Analysis of Basic Compounds in a Network Intrusion Detection System using NSL-KDD Data

Naglaa Badi Ibrahim **Hana Muhammad Usman**
College of Computer Science and Mathematics
University of Mosul, Mosul, Iraq

Received on: 15/10/2012 **Accepted on: 30/01/2013**

ABSTRACT

The increasing of security attacks and unauthorized intrusion have made network security one of the main subjects that should be considered in present data communication environment. Intrusion detection system is one of the suitable solutions to prevent and detect such attacks. This paper aims to design and implement a Network Intrusion Detection System (NIDS) based on genetic algorithm. In order to get rid of redundancy and inappropriate features principle component analysis (PCA) is useful for selecting features. The complete NSL-KDD dataset is used for training and testing data.

Number of different experiments have been done. The experimental results show that the proposed system based on GA and using PCA (for selecting five features) on NSL-KDD able to speed up the process of intrusion detection and to minimize the CPU time cost and reducing time for training and testing. C# programming language is used for system implementation.

Keywords: network security, Intrusion detection system, genetic algorithm.

تحليل المركبات الأساسية في نظام كشف التطفل الشبكي باستخدام بيانات NSL- KDD

نجلاء بديع إبراهيم **هناء محمد عصمان**
كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ استلام البحث: 2012/10/15 **تاريخ قبول البحث: 2013/01/30**

الملخص

إن أمنية الشبكات هي إحدى المواضيع الرئيسية التي يجب الاهتمام بها في بيئة الاتصالات الحالية بسبب زيادة التهديدات الأمنية والتطفلات الغير المخولة. نظم كشف التطفل هي إحدى الحلول المناسبة لصد وكشف هذه الهجمات. يهدف البحث إلى تصميم نظام كشف تطفل شبكي (NIDS) يعتمد على الخوارزمية الجينية. تم استخدام خوارزمية تحليل المركبات الأساسية PCA للتخلص من الميزات الفائضة وقليلة الفائدة. تم اعتماد مجموعة بيانات NSL-KDD كاملة في تدريب واختبار البيانات.

أجريت عدد من التجارب المختلفة وأظهرت النتائج التجريبية أنّ النظام المقترح للخوارزمية الجينية مع PCA باختيار خمس ميزات على بيانات NSL-KDD قادر على تسريع عملية الكشف عن التطفل وتصنيفها مع تقليل زمن المعالج وتقليل زمن التدريب والاختبار. استخدمت لغة فيجوال سي شارب (Visual C# 2008).

الكلمات المفتاحية: أمنية الشبكة، نظام كشف التطفل، خوارزمية جينية.

1. المقدمة

إن تعقيد أنظمة الحواسيب الموزعة وأهميتها وموارد المعلومات المتوفرة فيها نمت بسرعة كبيرة، استنادا لهذه الحقيقة فقد أصبحت الحواسيب وشبكاتها هدفاً لجرائم الحاسوب التي ازدادت أكثر فأكثر[1] .

استخدمت بعض التقنيات لحماية البيانات المهمة مثل الجدار الناري والتشفير وألخ. يعمل الجدار الناري مدافعاً لحماية البيانات الحساسة ولكنه يقلل الكشف فضلا عن المراقبة وتحديد نقاط الضعف في أنظمة التشغيل. إن الرسالة المشفرة ممكن فك شفرتها وفضلاً عن ذلك فإن التشفير يضيف عباً أضافياً على المستخدمين

والتطبيقات. إن أي تقنية أمنية جديدة تحوي في تصميمها على بعض العيوب مما يجعلها هدفا للهجمات، لذلك فإن من المهم أن نملك أنظمة كشف التطفل لحماية البيانات المهمة [2].

اتجه الباحثون إلى استخدام المفاهيم الذكائية لحل مشكلات أمنية الشبكات. تم بتقنية كشف إساءة الاستخدام اختيار الخوارزمية الجينية بسبب خصائصها الجيدة، مثل المرونة، وعدم الحاجة لمعلومات مرتبة لإيجاد حل امثل أو شبه امثل، وقابلية التعلم الذاتي، والأمثلية مع متغيرات مستمرة أو متقطعة، والتعامل مع عدد كبير من المتغيرات، وملاءمة للحواسيب المتوازية، والعمل مع مجموعة حلول لا حل منفرد، وتشفير المتغيرات، لذا تتم الأمثلية بمتغيرات مشفرة، والعمل مع بيانات مولدة عددياً، وبيانات تجريبية، أو دوال تحليلية [3,4].

تمت دراسة مشكلة كشف التطفل بشكل واسع ضمن حقل أمنية شبكة الحاسوب، إذ اقترح الباحثون في السنوات السابقة عدداً من الأساليب لتحديد المتطفلين وكشفهم في الأنظمة الحاسوبية، استخدم الباحث (Adhitya Chittur) [5] منهجاً فريداً للكشف عن التطفل باستخدام الخوارزمية الجينية واختبار كون هذه الخوارزمية خيارا ممكننا لتوليد الأنموذج في أنظمة كشف التطفل، معتمدا على الذكاء الاصطناعي. وأثبتت النتائج إن الخوارزمية الجينية قادرة بنجاح على توليد أنموذج سلوكي تجريبي دقيق من تدريب البيانات والقدرة على تطبيق المعرفة التجريبية بنجاح من بيانات لم يسبق التعامل معها.

في عام 2010 اقترح الباحثون (Mrutyunjaya Panda et. al.,) [6] أنموذجاً مبتكراً لنظام كشف التطفل يربط المصنف Naïve Bayes مع ثلاثة خوارزميات مختلفة لتقليل الميزات (تحليل المركبات الأساسية PCA، والتقدير العشوائي Random Projection (RP)، و(Nominal to Binary (N2B)، واستخدمت مجموعة مختارة من بيانات NSL− KDD، وأظهرت النتائج دقة كشف عالية للمصنف Naïve Bayes مع خوارزمية N2B بلغت (96.5%) ونسبة الإنذار الكاذب (3.0%)، في حين بلغت دقة الكشف للمصنف Naïve Bayes مع RP، PCA (81.4% و 94.8%) على التوالي، ونسبة الإنذار الكاذب (4.4% , 12.8%) على التوالي.

في عام 2010 قدم الباحثون (Shilpa lakhina et al.) [7] نظام كشف تطفل الشذوذ باستخدام خوارزمية تحليل المركبات الأساسية والشبكات العصبية الاصطناعية (PCANNA). قللت الخوارزمية المقترحة وقت التدريب والاختبار إلى (40%) و(70%) على التوالي، وقد استخدمت مجموعة مختارة من نماذج التدريب والاختبار NSL− KDD.

تم استخدام مجاميع NSL-KDD [8] على التطفلات التي أصبحت معياراً فعليا لاختبار أنظمة كشف التطفل. والغاية من استخدام هذه البيانات كونها القاعدة الأساسية المشتركة لأغلب الباحثين العاملين في مجال أنظمة كشف التطفل والتي من خلالها يتم المقارنة بين التقنيات الأخرى.

بصورة عامة تتعامل نظم كشف التطفل مع كمية هائلة من البيانات حتى بالنسبة لشبكة صغيرة، مما تحوي ميزات زائدة وغير ذات صلة. يمكن لهذه الميزات أن تجعل اكتشاف أنماط السلوك الشاذ صعباً مما يسبب عملية تدريب واختبار بطيئة، واستهلاكاً أعلى للمصادر، إضافة إلى نسبة كشف سيئة. لذلك يفضل تقليل أبعاد البيانات لاكتشاف أسهل وتحليل أسرع اعتماداً على خوارزمية PCA [9].

2. مدخل إلى أنظمة كشف التطفل

يمكن تصنيف أنظمة كشف التطفل من ناحية نظريات الكشف:

كشف إساءة الاستخدام (Misuse Detection)، كشف الشذوذ (Anomaly Detection)، كشف المحددات (Specification Detection). تقسم أنظمة كشف التطفل طِبقاً لمصدر معلوماتها إلى ثلاثة أقسام:
نظام كشف تطفل المضيف، ونظام كشف التطفل الشبكي وأنظمة كشف التطفل الهجيني.

استخدم العديد من الباحثين بيانات KDD Cup 1999 [10] لبناء نظم كشف التطفل. أظهرت الدراسات السابقة وجود بعض المشاكل الكامنة في هذه البيانات. إنّ التحديد المهم لهذه البيانات هو العدد الهائل للسجلات الزائدة بمعنى إن 78% من سجلات التدريب و75% من سجلات الاختبار متكررة ، هذه البيانات تعاني من بعض المشاكل وقد لا تكون مثلى للشبكات الفعلية الموجودة، ويمكن أن يكون محاكاة الهجوم ضمن واحد من الأصناف الأربعة (DoS, Probe, U2R, R2L) [11].

مجاميع البيانات المتولدة، KDD Test, KDD Train شملت (22544,125973) سجلاً على التوالي.
اقتُرحت بيانات NSL-KDD من قبل (Tavallaee et al.) حل مشاكل بيانات KDD المذكورة سابقاً. تعتبر NSL-KDDنسخة مختزلة من بيانات KDD الأصلية وتتألف من نفس ميزات بيانات KDD 99 التي تحوي في كل سجل اتصال TCP على إحدى وأربعين ميزة مع عنوان يوضح هل هذا الاتصال هو اتصال اعتيادي أو نوع من أنواع الهجمات، وهناك ثمانٍ وثلاثون ميزة رقمية وثلاث ميزات رمزية. فيما يأتي فوائد NSL-KDD مقارنة بمجموعة بيانات KDD الأصلية [12]:

- لا تشمل سجلات زائدة في مجموعة التدريب، لذا لن تميل المصنفات باتجاه سجلات أكثر حدوثا.
- عدد السجلات المختارة من كل مجموعة: مستوى الصعوبة يتناسب عكسيا ونسبة السجلات في مجموعة بيانات KDD الأصلية. نتيجة لذلك نسب تصنيف طرائق تعليم الآلة المتميزة تختلف بمدى واسع، مما يجعل من زيادة الفعالية امتلاك تقييم دقيق لتقنيات تعليم مختلفة.
- عدد السجلات في التدريب ومجاميع الاختبار معقول، مما يجعل من المحتمل إجراء تجارب على المجموعة الكاملة دون الحاجة للاختبار العشوائي لنسبة ضئيلة. ونتيجة لذلك سيكون تقييم نتائج البحوث المختلفة ثابتة ومشابهة.

3. تحليل المركبات الأساسية Principal Component Analysis (PCA)

إن PCA تقنية شائعة تحوّل عدداً من الميزات المترابطة (Correlated Features) إلى ميزات غير مترابطة (Uncorrelated Features) والتي تدعى المركبات الأساسية (Principal Components) [13].

يتطلب تطبيق تحليل المركبات الأساسية حساب مصفوفة التباين / التباين المشترك (Variance/Covariance matrix) للميزات، وبعدها يتم حساب المركبات الأساسية بطريقة جاكوبي (Jacobi Method).

يتطلب تحليل المركبات الأساسية رياضيا إيجاد مصفوفة التباين / التباين المشترك للميزات المتوفرة. يتم حساب مصفوفة التباين / التباين المشترك وفق المعادلة (1):

$$Cov_{ij} = \frac{1}{MN}\sum_{k=1}^{M}\sum_{l=1}^{N}\left(X_i(k,l) - M_i\right)\left(X_j(k,l) - M_j\right) \quad \ldots(1)$$

حيث M, N عدد السجلات الموجودة في نماذج التدريب وعدد الميزات الموجودة في كل سجل على الترتيب و j,i تمثل موقع الميزة لسجل وموقع السجل في الأنموذج على الترتيب وM_i وM_j هي المتوسط الحسابي للميزات i و j ويمكن حسابها على وفق المعادلة (2):

$$M_i = \frac{1}{MN}\sum_{k=1}^{M}\sum_{l=1}^{N} X_i(k,l) \quad \ldots(2)$$

نلاحظ من المعادلة إن قيم المصفوفة تكون تباينا (Variance) لعناصر القطر الرئيسي والتباين المشترك (Covariance) لبقية العناصر كذلك إن مصفوفة (Covariance) مصفوفة متناظرة أي إن العناصر فوق القطر الرئيسي تكون مساوية بالقيم للعناصر تحت القطر المناظر لها .

وهكذا فان مصفوفة التباين / التباين المشترك ل n عدد من الميزات هي مصفوفة nxn ويمكن ترتيبها على النحو الآتي:

$$A = \begin{bmatrix} a_{11} & a_{12} & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & a_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1} & a_{n2} & . & . & a_{nn} \end{bmatrix} \quad \ldots(3)$$

وحيث إن $a_{ij}=a_{ji}$ لكل $i \neq j$ فان المصفوفة A متناظرة.

مصفوفة متجه الأيكن T تشتق من مصفوفة A باستخدام طريقة جاكوبي الموضحة في الفقرة اللاحقة والتي نحصل عليها من D، وتمثل بالمعادلة (4):

$$D = \begin{bmatrix} \lambda_{11} & 0 & . & . & 0 \\ 0 & \lambda_{22} & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda_{nn} \end{bmatrix} \quad \ldots(4)$$

إن العناصر القطرية D تسمى قيمة المميزة (Eigen Value) لمصفوفة التباين / التباين المشترك حيث λ_{ii} لكل $i=1, 2, \ldots, n$ هي التباينات لمحاور المركبات الأساسية، إن العناصر خارج القطر (Off-Diagonal) للمصفوفة D هي صفر (أو قريب من الصفر) مما يدل على إن المركبات الخارجة غير مترابطة أي إنها مستقلة [14].

طريقة جاكوبي Jacobis Method

1. يتم إيجاد اكبر عنصر في مصفوفة معينة مربعة ولتكن A بحيث لا يكون من عناصر القطر الرئيسي أي أنّ $Max\_element=a_{ik}$ و i, k هي الصف والعمود في مصفوفة A و $i \neq k$
2. يتم إيجاد الزاوية θ وذلك عن طريق ما يأتي:

$$\theta = \frac{1}{2}\arctan\left(2a_{ik}/(a_{ii}-a_{kk})\right) \qquad \text{If } a_{ii} \neq a_{kk} \quad \ldots(5)$$

$$\theta = \begin{cases} \pi/4 & when\ a_{ik} > 0 \\ -\pi/4 & when\ a_{ik} < 0 \end{cases} \qquad \text{If } a_{ii} = a_{kk} \quad \ldots(6)$$

3. إجراء عملية التدوير (Rotation) على المصفوفة C وإرجاع الناتج في مصفوفة وكما يأتي:

$$d_{ii} = \frac{1}{2}\left(a_{ii} + a_{kk} + \sigma R\right) \quad \ldots(7)$$

$$d_{kk} = \frac{1}{2}(a_{ii} + a_{kk} - \sigma R) \quad \ldots(8)$$

$$d_{ik} = d_{ki} = 0 \quad \ldots(9)$$

حيث يتم إيجاد قيمة R على وفق المعادلة الآتية:

$$R = \sqrt{(a_{ii} - a_{kk})^2 + 4a_{ik}^2} \quad \ldots(10)$$

ويتم إيجاد قيمة σ على وفق المعادلة الآتية:

$$\sigma = \begin{cases} 1 & if \quad a_{ii} \geq a_{kk} \\ -1 & if \quad a_{ii} < a_{kk} \end{cases} \quad \ldots(11)$$

أما بقية عناصر مصفوفة التدوير فنجدها باستخدام المعادلتين الآتيتين:

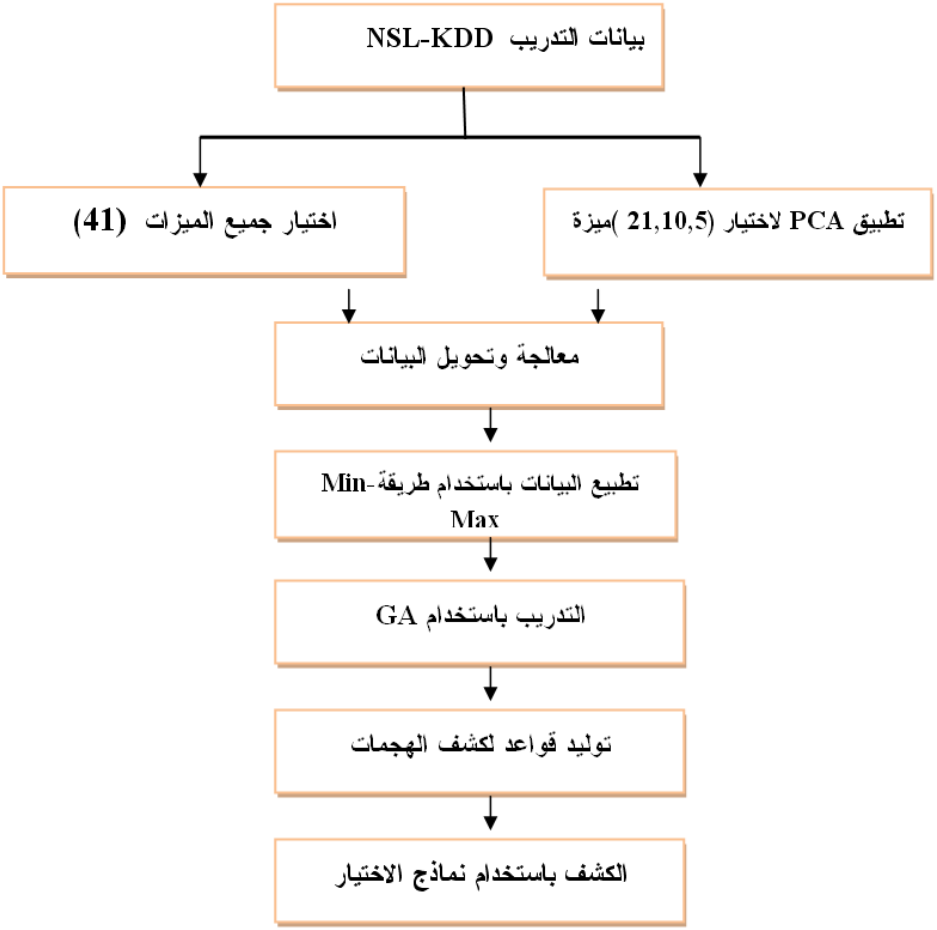
$$d_{ir} = a_{ir}\cos\theta + a_{kr}\sin\theta \quad \ldots(12)$$

$$d_{kr} = -a_{ir}\sin\theta + a_{kr}\cos\theta \quad \ldots(13)$$

حيث $r \neq i$ و $r \neq k$ و i, k تمثل قيمة الصف والعمود داخل المصفوفة.

4. إعادة الخطوات من 1 إلى 3 على المصفوفة الناتجة إلى أن يتم الحصول على قيم عناصر خارج القطر قريبة من الصفر [14] .

4. الخوارزمية المقترحة: الشكل (1) يوضح معمارية الخوارزمية المقترحة



الشكل (1) معمارية الخوارزمية المقترحة

1-4 خطوات تنفيذ خوارزمية تحليل المركبات الأساسية

أولا: قراءة مجموعة نماذج التدريب NSL-KDD .

ثانيا: معالجة بيانات التدريب، حيث يتم تحويل الميزات الحرفية إلى عددية فالميزات (Protocol type, Service ,Flag) يتم تحويلها من [1 .. عدد القيم ضمن الميزة]، وكذلك تحويل حقل Attack Type من (Normal وAnomaly) إلى (1,0) على التوالي.

ثالثا: حساب مصفوفة التباين / التباين المشترك Variance/ Covariance Matrix للميزات الموجودة في كل سجل من نماذج التدريب.

رابعا: حساب متجه المميزة من مصفوفة التباين / التباين المشترك على النحو الآتي:

1. إيجاد أكبر عنصر في المصفوفة.
2. إيجاد زاوية التدوير.
3. إيجاد عناصر مصفوفة التدوير.
4. إعادة الخطوات من 1 إلى 3 على المصفوفة الناتجة من الخطوة السابقة حتى يتم الحصول على عناصر خارج القطر قريبة من الصفر.

خامسا: حساب قيمة متجه المميزة من المصفوفة الناتجة ووضعها في مصفوفة المميزة.

سادسا: ترتيب مصفوفة المميزة.

2-4 معالجة البيانات وتطبيعها

إن قيم الميزات (Protocol type, Service, Flag) تحتوي على بيانات حرفية فيتم تحويلها إلى بيانات عددية، وتتم عملية التحويل بإيجاد تكرارات كل ميزة في كل سجلات الاتصال ثم تعطى القيمة الجديدة ضمن المدى [1 .. عدد القيم ضمن الميزة]، إذ القيمة الأقل تكرار سوف تأخذ العدد 1 والقيمة الأكثر تكرار سوف تأخذ عدد مساوي إلى عدد القيم ضمن الميزة. كذلك تم تحويل حقل Attack type من (Normal, Anomaly) إلى القيم (1, 0) على التوالي.

تم تطبيع البيانات بطريقة Min-Max وتقوم هذه الطريقة بإجراء عمليات تحويل خطية على قيم البيانات الأصلية، ثم تطبق المعادلة الخطية على كل قيم الميزة X للحصول على القيمة الجديدة، المعادلة الآتية تستعمل لتطبيع بيانات التدريب والاختبار [2]:

$$X_n = (X - MinX) / (MaxX - MinX) \quad \ldots(14)$$

حيث إنّ MinX وMaxX هما اقل وأكبر قيمة للإدخال الأصلي X على التوالي، X_n هي ناتج عملية التطبيع.

بعد انتهاء المعالجة الأولية لبيانات نماذج التدريب يتم إجراء مرحلة التدريب.

3-4 التدريب باستخدام الخوارزمية الجينية

أولا: إدخال المجتمع الابتدائي.

ثانيا: حساب دالة التقييم لكل كروموسوم في المجتمع الابتدائي.

يتم ذلك بمقارنة كل كروموسوم بسجلات نماذج التدريب، إذا وجد تطابق بينهما يزداد سجل التطابق بمقدار واحد، وحسب المعادلة الآتية:

$$F = WT1 * \log (NA) + WT2 \quad \ldots(15)$$

حيث (NA) هي عدد تطابق الكروموسومات مع سجلات بيانات التدريب، WT1 وWT2 قيم ثابتة للدالة. إن قيم دالة التقييم تتراوح بين [0.0 – 1.0].

ثالثاً: الاحتفاظ بأفضل الآباء

رابعاً: عملية تكوين الأجيال الجديدة

- استخدام عجلة الروليت لاختيار أفضل الآباء الذين لديهم صلاحية عالية باحتمالية كبرى.
- تتم الطفرة بنسبة (5%)، وتستخدم طريقة التبادل بين قيم الكروموسوم باختيار عشوائي للكروموسوم واختيار عشوائي لاثنين من قيم الكروموسوم حيث يتم التبادل بينهما.
- حساب دالة التقييم للكروموسومات الجديدة وترتيبها تصاعدياً بعد إجراء عمليتي التزاوج والطفرة.

خامساً: اختبار شرط التوقف

يتحقق شرط التوقف بانتهاء عدد الأجيال أو يتم الوصول إلى عدد القواعد المطلوب توليدها، إذا لم يتحقق شرط التوقف يتم استبدال الكروموسومات السيئة للآباء بالكروموسومات الجيدة، ونسبة اختيار الكروموسومات الأفضل هي 10%.

وتتكرر عملية تكوين الأجيال الجديدة بالرجوع إلى الخطوة الرابعة إلى أن يتحقق شرط التوقف.

4-4 مرحلة الاختبار Testing Stage

بعد انتهاء مرحلة التدريب والحصول على قائمة قواعد التصنيف المكتشفة تبدأ مرحلة تصنيف نماذج الاختبار بإتباع ما يأتي:

أولا: تبدأ عملية التصنيف بإدخال مجموعة نماذج الاختبار (NSL-KDD).

ثانيا: إدخال قواعد التصنيف المكتشفة وعددها إلى النظام .

ثالثا: اختيار الميزات (41, 21, 10, 5) ميزة.

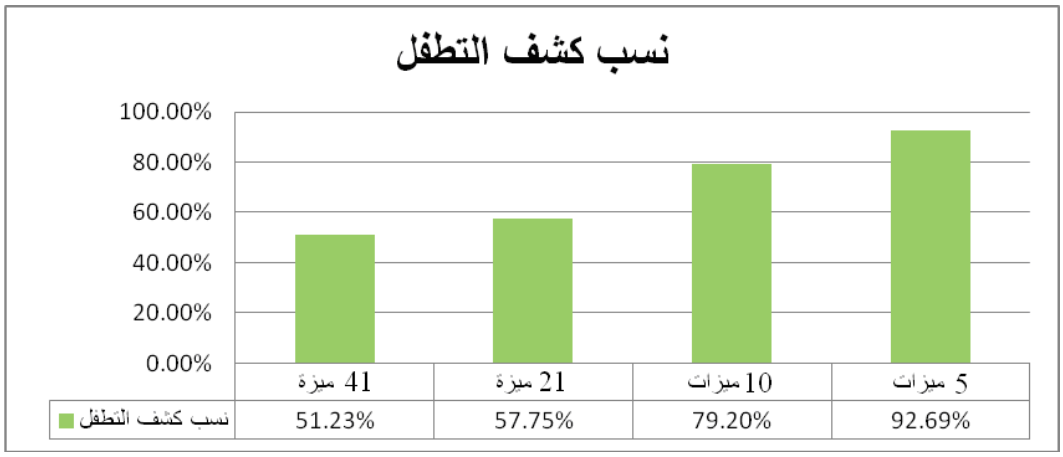
رابعاً: المعالجة الأولية لبيانات نماذج الاختبار ثم تطبيع البيانات باستخدام طريقةMin-Max Normalization

خامساً: في هذه المرحلة يتم اختبار كل نماذج الاختبار المدخلة إلى النظام، حيث إن كل أنموذج اختبار يقارن بقواعد التصنيف، فإذا وجد تطابق سوف يحدد نوع الهجوم (طبيعي أو شاذ).

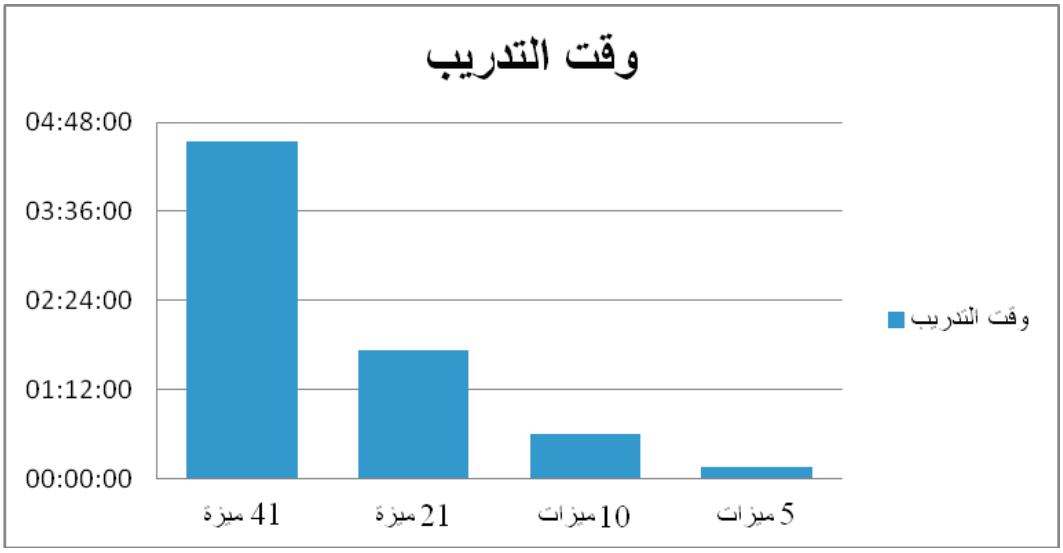
5. النتائج والاستنتاجات

استخدمت تقنيات تقليل الأبعاد كتحليل المركبات الأساسية مع بيانات NSL-KDD لتقليل الميزات، واعتمدت على طريقة جاكوبي في إيجاد قيمة المميزة، وأظهرت النتائج انجازاً جيداً لتحليل المركبات الأساسية لاختياره الميزات التي تحمل أعلى قيم في متجه المميزة. تم اختيار (5 ,10 ,21) ميزة، وبذلك تم تقليل تعقيد العمليات الحسابية.

تم مقارنة نسب كشف التطفل قبل تحليل المركبات الأساسية وبعده كما موضح في الشكل (2) وكذلك تم مقارنة وقت التدريب قبل PCA وبعده والشكل (3) يوضح ذلك حيث أن الميزات القليلة قللت وقت التدريب والاختبار.



الشكل (2). نسب كشف التطفل قبل تحليل المركبات الأساسية وبعده



الشكل (3). وقت التدريب قبل تحليل المركبات الأساسية وبعده

قورنت النتائج التي تم الحصول عليها من التجارب مع نتائج عدد من الباحثين الذين يعملون في المجال نفسه إذ إن الجدول (1) يبين أنّ تطبيق الخوارزمية الجينية وخوارزمية تحليل المركبات الأساسية باستخدام (5 ميزات) أعطت نتائج جيدة في كشف التطفل.

الجدول (1). مقارنة نتائج الخوارزمية الجينية مع عدد من الباحثين

Model	Data Base	DR %	False alarm%
SOM [15]	NSL-KDD	64	
K-Means [15]	NSL-KDD	60	
Online BPN[16]	KDD	91.50	2.68

Ant-Miner [2]	KDD	92.42	
PCA-GA [14]	NSL-KDD	92.69	0

من خلال تصميم النظام المقترح وتطبيقه على بيانات NSL-KDD، وبعد إجراء التجارب المختلفة لقياس كفاءة النظام وأداءه، تم استنتاج الآتي:

1. تم اختبار بيانات NSL-KDD التي حلت مشاكل KDD 99، وأوضحت التجارب أن هذه البيانات يمكن وضعها لمساعدة الباحثين لمقارنة مختلف نماذج كشف التطفل.
2. أوضحت التجارب أن النظام المقترح قادر على تسريع عمليات التدريب والاختبار لكشف وتصنيف التطفل والتي تعمل على زيادة سرعة تطبيقات الشبكة، كما قلل النظام المقترح وقت التدريب والاختبار.
3. النظام المقترح المعتمد على الخوارزمية الجينية مع PCA أسرع بالتدريب والاختبار من الخوارزمية الجينية بدون PCA.

6. التوصيات والأعمال المستقبلية

1. اعتماد قاعدة البيانات NSL-KDD، بالرغم من إنها لم تعط حلولاً لجميع المشاكل السابقة بصورة كاملة إلا أنها أثبتت فائدتها البحثية في بناء نماذج كشف التطفل على بيئة المحاكاة.
2. اختيار ميزات مناسبة لكل نوع من هجمات الشبكة.
3. تطوير النظام لكي يعمل على البيئة الحقيقية (On-line).
4. استخدام تقنيات أخرى في استخلاص الميزات واختيار الميزات مثل LDA, وIDA.

المصادر

[1] د.علاء حسين الحمامي، د.سعد عبد العزيز العاني، 2007، "تكنولوجيا أمنية المعلومات وأنظمة الحماية"، جامعة عمان الأهلية.

[2] Mahmood S. Mahmood, 2011, "Using Ant and Self-Organization Maps Algorithms To Detect and Classify Intrusion In Computer Networks", MSc. Thesis, Computer Science College, University of Mosul, Iraq.

[3] Gong RH, Zulkernine M and Abolmaesumi P., 2005, "A Software Implementation of a Genetic Algorithm based approach to Network Intrusion Detection". In: Proceedings of the sixth international conference on software engineering, artificial intelligence, networking and parallel/distributed computing and first ACIS international workshop on self-assembling wireless networks (SNPD/SAWN‘05).

[4] Randy L. Haupt and Sue Ellen Haupt, 2004 , "Practical Genetic Algorithms", Second Edition, A John Willey &Enetic Sons, Inc., Publication.

[5] Chittur A., 2001, "Model Generation for an Intrusion Detection System Using Genetic Algorithms", Ossining High School, NY.
http://www1.cs.columbia.edu/ids/publications/gaids-Thesis1.pdf

[6] Mrutyunjaya Panda, Ajith Abraham, Manas Ranjan Patra, 2010, "Discriminative Multinomial Naïve Bayes for Network Intrusion Detection.
http://www.softcomputing.net/ias10_panda.pdf

[7] Shilpa lakhina, Sini Joseph and Bhupendra Verma, 2010, "Feature Reduction Using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD", International Journal of Engineering Science and Technology, Vol. 2(6), 1790-1799.

[8] G. Meera Gandhi, Kumaravel Appavoo and S.K. Srivatsa, 2010,"Effective Network Intrusion Detection Using Classifiers Decision Trees and Decision rules", Int. J. Advanced Networking and Applications Volume: 02, Issue: 03, Pages: 686-692.

[9] Heba F. Eid, Ashraf Darwish, Aboul Ella Hassanien and Ajith Abraham, 2010," Principle Components Analysis and Support Vector Machine based Intrusion Detection System", 978-1-4244-8136-1/10 IEEE

[10] KDDCup 1999 Dataset. Available at:
http:// kdd.ics.uci.edu/databases/kddcup1999.html.

[11] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, 2009, "A Detailed Analysis of the KDD CUP 99 Data Set", Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA).

[12] Shilpa lakhina, Sini Joseph and Bhupendra Verma, 2010, "Feature Reduction Using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD", International Journal of Engineering Science and Technology, Vol. 2(6), 1790-1799.

[13] Ansam O. Abdul-Majeed, 2011, "New Steganographic Method for VQ-Compressed Images", MSc. Thesis, Computer Science College, University of Mosul, Iraq.

[14] Hana M. Osman, 2012, "Investigation of Applying Genetic Algorithm Based - Intrusion Detection and Classification System to NSL-KDD Dataset", MSc. Thesis, Computer Science College, University of Mosul, Iraq.

[15] Ritu Ranjani Singh, Prof. Neetesh Gupta, 2010, " To Reduce the False Alarm in Intrusion Detection System using self Organizing Map", International journal of Computer Science and its Applications.

[16] Ibraheem M. Ahmed Al-Haleema, 2011, "Development of Network-Based Intrusion Detection System Using Artificial Neural Networks", M.Sc. Thesis, Computer Science College, University of Mosul, Iraq.