate	40

يعرض الجدول (9) أزواج السمات ذات الارتباط القوي فقط أي التي كانت قيمة الترابط فيما بينها أكبر من 0.9 حيث تتراوح قيمة الترابط بشكل عام بين الصفر والواحد كما تم ذكره سابقا، كلما كانت أقرب من الواحد كان الترابط أقوى. تم على أساس ذلك الإبقاء على سمة وحيدة من كل زوج مترابط لأن تأثير إحداها يمثل تأثير الأخرى. على سبيل المثال السطر الأول من الجدول يمثل الارتباط بين السمتين (13) و(16) تم الاستغناء عن كل سمة من هذا الزوج على حدة ومقارنة النتائج عند غياب كل منها لنلاحظ أن النتائج متماثلة أي وجود واحدة من السمتين يعوض عن وجود السمة الأخرى بالتالي تم اختيار إحدى السمتين. بنفس الطريقة بالنسبة للسطر الثاني من الجدول نلاحظ أن الارتباط بين السمتين (25) و (26) قوي وكل من هاتين السمتين مترابطة بقوة مع مجموعة السمات ذاتها كما هو موضح في الجدول بالتالي يكفي الإبقاء على السمة (25) أو السمة (26) والاستغناء عن باقي السمات المرتبطة معها. على هذا الأساس تم حذف السمات: num_compromised(13), sev_serror_rate(26), rerror_rate(27), dst_host_same_srv_rate(34), dst_host_serror_rate(38), dst_host_srv_serror_rate(39), dst_host_rerror_rate(40), dst_host_srv_rerror_rate(41)، والإبقاء على السمات : num_root(16), serror_rate(25), sev_rerror_rate(28), dst_host_srv_count(33).

تم على أثر ذلك استخدام 31 سمة كأشعة دخل لمصنف الشبكة العصبونية الثالثة واختبارها، يعرض الجدول (11) نتائج المصنف الناتج.

- **التجربة الرابعة** : تم في هذه التجربة الاعتماد على تقنية الشبكات العصبونية لدراسة تأثير كل سمة بشكل منفرد على عملية الكشف. تم القيام بـ 31 تجربة فرعية ضمن هذه التجربة في كل مرة تم حذف واحدة من السمات على حدة وتدريب الشبكة العصبونية ودراسة مدى تأثير عدم وجود هذه السمة على عملية الكشف. نتيجة لما سبق تم حذف السمة رقم 12 السمة "Loggrd_in" بعد ملاحظة أن نتائج التصنيف تتحسن عند الاستغناء عنها. ثم تم تدريب المصنف الرابع والأخير على 30 سمة، وحساب نتائج النظام من دقة و معدل كشف و معدلات أخطاء كما في التجارب السابقة كما هو موضح في الجدول (11).

يبين الجدول (10) السمات النهائية (30 سمة) المستخدمة في بناء مصنف الشبكة العصبونية للنظام المقترح:

الجدول (10) وصف لمجموعة السمات النهائية المشاركة في عملية التصنيف

No	اسم السمة Feature Name	الوصف Description
1	Duration	عدد الثواني في الاتصال
2	Protocol_type	نوع البروتوكول، مثل ....,TCP,UDP
3	Service	خدمة الشبكة في الوجهة، مثل ...,http,telnet
4	Flag	حالة الاتصال : طبيعي أو خاطئ
5	Src_bytes	عدد بايتات البيانات من المصدر إلى الوجهة
6	Dst_bytes	عدد بايتات البيانات من الوجهة إلى المصدر
7	Land	1: الاتصال من\إلى نفس المضيف\المنفذ. 0: غير ذلك
8	Wrong_fragment	عدد التجزئات "الخاطئة"
9	Urgent	عدد الحزم الطارئة (المستعجلة)
10	Hot	عدد الدخولات إلى أدلة النظام، إنشاء وتنفيذ البرامج
11	Num_failed_logins	عدد محاولات الدخول الفاشلة
14	Root_shell	1: تم الحصول على الـ root shell. 0: غير ذلك
15	Su_attempted	1: تمت محاولة الأمر 'su root'. 0: غير ذلك
16	Num_root	عدد وصولات الـ 'root'
17	Num_file_creations	عدد عمليات إنشاء الملف
18	Num_shells	عدد مطالبات الـ shell
19	Num_access_files	عدد عمليات الكتابة والحذف والإنشاء على ملفات التحكم بالوصول
22	Is_guest_login	1: تسجيل الدخول هو 'guest' (مثل ...,guest,anonymous) 0: غير ذلك
23	Count	عدد الاتصالات إلى نفس المضيف حيث الاتصال الحالي بآخر ثانيتين
24	Srv_count	عدد الاتصالات إلى نفس الخدمة حيث الاتصال الحالي في آخر ثانيتين
25	Serror_rate	النسبة المئوية للاتصالات التي لها 'SYN' خطأ إلى نفس المضيف
28	Sev_rerror_rate	النسبة المئوية للاتصالات التي لها 'REJ' خطأ إلى نفس الخدمة
29	Same_srv_rate	النسبة المئوية للاتصالات إلى نفس الخدمة و نفس المضيف
30	Did_srv_rate	النسبة المئوية للاتصالات إلى خدمات مختلفة و نفس المضيف
31	Srv_diff_host_rate	النسبة المئوية للاتصالات إلى نفس الخدمة و مضيفين مختلفين
32	Dst_host_count	عدد الاتصالات إلى نقس المضيف إلى مضيف الوجهة حيث الاتصال الحالي بآخر ثانيتين
33	Dst_host_srv_count	عدد الاتصالات من نفس الخدمة إلى مضيف الوجهة حيث الاتصال الحالي في آخر ثانيتين
35	Dst_host_diff_srv_rate	النسبة المئوية للاتصالات من خدمات مختلفة إلى مضيف الوجهة
36	Dst_host_sane_src_port_rate	النسبة المئوية للاتصالات من منفذ الخدمات إلى مضيف الوجهة
37	Dst_host_srv_diff_host_rate	النسبة المئوية للاتصالات من مضيفين مختلفين من نفس الخدمة إلى مضيف الوجهة

1- النتائج العملية :

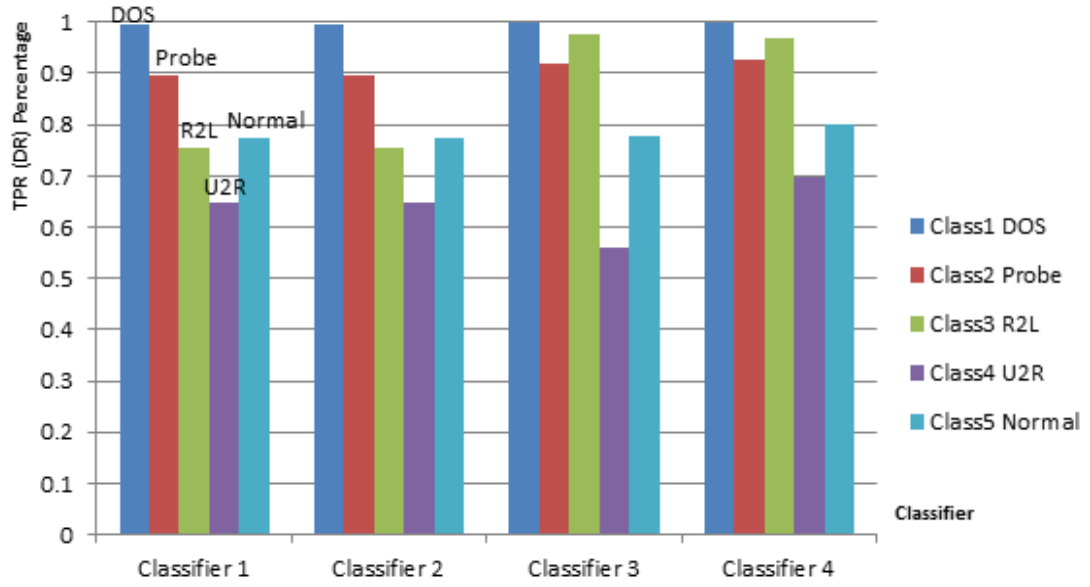
يبين الجدول (11) مقارنة بين المصنفات الأربعة من حيث معدل الإيجابيات الصحيحة (TPR)، معدل السلبيات الصحيحة (TNR)، معدل الأخطاء الإيجابية (FPR)، معدل الأخطاء السلبية(FNR) ، ودقة النظام الكلية (Acc).

سيتم التركيز على ثلاث قيم، معدل الإيجابيات الصحيحة TPR الذي يعبر عن الحالات التي تم تصنيفها بشكل صحيح من قبل المصنف، معدل الإيجابيات الخاطئة FPR أو ما يسمى معدل الإنذارات الكاذبة الذي يعبر عن الحالات التي تم فيها إطلاق إنذار من قبل المصنف على وجود صنف ما ولكن في الحقيقة كان إنذارا خاطئا، بالإضافة إلى معيار الدقة الكلية Acc للمصنف وهي نسبة التصنيفات الصحيحة إلى التصنيفات الكلية.

الجدول (11) المقارنة بين أداء كل مصنف من المصنفات الأربعة من حيث الدقة ومعدل الكشف الإيجابي والسلبي ومعدلات الأخطاء الإيجابية والسلبية لكل صنف من الأصناف الخمسة.

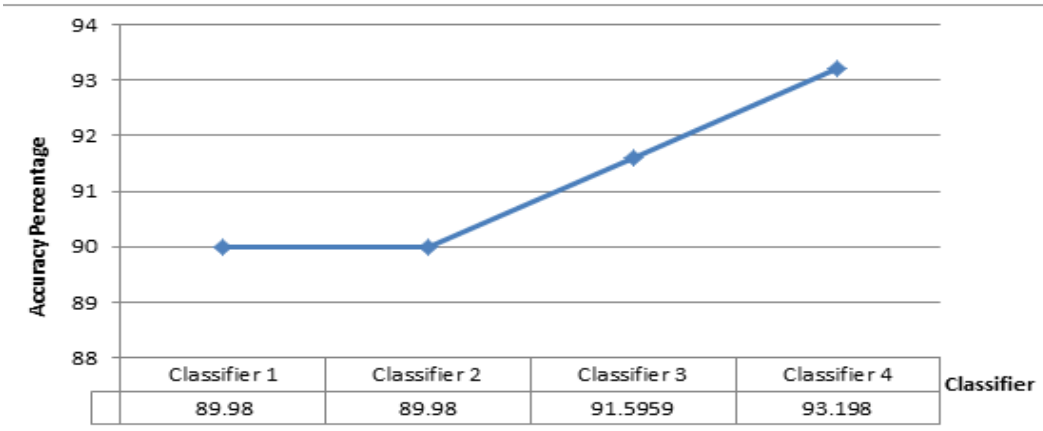
Accuracy	TPR(DR)	TNR	FPR	FNR	المصنف الأول
%89.98	0.9973	0.9886	0.0027	0.0114	Class1 (DoS)
	0.8943	0.9969	0.1057	0.0031	Class2 (Probe)
	0.7543	0.9471	0.2457	0.0529	Class3 (R2L)
	0.6471	0.9998	0.3529	0.0002	Class4 (U2R)
	0.7747	0.9980	0.2253	0.0020	Class5(Normal)
Accuracy	TPR(DR)	TNR	FPR	FNR	المصنف الثاني
%89.98	0.9973	0.9886	0.0027	0.0114	Class1 (DoS)
	0.8943	0.9969	0.1057	0.0031	Class2 (Probe)
	0.7543	0.9471	0.2457	0.0529	Class3 (R2L)
	0.6471	0.9998	0.3529	0.0002	Class4 (U2R)
	0.7747	0.9980	0.2253	0.0020	Class5(Normal)
Accuracy	TPR(DR)	TNR	FPR	FNR	المصنف الثالث
%91.5959	0.9988	0.9860	0.0012	0.0140	Class1 (DoS)
	0.9178	0.9968	0.0822	0.0032	Class2 (Probe)
	0.9775	0.9493	0.0225	0.0507	Class3 (R2L)
	0.5588	0.9998	0.4412	0.0002	Class4 (U2R)
	0.7769	0.9988	0.2231	0.0012	Class5(Normal)
Accuracy	TPR(DR)	TNR	FPR	FNR	المصنف الرابع
%93.198	0.9987	0.9849	0.0013	0.0151	Class1 (DoS)
	0.9249	0.9977	0.0751	0.0023	Class2 (Probe)
	0.9682	0.9496	0.0318	0.0504	Class3 (R2L)
	0.6982	0.9998	0.3684	0.0002	Class4 (U2R)
	0.7988	0.9987	0.2012	0.0013	Class5(Normal)

يبين الشكل (1) مقارنة دقة كشف النظام DR ضمن كل مصنف من المصنفات الأربعة مع دقة كشف كل صنف من الأصناف الخمسة على حدة:



الشكل (1) مقارنة بين المصنفات من حيث معدل كشف كل صنف

يبين الشكل (2) مدى تحسن دقة المصنف من التجربة الأولى بوجود السمات كاملة إلى التجربة الرابعة مع التخفيض الأخير لعدد السمات المشاركة في عملية التصنيف، حيث تحسن بنسبة 3.218%.



الشكل (2) مقارنة بين المصنفات الأربعة من حيث الدقة

نلاحظ عند دراسة النتائج فيما يتعلق بنظام كشف التسلل أن نتائج المصنفين الأول والثاني متماثلة وهذا يؤكد عدم مساهمة السمتين 20 و21 في عملية التصنيف (قيم صفرية لتباينات السمتين كما هو موضح في الجدول (8)). بالنسبة للمصنف الثالث في التجربة الثالثة بعد حساب قيم الارتباط للسمات (كما هو موضح في الجدول (9)) نلاحظ تحسن دقة التصنيف وحساسية نظام كشف التسلل، وانخفاض معدلات الأخطاء الإيجابية والسلبية بالمقابل، يوضح الجدول (11) ارتفاع معدل كشف الصنف الأول DoS والثاني Probe والثالث R2L والخامس Normal وانخفاض في معدل كشف الصنف الرابع U2R. تكون دقة المصنف الأخير بعد اعتماد مجموعة السمات النهائية (30 سمة) هي الأفضل حيث دقة نظام كشف التسلل 0.943 وحساسية النظام 0.93، يعود ليتحسن معدل كشف الصنف الرابع U2R، والصنف الخامس Normal، وتبقى الأصناف الباقية بنفس معدل الكشف تقريبا.

تبين النتائج أولا مدى فعالية استخدام الشبكات العصبونية لكشف التسلل في الشبكات اللاسلكية، وأوضحت النتائج من ناحية أخرى أن أداء مصنف الشبكة العصبونية لنظام كشف التسلل هو الأفضل مع عدد سمات أقل.

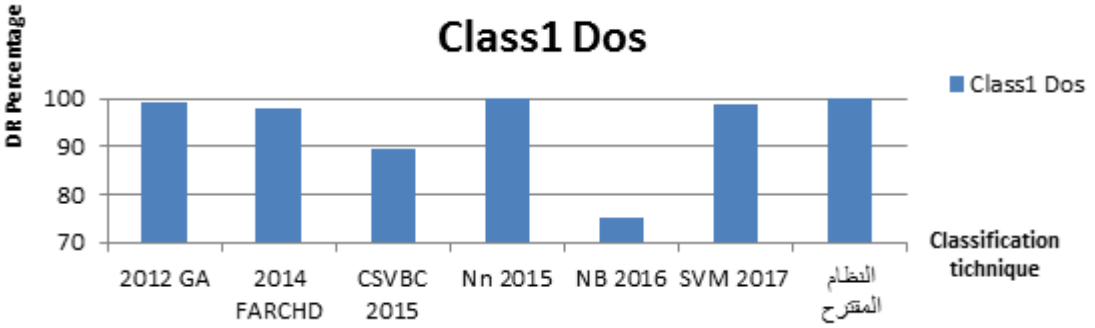
باعتماد النموذج الأخير الذي تم التوصل إليه في التجربة الرابعة كمصنف لنظام كشف التسلل ضمن هذا البحث. يمكن مقارنة النظام المقترح مع أنظمة أخرى تم اقتراحها في أبحاث سابقة استخدمت نفس مجموعة البيانات المستخدمة KDD99 ولكن بالاعتماد على خوارزميات مختلفة. يبين الجدول (12) هذه المقارنة بين النظام المقترح الذي استخدم خوارزمية الانتشار الخلفي للشبكات العصبونية والتقنيات المستخدمة في عدة دراسات سابقة من حيث معدلات كشف الأصناف الهجومية والصنف الطبيعي.

الجدول (12) مقارنة بين معدلات كشف الأصناف للنظام المقترح ضمن هذا البحث و لأنظمة أخرى ضمن دراسات سابقة.

Class5Normal	Class4U2R	Class3R2L	Class2Probe	Class1Dos	DR
69.5%	18.9%	5.4%	71.1%	99.4%	GA 2012
99.81%	65.38%	87.54%	95.83%	98.05%	FARCHD 2014
60.54%	5.26%	82.05%	72.43%	89.66%	CSVBC 2015
Not Included	0%	82.88%	95.12%	99.99 %	Nn 2015
74.9%	72.3%	70.1%	74.1%	75.2%	NB 2016
60%	57.89%	59.4%	81.62%	98.67%	SVM 2017
79.88%	69.82%	96.82%	92.49%	99.87%	النظام المقترح

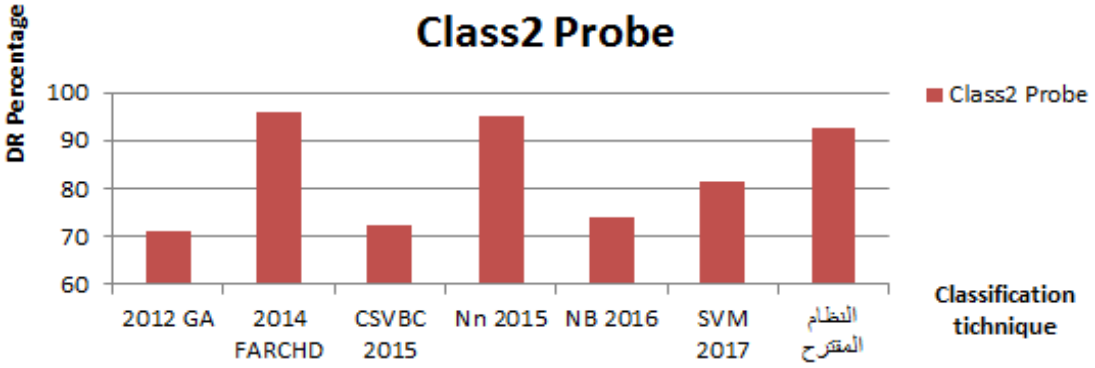
نلاحظ من الجدول (12) بمقارنة النظام المقترح (السطر الأخير) مع الدراسة [27] عام 2012 (السطر الأول) التي اقترحت نظام كشف تسلل IDS بتطبيق الخوارزمية الجينية GA، أن نتائج النظام المقترح أعلى بالنسبة لمعدلات كشف الأصناف الهجومية الأربعة والصنف الطبيعي. بالنسبة للدراسة [28] عام 2014 (السطر الثاني) التي استخدمت تقنية جديدة في تصميم نظام كشف التسلل تعتمد على الأنظمة الضبابية الجينية GFS، هي خوارزمية FARCHD، كانت نتائج هذه الخوارزمية أعلى من نظامنا بالنسبة لمعدل كشف كل من الصنف الطبيعي (99.81%) والصنف الهجومي Probe (95.83%)، بالمقابل كانت نتائج النظام المقترح (السطر الأخير) أعلى من حيث معدلات كشف الأصناف الهجومية الثلاثة الباقية (DOS, R2L,U2R). في العام 2015 كانت الدراسة [29] التي اقترحت تصميم نظام كشف تسلل بالاعتماد على الدمج بين تقنيتين SVM و Bee Colony وكانت النتيجة تقنية CSVBC) (، نلاحظ من الجدول أن نتائج نظامنا أعلى من نتائج التقنية CSVBC بالنسبة لمعدلات كشف الأصناف الهجومية الأربعة والصنف الطبيعي. في نفس العام كانت الدراسة [13] التي استخدمت نفس الخوارزمية المستخدمة في هذا البحث، وحققت نتائج أعلى من بحثنا بالنسبة للصنف Probe بنسبة (% 2.63)، في حين فشلت الدراسة [13] في التعرف على الصنف U2R وكانت نتائج كشف الصنف R2L أقل من النظام المقترح بنسبة 13.94%. في العام 2016 كانت الدراسة [19] التي استخدمت مصنف بايز البسيط وحققت نتائج متقاربة بالنسبة لمعدلات كشف الأصناف الخمسة، كانت نتائج هذه الدراسة أقل من النظام المقترح بالنسبة للأصناف جميعها باستثناء الصنف U2R (72.3%). استخدمت الدراسة [30] عام 2017 تقنية آلة شعاع الدعم SVM في بناء نظام كشف تسلل، كانت نتائج هذه الدراسة أقل من النظام المقترح في هذا البحث بالنسبة لمعدلات كشف الأصناف الهجومية الأربعة والصنف الطبيعي.

تبين الأشكال (3)، (4)، (5) و(6) مخططات توضيحية تبين المقارنة التي تمت في الجدول (12) بين النظام المقترح والأنظمة السابقة التي تعتمد استخدمت نفس مجموعة البيانات، من حيث معدلات الكشف لكل من الأصناف الهجومية الأربعة (DoS, Probe, R2L,U2R) على الترتيب.



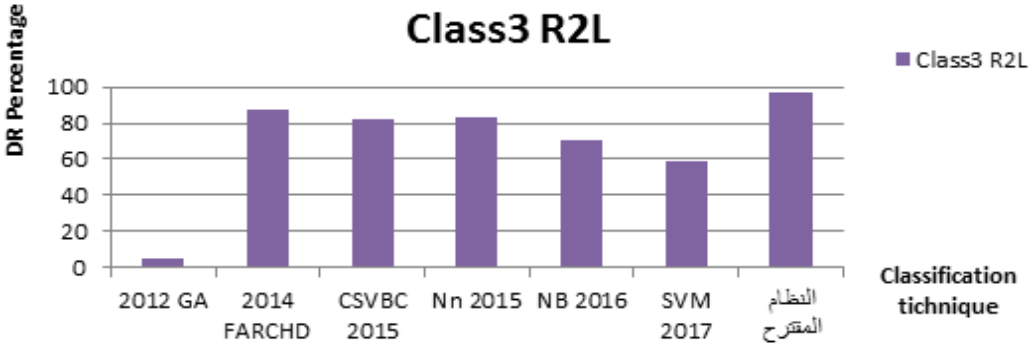
الشكل (3) مقارنة بين أنظمة كشف التسلل و النظام المقترح من حيث معدل كشف الصنف DoS

يبين الشكل (3) أن معدل كشف النظام المقترح للصنف DoS (99.87%) أعلى من معدلات الكشف في الدراسات السابقة، وكانت النتيجة متقاربة مع الدراسة [13] عام 2015.



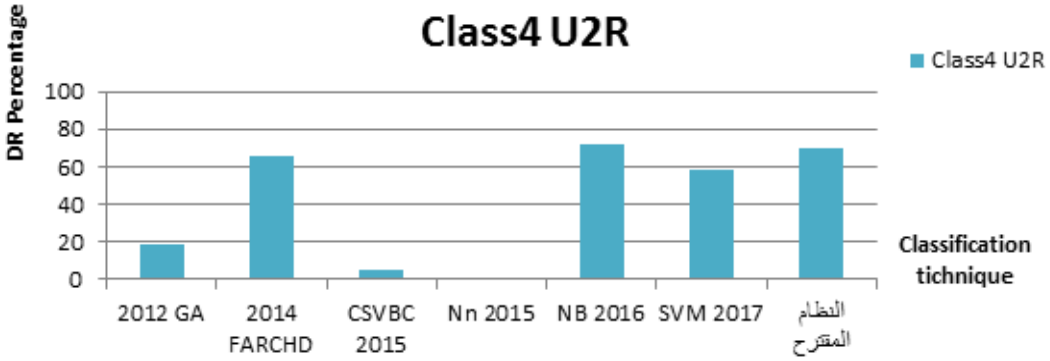
الشكل (4) مقارنة بين أنظمة كشف التسلل و النظام المقترح من حيث معدل كشف الصنف Probe

نلاحظ من الشكل (4) أن معدل كشف النظام المقترح للصنف Probe (92.49%) أقل من معدل كشف الصنف في الدراستين [28] و [13] عام 2014 و 2015، وأعلى من الدراسات الباقية عام 2012، 2016 ، 2015 و 2017.



الشكل (5) مقارنة بين أنظمة كشف التسلل و النظام المقترح من حيث معدل كشف الصنف R2L

يوضح الشكل (5) أن معدل كشف النظام المقترح للصنف R2L (96.82%) أعلى من معدل كشف الصنف في جميع الدراسات المذكورة.



الشكل (6) مقارنة بين أنظمة كشف التسلل و النظام المقترح من حيث معدل كشف الصنف U2R

يوضح الشكل (5) أن معدل كشف النظام المقترح للصنف U2R (69.82%) أعلى من معدل كشف الصنف في جميع الدراسات باستثناء الدراسة [19] عام 2016.

الاستنتاجات والتوصيات:

اعتمادا على الدراسة التي أُجريت ضمن هذا البحث على مجموعة بيانات القياسية KDD99 والتي اعتمدت على خوارزمية الانتشار الخلفي للشبكات العصبونية لبناء نظام كشف تسلل، أظهرت النتائج أن استخدام الشبكات العصبونية يمكّن نظام كشف التسلل من كشف أصناف الهجومات الأربعة الموجودة ضمن مجموعة بيانات الاختبار بدقة عالية، كما أظهرت تحسّن في دقة الكشف وانخفاض في معدلات الأخطاء الإيجابية والسلبية عند استخدام مجموعة السمات المخفّضة.

نظرا لأهمية البيانات المستخدمة في بناء نظم كشف التسلل، يُنصح البحث في قواعد بيانات أخرى والتركيز على إدخال أصناف جديدة من الهجمات، مع الاعتماد على نفس الخوارزمية المتبعة في هذا البحث أي تقنية الشبكة العصبونية ذات خوارزمية الانتشار الخلفي بهدف مقارنة النتائج.

يمكن من جهة أخرى دراسة تأثير الخوارزمية المستخدمة على كشف التسلل، من خلال الاعتماد على نفس مجموعة البيانات المستخدمة في هذا البحث KDD99، لكن مع تطبيق تقنيات أخرى غير تقنية الشبكة العصبونية، أو حتى باستخدام خوارزميات أخرى للشبكة العصبونية غير خوارزمية الانتشار الخلفي، ودراسة النتائج و مقارنتها.

المراجع:

[1] BROWN, C.; et al. Analysis of the 1999 DARPA/Lincoln Laboratory IDS Evaluation Data with NetADHICT. In Proc. of the Second IEEE international conference 10 –on Computational intelligence for security and defense applications, Canada, July 08 2009, 67-73.

[2] KDD Cup 1999 Data. https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. 21/May/201

[3] TSANG, C.; et al. Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. The Journal of The Pattern 2391.–Recognition Society, Vol. 40, 2007, 2373

[4] REHMAN, R.U. Intrusion Detection Systems with Snort Advanced IDS Techniques Using Snort, Apache, MySQL, PHP, and ACID. 1st.ed., Prentice Hall PTR, United States of America, 2003, 275.

[5] RAJU, P. N. State-of-the-art Intrusion Detection: Technologies, Challenges, and Institutionen för systemteknik, Linkoping, Feb 2005. Evaluation.

[6] LEE, W.; et al. Adaptive Intrusion Detection: a Data Mining Approach. Springer Netherlands, Vol. 14 No. 6, 2000, 533-567.

[7] SABHNANI, M.; et al. Why Machine Learning Algorithms Fail in Misuse Detection on KDD Intrusion Detection Data Set. Journal Intelligent Data Analysis, USA, 415.–Vol. 8 No. 4, September 2004, 403

[8] CHANDOLA, V.; et al. Anomaly Detection : A Survey. ACM Computing Surveys, USA, Vol. 41, No. 15, July 2009.

[9] ADSUL, A.P.; et al. Attacks classification in Network Intrusion Detection System Using ANN. International Journal of Application or Innovation in Engineering & Management (IJAIEM), Vol. 3 , April 2014, 397-400.

[10] GYANCHANDANI, M.; et al. Taxonomy of Anomaly Based Intrusion Detection System: A Review. International Journal of Scientific and Research Publications, Vol. 2, December 2012.

[11] NOVIKOV, D.; et al. Anomaly Detection Based Intrusion Detection. In Proc. of the Third International Conference on Information Technology: New Generations, April 10 12 2006, 420-425.–

[12] DIAMANTARAS, K.; et al. Principle Component Neural Networks: Theory and Applications. New York: John Wiley & Sons Inc., 2006.

[13] SHAKYA, S.; et al. Intrusion detection system using back Propagation s performance with Self Organizing Map. Journal of Advanced 'Algorithm and Compare it College of Engineering and Management, Vol. 1, 2015.

[14] OZGUR, A.; et al. A Review of KDD99 Dataset Usage in Intrusion Detection and Machine Learning between 2010 and 2015. PeerJ Preprints, April 14 2016.

[15] SHYU, M.; et al. Handling Nominal Features in Anomaly Intrusion Detection Problems. In Proc. of the 15th International Workshop on Research Issues in Data Engineering: Stream Data Mining and Applications, April 03 - 04 2005, 55-62.

[16] ABDULLAH, M.; ALSANEE, E.; ALSEHEYMI, N. Energy Efficient Cluster-Based Intrusion Detection System for Wireless Sensor Networks. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 5, No. 9, 2014, 10-15.

[17]BHUYAN, M. H.; et al. Towards Generating Real-life Datasets for Network Intrusion Detection.IJ Network Security, 2015, Vol. 17, No.6, 683-701.

[18] BEKKAR, M.; et al. Evaluation Measures for Models Assessment over Imbalanced Data Sets. Information Engineering and Applications, Vol.3 No.10, 2013, 27-39.

[19] TIWARI, V. N. Enhanced Method for Intrusion Detection over KDD Cup 99 Dataset. International Journal of Current Trends in Engineering & Technology, Vol. 02 No. 02, Mar-Apr 2016.

[20] HUSENYNOV, Kh.; et al. Evaluation of Public Datasets for Intrusion Detection/Prevention System Benchmark. under the R&D program , KUSTAR-KAIST Institute, Korea, 27-09-2013.

[21] ZUECH, R.; et al. A New Intrusion Detection Benchmarking System. In Proc. of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, 2015-04-07, 253-255.

[22] DHANABAL, L.; et al. A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms. International Journal of Advanced Research in Computer and Communication Engineering. Vol. 4, June 2015, 446-452.

[23] DARPA Intrusion Detection Evolution. https://www.ll.mit.edu/ideval/data/2000/LLS_DDOS_2.0.2.html. 1/11/2016.

[24] The UNSW-NB15 data set description. https://www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-NB15-Datasets/. 25/Dec/2016.

[25] LBNL/ICSI Enterprise Tracing Project. http://www.icir.org/enterprise-tracing/ . 12/Dec/2016.

[26] CAIDA Data, Overview of Datasets, Monitors, and Reports. http://www.caida.org/data/overview/. 1/Dec/2016.

[27] HOQUE, M. S.; et al. An Implementation of Intrusion Detection System Using Genetic Algorithm. International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012.

[28] ELHAG, S.; et al. On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems. ELSEVIER, 11 August 2014, 193-202.

[29] GUPTA, M.; et al. Intrusion Detection System based on SVM and Bee Colony. International Journal of Computer Applications, Vol.111, No.10, February 2015, 27-32.

[30] YENDOLE, S; et al. Identifying Intrusion Detection System using Hybrid technique with Support Vector Machine. International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 5 , No. 3, March 2017.