Journal of University of Babylon for Pure and Applied Sciences, Vol. (27), No. (5): 2019

Design a Hybrid Algorithm Based on Tabu Search and Misuse Detection for Intrusion Dataset (KDD99 10%)

Yahya Ismail Ibrahim
Department of Computer Science, College of Education for Pure Sciences, University of Al Mosul, Ninava, Iraq
yahyaismail4@yahoo.com
Fawzia Mahmoud Remo
Department of Computer Science, College of Computer Science and Mathematics, University of Al Mosul, Ninava, Iraq
fawziyaramo@uomosul.edu.iq
Younis Samir Younis
Department of Computer Science, College of Computer Science and Mathematics, Ministry of Education Directorate of Education Nineveh, Ninava, Iraq
fajirnet1@yahoo.com

ARTICLE INFO
Submission date: 10/4/2019
Acceptance date: 19/5/2019
Publication date: 1/10/2019

Abstract

The growth and development of the Internet has led to the rapid expansion of Local network systems and World Wide Web, has reflected a very large increase in Internet users. Data and computer hardware are becoming more relevant to the global network, making it more vulnerable to attackers. so there is an urgent need to further protect the Devices and data users of attacks or unauthorized access. Including the detection and classification of data transmitted over the World Wide Web, especially that connect users' devices to prevent illegal access. In this sense, the research was based on the use of one of the smart algorithms of meta heuristic algorithm (Tabu) to find the most efficient rules in the detection process and use the best ways to neglect the repeated rules, which helps to reduce the storage space and reduce the time addition. Algorithm was hybridized with the artificial intelligence (Fuzzy Logic) Which in turn can reduce the rules required to hybridize in order to obtain new rules. Experiments were conducted on dataset (KDD99 10%) for detection and determination through algorithm if dataset is normal or attacks, Thus, this algorithm gave the results of a Detection Rate of 87.7%, a Detection accuracy of 96.36%, and the time taken for the results obtained was 13 minutes and 26 seconds.

Keywords: - Tabu algorithm, Misuse Detection, Detection Accuracy, Data Normalization, stochastic, Anomaly Detection.

تصميم خوارزمية هجينة بالاعتماد على خوارزمية تابو وتقنية كشف سوء الاستخدام على بيانات (KDD99 10%)

يحيى اسماعيل ابراهيم

قسم علوم الحاسوب، كلية التربية للعلوم الصرفة، جامعة الموصل، نينوى، العراق

yahyaismail4@yahoo.com

فوزية محمود رمو

قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، نينوى، العراق

fawziyaramo@uomosul.edu.iq

يونس سمير يونس

قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، وزارة التربية/مديرية تربية نينوى، نينوى، العراق

fajirnet1@yahoo.com

الخلاصة

ادى النمو والتطور في شبكة الانترنت، الى سرعة التوسع في انظمة الشبكات المحلية وشبكة الويب العالمية، وانعكس على زيادة كبيرة جدا لمستخدمي شبكة الانترنت، وبهذا اصبحت البيانات واجهزة الحاسوب ذات اهمية أكبر المرتبطة بالشبكة العالمية التي جعلتها أكثر عرضة للمهاجمين، لهذا فهنالك حاجة ملحة لزيادة حماية اجهزة وبيانات المستخدمين من الهجمات. ومنها كشف وتصنيف للبيانات المنقولة عبر الشبكة العالمية وبالأخص التي تصل اجهزة المستخدمين لمنع الوصول غير المشروع. ومن هذا المنطلق استند هذا البحث على استخدام احدى الخوارزميات الذكائية المتمثلة بخوارزمية ما بعد الحدسية (تابو) لإيجاد القواعد الاكثر كفاءة في عملية الكشف واستخدام افضل الطرق لإهمال القواعد المتكررة مما يساعد على تقليل مساحات الخزن وتقليل الوقت مضاف الى ذلك تم تهجينها مع الذكاء الصناعي المتمثل بالمنطق المضبب الذي بدوره مكن من اختزال القواعد المطلوبة للقيام بتهجينها من اجل الحصول على قواعد جديدة ، وتم اجراء التجارب على بيانات KDD99 10% لإجراء عمليات الكشف وتحديد من خلال الخوارزمية فيما لو كانت البيانات طبيعية ام هجمات، وبهذا أعطت هذه الخوارزمية نتائج معدل كشف (Detection Rate) ما مقداره (87.7%) ودقة كشف (Detection accuracy) ما مقداره (96.36%) والوقت المستغرق للنتائج التي تم الحصول عليها 13 دقيقة و 26 ثانية.

الكلمات الدالة: – خوارزمية تابو، اساءة الاستخدام، دقة الكشف، تطبيع البيانات، العشوائية، كشف الشذوذ

1-المقدمة

ان التقدم الحاصل والتوسع الكبير في شبكة الحاسبات ومنظوماتها وزيادة اهميتها واهمية موارد المعلومات التي تملكها، جعلت منها السبب الرئيسي لتكون شبكة الحواسيب عرضة لجرائم الحاسوب المتزايدة اكثر واكثر، ولما يعتمده مستخدم الانترنت على شبكة الحواسيب بشكل كبير لجميع المجالات المتمثلة بالمصارف والاتصالات الداخلية والخارجية المغلقة والاسواق التجارية من شبكة الانترنت بشكل دوري يوميا والتي تحتاج الى عمليات انتقال امينة ضمن الشبكة الحاسوبية والتي تحتاج لوصول امين عن طريق كلمات المرور التي تحمي المواقع والحسابات الشخصية. وبهذه الزيادة من عمليات التنقل خلال شبكة الانترنت بشكل شامل سوف تكون البيانات المهمة معرضة للهجمات وصولا للتدمير [1]. نظام كشف التطفل يتكون من صنفين رئيسيين

أ – كشف إساءة الاستخدام ((Misuse Detection: تعمل على نماذج تدخلات معلومة ويتم الكشف عن الاختراقات عن طريق المطابقة وبهذا يكتشف الهجوم، حيث ان كشف الاساءة الاستخدام يتم بمطابقة البيانات المراقبة مع معلومات التطفل المخزونة سابقا، بهذه الحالة الهجمات المعروفة مباشرة تكتشف بعد حدوثها مع وجود نسب قليلة جدا من الانذارات الكاذبة (False Alarm) [2],[1].

ب – كشف الشذوذ (Anomaly Detection): بهذه الطريقة من الكشف يتم متابعة التغيرات الكبيرة والغير مألوفة في نماذج الاتصال، ليتكون ملف للتعريف العادي يمتلك مقاييس مأخوذة من انظمة التشغيل. ليتم بعدها مراقبة النظام للاتصالات واجراء المقارنة مع الملفات التعريفية لآجل اكتشاف اي تغييرات في الانماط المستخدمة وسلوك النظام [2],[1]. بحيث هنالك قدرة لكشف الشذوذ لتمييز انواع التطفل الجديدة، في هذا النوع من الانظمة يتم اختبار او الكشف عن حزم البيانات المشوهة او المشبوه ولا تكون عملية الكشف محددة ولكن مجرد تحذر من انها حزمة غريبة [3],[1].

2- الدراسات السابقة

في عام 2011 اعتمد الباحث (محمود صبحي محمود) استخدم بها خوارزمية النمل واستند بهذه الخوارزمية على بيانات KDD99 10% واظهرت نتائج معدل كشف 92.42% [2],[3].

في عام 2013 اعتمد الباحثون (LAHEEB M. IBRAHIM, DUJAN T. BASHEER, MAHMOD S. MAHMOD) نموذجا مبتكرا لنظام كشف التطفل استخدم خوارزمية Self-organization map(SOM) واستخدم بذلك بيانات KDD99 10% وبهذا اظهرت نتائج معدل كشف 92.37%ومعدل كشف خاطئ % 5.77 [4].

3-خوارزمية البحث الممنوع Tabu Search Algorithm

من احدى الخوارزميات التي يطلق عليها ما بعد الحدسية هي خوارزمية البحث الممنوع والتي ساعدت على معالجة المسائل ذات الاهمية العملية. وامتدت او توسعت تطبيقات خوارزمية البحث الممنوع الى عدت مجالات ومن امثلتها تحليل الاقتصاد، الاتصالات، تخطيط المصادر، توزيع الطاقة ومجالات مختلفة اخرى في حياتنا اليومية الحالية. وتعني كلمة Tabu الاشياء المخيفة او المقدسة التي لا يجوز الاقتراب منها [5],[1].

طريقة عمل هذه الخوارزمية يكون على قيامها ببحث محلي متقدم لاستكشاف نطاق الحلول حتى يتم الوصول للحل الامثل. تعتمد خوارزمية البحث الممنوع بحل المشاكل عند البحث بنطاق الحلول بشكل اقتصادي وبفعالية.

وممكن تطبيق البحث الممنوع مباشرة لجميع انواع المشاكل التي تحتاج لتحسين الحل. ممكن تمثيل المشاكل بالشكل التالي، حيث (optimize) يعني تقليل او تكبير:

$$\text{Optimize } f(x) \quad \text{where} \quad x \in X$$

دالة f(x) ممكن أن تكون خطية، غير خطية، أو عشوائية (stochastic)، المتغير X عبارة عن متجه يمثل مجموعة متغيرات لقيم x التي يتم اختبارها باستخدام معادلة optimize [6].

4-المنطق المضبب Fuzzy Logic

المنطق المضبب يعمل بشكل مشابه لعمل العقل البشري مع التطبيقات اي انه يمتلك قيم منطقية بين [1,0]، حيث انه يعمل على استقبال المعلومات من التقييم او المقاييس. من المعروف لدى علماء الحاسوب الوصف يكون بـ 0 و1 أي صح او خطا. وتعمل الانظمة الخبيرة وتطبيقات الذكاء الاصطناعي على المنطق المضبب، عندما كان هناك حاجة لإيجاد طرق افضل لمعالجة البيانات، وعندما لاحظ انه لا يمكن التعامل مع مختلف الاشكال المنطقية بحالة [1,0] وبالأخص للمشاكل التي تحدث حاليا عندما يكون هناك حالات ممكن اعتبارها صحيحة واحيان اخرى خاطئة او تمتلك نسبة صحة مختلفة عن الاخرى، وان أي تغيير بقيم الادخال وان كان صغير يؤدي الى زيادة بالتغيير ولكن ليس تغيير تام ، ويستخدم بذلك دالة العضوية Membership Function وهي الدالة التي تمثل نسبة انتماء العنصر للمجموعة [7],[1]. وبسبب التطور والحاجة لتطبيقات المنطق المضبب صنعت شريحة المنطق المضبب Fuzzy Logic Chip حيث استخدمت بعديد من المنتجات كالات التصوير والتي تعمل مع بيانات غير دقيقة، لهذا المنطق المضبب يعتبر العنصر الاساسي الذي تعمل عليه منظومات الذكاء الاصطناعي والانظمة الخبيرة [8].

5-مقاييس تقييم أداء أنظمة كشف التطفل

نسبة كشف التطفل Detection Rate (DR) تعتمد عليها انظمة كشف التطفل المختلفة لتقييم كفاءة نظام الكشف، والمتمثل بتحديد او اكتشاف مرور الحزم الاعتيادية من الهجوم، وايضا نسبة الانذارات الكاذبة False Alarm Rate(FAR) (التي تمثل مرور الحزم الاعتيادية والاشتباه بانها هجوم) يوضح الجدول (1) مقياس تقييم كفاءة كشف التطفل [1],[9],[14].

الجدول (1) مقاييس تقييم أداء أنظمة كشف التطفل

المعايير القياسية		الصنف المتنبئ للاتصال	
		اعتيادي	هجوم
الصنف الفعلي للاتصال	اعتيادي	(TN)	(FP)
	هجوم	(FN)	(TP)

إذ أن:

- TN (True Negative): يمثل عدد الحزم الاعتيادية مكتشفة بشكل صحيح.
- FP (False Positive): يمثل عدد الحزم الاعتيادية مكتشفة على أنها هجوم.
- TP (True Positive): يمثل عدد حزم الهجوم مكتشفة بشكل صحيح.
- FN (False Negative): يمثل عدد حزم الهجوم مكتشفة على أنها اتصالات اعتيادية.

* المعادلات المستخدمة لقياس أداء نظام كشف التطفل[1] ,[11],[10]:

نسبة الكشف (Detection Rate (DR)) $\frac{TN+TP}{\text{العدد الكلي لبيانات الاختبار}}$ % ...= (1)

نسبة الإنذار الكاذب الايجابي= $\frac{FP}{FP+TN}$ %...(2)

نسبة الإنذار الكاذب السلبي= $\frac{FN}{FN+TP}$ % ...(3)

نسبة الاتصالات الغير طبيعية المكتشفة= $\frac{TP}{\text{العدد الكلي لسجلات الهجوم في بيانات الاختبار}}$ %... (4)

نسبة الاتصالات الطبيعية المكتشفة= $\frac{TN}{\text{العدد الكلي لسجلات الاعتيادية في بيانات الاختبار}}$ %...(5)

نسبة الاتصالات الغير مكتشفة = (مجموع الاتصالات الكلي –الاتصالات المكتشفة) %... (6)

دقة الكشف= $\frac{\text{اتصالات اعتيادية مكتشفة+اتصالات هجوم مكتشفة}}{\text{اتصالات اعتيادية مكتشفة+اتصالات اعتيادية مكتشفة على أنها هجوم+اتصالات هجوم مكتشفة على أنها اعتيادية+اتصالات هجوم مكتشفة}}$ % .... (7)

6- خوارزمية البحث الممنوع المهجنة

استخدم لكشف التطفل الشبكي والمستند على كشف اساءة الاستخدام (Misuse Detection) على خوارزمية البحث الممنوع المبتكرة مدمجة مع المنطق المضبب وطبقت على بيانات KDD99 10% وبعدد نماذج التدريب والاختبار (805050) [1],[12]. تتألف الخوارزمية المبتكرة من مجموعة من الخطوات التالية: -

أولاً- إدخال مجموعة بيانات KDD99 10% واستخدام نماذج التدريب والتي عددها (494021) نموذج, ونماذج الاختبار والتي عددها (311029) نموذج ويحتوي كل سجل على 41 ميزة كما مبين في الشكل (1) [12].

نوع البيانات	نموذج بيانات التدريب المدخلة قبل المعالجة
Normal	0,udp,other,SF,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,13,1,0.00,0.00,0.00,0.00,0.08, 0.15,0.00,255,1,0.00,0.60,0.88,0.00,0.00,0.00,0.00,0.00
Attack	0,tcp,private,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,123,6,1.00,1.00,0.00,0.00,0.05, 0.07,0.00,255,26,0.10,0.05,0.00,0.00,1.00,1.00,0.00,0.00

شكل (1) نموذج بيانات KDD99 10%

ثانيا– إجراء المعالجة الأولية للبيانات (Data preprocessing)

أ- تعالج البيانات التي تمتلك بيانات حرفية وتبديلها إلى بيانات عددية، ويتم على مميزات (protocol, service, flag) ويكون ذلك التحويل بإيجاد تكرار كل عنصر في كل ميزة في جميع سجلات الاتصال ثم تعطى قيمة لكل عنصر في الميزة ضمن المدى [1.. عدد القيم ضمن الميزة]، إذ العنصر الأكثر تكرار يأخذ العدد 1 والعنصر الأقل تكرار يأخذ عدد مساوي إلى عدد القيم ضمن الميزة.

ب- تطبيع البيانات ((Data Normalization

بعد تغيير جميع البيانات الحرفية إلى بيانات عددية، يتم تطبيع البيانات لتقع ضمن مدى معين بالاعتماد على طريقة Min-Max Normalization حيث البيانات تقع ضمن المدى [0.0–1.0]. كما في المعادلة التالية [8].

$$X_n = (X - MinX) / (MaxX - MinX) \quad ....(7)$$

بعد تطبيق المعادلة (7) نحصل على النموذج الموضح في شكل (2)

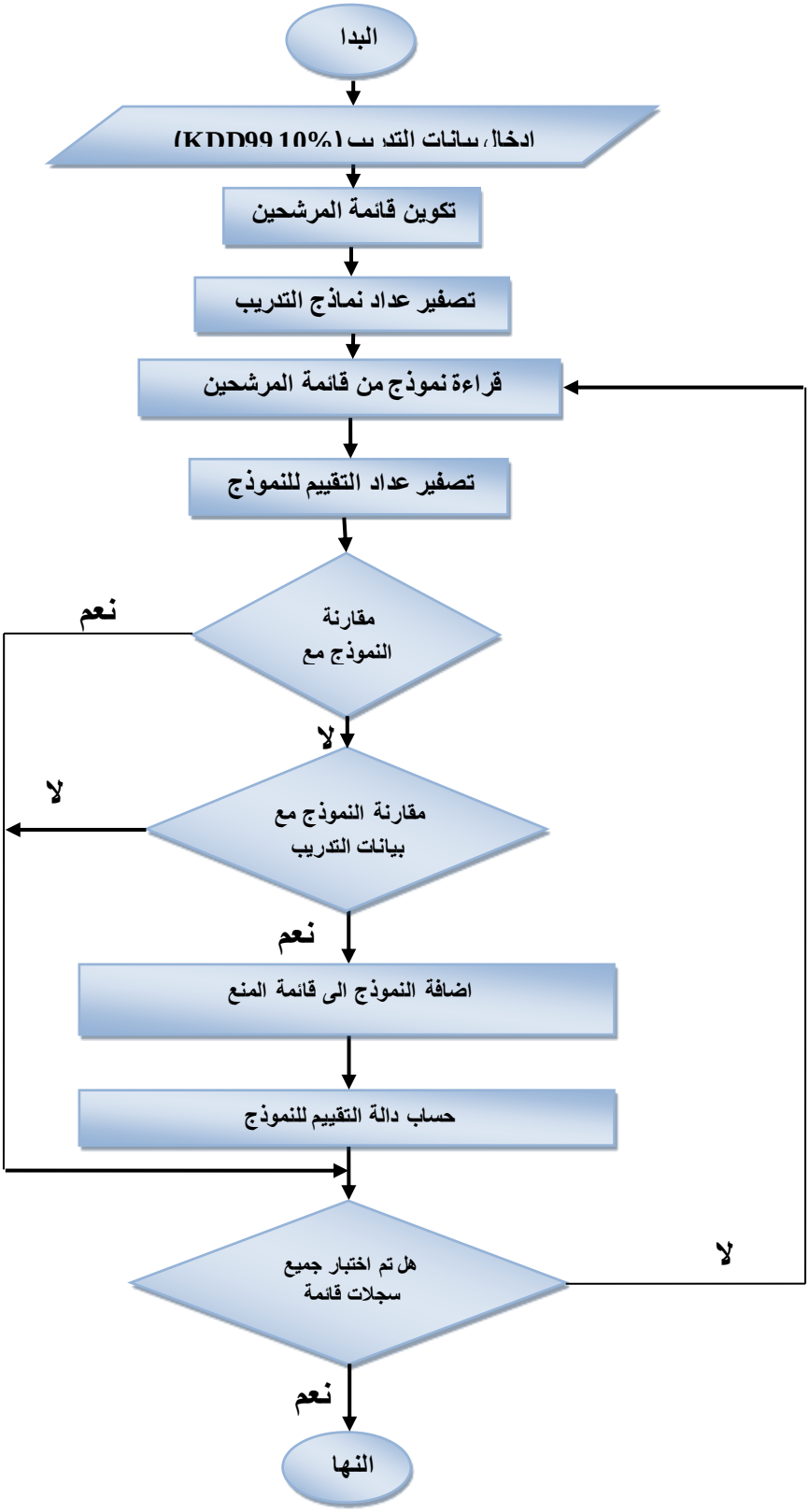
نوع البيانات	نموذج بيانات التدريب بعد عملية التحويل والتطبيع
Normal	0,0.5,0.1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.1,0.1,0,1,0,0,0.6, 0.9,0,0,0,0,0
Attack	0,0,0,0.1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.2,0,1,1,0,0,0.1,0.1,0,1,0.1,0.1, 0.1,0,0,1,1,0,0

شكل (2) نموذج بيانات KDD99 10% بعد عملية التحويل والتطبيع

ثالثاً– مرحلة التدريب

تتألف عملية التدريب لقائمة قواعد التصنيف المكتشفة من مرحلتين: –

المرحلة الاولى: – اعتماد خوارزمية البحث الممنوع بعملية التدريب لاكتشاف قواعد التصنيف كما مبين بالمخطط الانسيابي في الشكل (3).



الشكل (3): المخطط الانسيابي لعملية التدريب باستخدام خوارزمية البحث الممنوع

واعتمد بهذا (494021) نموذج والمتمثلة بنماذج بيانات التدريب وبناء قائمة مرشحين التي تتألف من 3000 نموذج وكل نموذج منها يمتلك 41 ميزة، ويتم انشاء قائمة المنع التي تكون بالبداية فارغة بعدها يضاف

لها جميع القواعد المكتشفة من قائمة المرشحين. يؤخذ في كل مرة نموذج من قائمة المرشحين، ويفحص النموذج مع قائمة المنع ان وجد النموذج ضمن قائمة المنع يهمل النموذج ويعتبره قاعدة مكتشفة سابقا ويجب اختبار جميع عناصر قائمة المرشحين[13] . في حالة عدم انهاء فحص نماذج قائمة المرشحين يتم استدعاء المرشح التالي واختباره مع قائمة المنع وبحالة عدم ايجاده في قائمة المنع يؤكد ذلك ان النموذج لم يكتشف بالسابق ويستخدم لاختباره مع بيانات التدريب وبحالة عدم تطابقه مع اي من بيانات التدريب سوف يهمل ويتم الانتقال للنموذج التالي وبحالة ظهوره سوف يضاف الى قائمة المنع، وتحسب دالة التقييم للنموذج المكتشف والمتمثلة بعد مرات تكرار النموذج ببيانات التدريب، ويتم تكرار هذه العملية حتى تفحص جميع نماذج قائمة المرشحين.[15]

المرحلة الثانية: –

تستخدم بهذه المرحلة خوارزمية المنع المدمجة مع المنطق المضبب في التدريب وكما مبين بالشكل (4).

يبين المخطط الانسيابي اختيار القاعدة بالاستناد على دالة اللياقة للنماذج بحيث انها لم تستخدم سابقا. بعدها يطبق المنطق المضبب واعتمدت بهذا النظام دالة العضوية (Membership Function) ذات الشكل المثلثي (Triangular), ويتم تخزين افضل قيم الادخال بالاستناد على الاخراج ليتم بعدها مقارنة النتائج المخرجة بحد التعبة (وقيمته =0.6) فاذا كانت قيمه اقل منه نستخدم القاعدة المختارة بانشاء قواعد جديدة التي تمثل قائمة المرشحين، وبخلاف ذلك يتم استخدام احدى القواعد التي استخدمت في انشاء القواعد بالدورات السابقة وعلى ان تكون قيمة العتبة2 للقاعدة اقل من (3) وتستمر حتى تحقق شرط التوقف[1],[12].

وبعد اختيار القاعدة التي ستستخدم من خلالها لإنشاء مجموعة سجلات جديدة لقائمة المرشحين لاكتشاف قواعد جديدة يستمر تكوين قائمة المرشحين بكل مرة تفحص جميع سجلاتها بالاستناد على عدد الدورات التي يحددها النظام.

رابعاً – مرحلة الاختبار

تبدأ هذه المرحلة بعد اكتشاف القواعد في مرحلة التدريب واختبارها على مجموعة نماذج الاختبار (Testing dataset) ويتم ذلك بعد اجراء المعالجة المبدئية (المتمثلة بالتبديل والتطبيع) حتى تصبح النماذج مطابقة مع القواعد المكتشفة. ليتم بعدها التأكد من كفاءة النظام عن طريق ايجاد نسب كشف التطفل، ونسب الانذار السلبي، والانذار الكاذب، ونسب كل من الاتصالات الطبيعية والغير طبيعية المكتشفة.

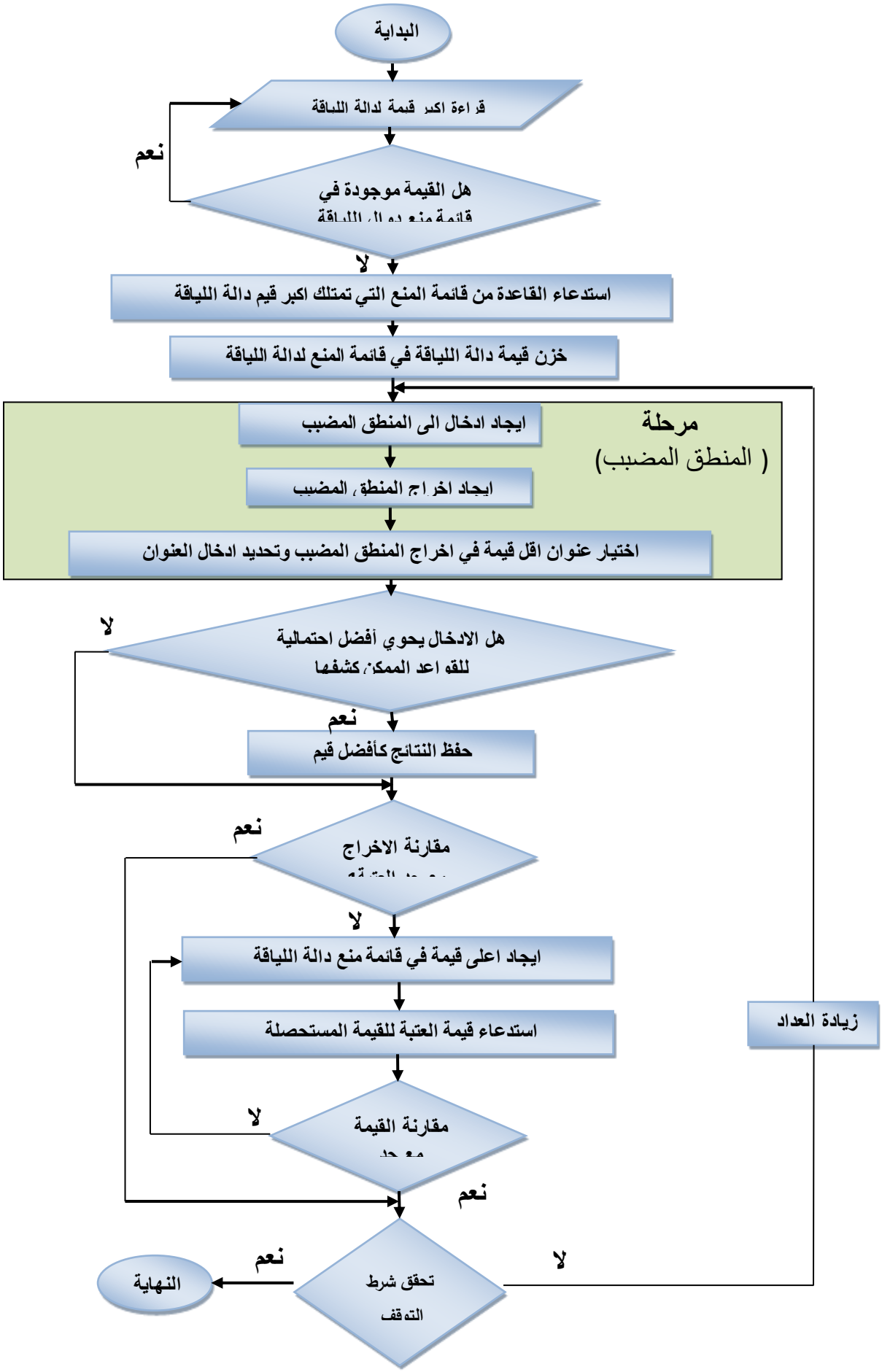
7– التجارب المنجزة

اجريت تجارب التدريب والاختبار لنظام كشف التطفل المعتمد على KDD99 10% Dataset والمتمثلة بـ (494021) نموذج لبيانات التدريب بمرحلة تدريب النظام، والجدول التالي يبين عدد سجلات بيانات التدريب واصنافها.

الجدول (2) عدد السجلات الموجودة في بيانات التدريب KDD99 10% وأصنافها

KDD99 10% DataSet	السجلات الطبيعية	سجلات الهجوم
عدد السجلات	97278	396743
النسبة المئوية	19.69%	80.31%

في حين بيانات الاختبار، بلغ عددها (311029) سجل اتصال. والجدول (3) يبين عدد سجلات بيانات الاختبار وأصنافها.



الشكل (4): مخطط انسيابي يمثل مرحلة التدريب باستخدام خوارزمية البحث الممنوع المهجنة مع المنطق المضبب

الجدول (3) عدد السجلات الموجودة في بيانات الاختبار KDD99 10% وأصنافها

KDD9910% Dataset	السجلات الطبيعية	سجلات الهجوم
عدد السجلات	60593	250436
النسبة المئوي	19.48 %	80.52%

التجربة الاولى

أولا: مرحلة التدريب.

باستخدام الخوارزمية المبتكرة بالتجربة الاولى اعتمدت قيم المعاملات بمرحلة التدريب للنظام بالشكل الاتي:

- حجم قائمة المرشحين = 3000
- عدد الدورات = 100
- حجم قائمة المنع التي تحوي القواعد المكتشفة = 7000

استغرق الوقت لعملية التدريب لدورات المئة (7 دقيقة و13 ثانية) لتكون القواعد المكتشفة كما مبين في الجدول (4): -

الجدول (4) القواعد المتولدة

الوصف	العدد	بيانات مكتشفة بالتدريب	نسبة مكتشفة من بيانات التدريب
القواعد الاعتيادية المكتشفة	**2223**		
قواعد الهجمات المكتشفة	**1196**		
اجمالي القواعد المكتشفة	**3419**		
الاتصالات الطبيعية	**97278**	**76216**	**78.4%**
الاتصالات الغير طبيعية	**396743**	**391698**	**98.7%**

ثانياَ: مرحلة الاختبار

باختبار النظام لبيانات الاختبار بالاستناد على قواعد المكتشفة بمرحلة تدريب النظام، تم الحصول على نتائج الاختبار كما في الجدول (5).

الجدول (5) نتائج مرحلة الاختبار

النتائج		
الوصف	**العدد**	**النسبة**
الاتصالات الطبيعية المكتشفة	**50727**	**83.7%**
الاتصالات الغير طبيعية المكتشفة	**220782**	**88.2%**
نسبة الكشف (DR)	**271509**	**87.3%**
الاتصالات الغير المكتشفة	**39520**	**12.7%**
الإنذار الكاذب الايجابي	**6**	**0.01%**
الإنذار الكاذب السلبي	**10311**	**4.46%**
دقة الكشف		**96.34%**

التجربة الثانية

أولا: مرحلة التدريب.

باستخدام الخوارزمية المبتكرة بالتجربة الثانية اعتمدت قيم المعاملات بمرحلة التدريب للنظام بالشكل الاتي:

- حجم قائمة المرشحين = 3000
- عدد الدورات = 200
- حجم قائمة المنع التي تحوي القواعد المكتشفة = 7000

استغرق الوقت لعملية التدريب للدورات المئتان (13 دقيقة و26 ثانية) لتكون القواعد المكتشفة كما مبين في الجدول (6): -

الجدول (6) القواعد المتولدة

الوصف	العدد	بيانات مكتشفة بالتدريب	نسبة مكتشفة من بيانات التدريب
القواعد الاعتيادية المكتشفة	**2402**		
قواعد الهجمات المكتشفة	**1394**		
اجمالي القواعد المكتشفة	**3796**		
الاتصالات الطبيعية	**97278**	**77162**	**79.3%**
الاتصالات الغير طبيعية	**396743**	**392558**	**99%**

ثانياَ: مرحلة الاختبار

باختبار النظام لبيانات الاختبار بالاستناد على قواعد المكتشفة بمرحلة تدريب النظام، تم الحصول على نتائج الاختبار كما في الجدول (7).

الجدول (7) نتائج مرحلة الاختبار

النتائج		
الوصف	**العدد**	**النسبة**
الاتصالات الطبيعية المكتشفة	**50957**	**84.1%**
الاتصالات الغير طبيعية المكتشفة	**221700**	**88.5%**
نسبة الكشف (DR)	**272657**	**87.7%**
الاتصالات الغير المكتشفة	**38372**	**12.3%**
الإنذار الكاذب الإيجابي	**20**	**0.04%**
الإنذار الكاذب السلبي	**10295**	**4.4%**
دقة الكشف		**96.36%**

التجربة الثالثة: –

أولا: مرحلة التدريب.

باستخدام الخوارزمية المبتكرة بالتجربة الثالثة اعتمدت قيم المعاملات بمرحلة التدريب للنظام بالشكل الاتي:

- حجم قائمة المرشحين = 3000
- عدد الدورات = 400
- حجم قائمة المنع التي تحوي القواعد المكتشفة = 7000

استغرق الوقت لعملية التدريب للدورات الاربع مئة (16 دقيقة و22 ثانية) لتكون القواعد المكتشفة كما مبين في الجدول (8): –

الجدول (8) القواعد المتولدة

الوصف	العدد	بيانات مكتشفة بالتدريب	نسبة مكتشفة من بيانات التدريب
القواعد الاعتيادية المكتشفة	2753		
قواعد الهجمات المكتشفة	1414		
اجمالي القواعد المكتشفة	4167		
الاتصالات الطبيعية	97278	78575	80.8%
الاتصالات الغير طبيعية	396743	392555	99%

ثانياَ: مرحلة الاختبار

باختبار النظام لبيانات الاختبار بالاستناد على قواعد المكتشفة بمرحلة تدريب النظام، تم الحصول على نتائج الاختبار كما في الجدول (9).

الجدول (9): نتائج مرحلة الاختبار

النتائج		
الوصف	العدد	النسبة
الاتصالات الطبيعية المكتشفة	51378	84.8%
الاتصالات الغير طبيعية المكتشفة	221679	88.5%
نسبة الكشف (DR)	273057	87.8%
الاتصالات الغير المكتشفة	37972	12.2%
الإنذار الكاذب الإيجابي	15	0.03%
الإنذار الكاذب السلبي	10433	4.5%
دقة الكشف		96.32%

8- مناقشة النتائج

بعد الحصول على النتائج المستحصلة من الخوارزمية المبتكرة ومقارنتها مع النتائج لعدد من الباحثين الذين يعملون بنفس المجال واعتماد نفس نوعية البيانات KDD99 10% وايضا نفس كمية البيانات بشكل كامل بدون استقطاع او اختزال في البيانات. حيث وضعت النتائج في الجدول (10) وجد ان الخوارزمية قدمت نتائج كفؤة لكشف التطفل من ناحية الكشف والفترة الزمنية ويضاف الى ذلك ان الاختبار الثالث الذي يشمل 400 دورة ان النظام توقف عند 254 بعد عدم حصوله على قواعد جديدة مما وفر بالوقت، اضافة الى ذلك من نتائج الاختبار للتجارب المذكورة تبين ان معدل الكشف متقارب جدا حيث ان في 100 دورة نسبة الكشف كانت 87.3% ودقة

الكشف 96.34% وعند تحديد 400 دورة كانت نسبة الكشف 87.8% ودقة الكشف 96.32% وما يثبت ان الخوارزمية لها امكانية بإيجاد افضل القواعد التي لها القدرة على الكشف على اكبر عدد ممكن من البيانات بعدد دورات قليلة وفترة زمنية قياسية واظهر بزيادة عدد الدورات سوف يتم الحصول على قواعد جديدة ولكنها تؤثر على دقة الكشف بشكل سلبي وان افضل النتائج ظهرت عند اختبار النظام على 200 دورة من ناحية معدل الكشف ودقة الكشف والفترة الزمنية.

الجدول (10) مقارنة نتائج الخوارزمية المبتكرة مع عدد من الباحثين

Model	Dataset	DR %	FP %	% ACC	Time(H:M:S)
Adaboost [13]	KDD99 10%	90.88	1.79	*	*
Ant-Miner [2]	KDD99 10%	92.42	*	*	*
SOM [4]	KDD99 10%	92.37	4.67	*	00:35:00
Proposed	KDD99 10%	87.7	0.03	96.36	00:13:26

9- الاستنتاجات

تبين من استخدام خوارزمية تابو لإيجاد طريقة أفضل لتحديد أفضل القواعد اثناء عملية التدريب للبيانات من خلال تقليل مساحات الخزن المطلوبة لاختيار القواعد ذات الكفاءة العالية واهمال القواعد التي تم ايجادها سابقا عند الحصول عليها مما ادى بدوره بتقليل الوقت المطلوب لإجراء عمليات التدريب مضاف لذلك ايجاد النتائج اثناء عمليات الاختبار. تبع بعد ذلك استخدام المنطق المضبب والذي ظهر بشكل واضح من حيث استهداف معالجة بيانات كبيرة ومعقدة منها التي اعتمدت بالبحث، حيث بوجوده اعطى امكانية لاختيار خصائص تمكن من ايجاد نسبة كشف اعلى باتخاذ القرار الصائب بتحديد نوعية القواعد المطلوبة، وبهذا وفرت امكانية تهجين المنطق المضبب مع خوارزمية تابو على زيادة الدقة باختيار القواعد مما يؤدي لتقليل الوقت بالتدريب والاختبار والذي اضاف بذلك وقت اقل لما وفرته خوارزمية تابو من تقليل الوقت بالاختبار والتدريب. وبهذا شملت عملية التهجين الى تقليل باستهلاك المصادر المادية من ناحية المعالجة والخزن ومتبوع بزيادة القواعد التي توفر نسبة كشف اعلى وقلة القواعد عدد القواعد بالوقت نفسه واخيرا الدقة باختيار القواعد التي تعطي اعلى نسبة كشف. ظهر هذا الامر بشكل واضح اثناء عمليات التدريب والاختبار على بيانات KDD99 10% التي اعطت نتائج متميزة على الرغم من الحجم الكبير لهذه البيانات وبهذا نتائج التجارب ممكن ان تساعد الباحثين لمقارنة مختلف نماذج الكشف للبيانات.

Conflict of Interests.

There are non-conflicts of interest .

المصادر

1. فوزية محمود رمو، "نظام ذكائي هجين لكشف التطفل باستخدام بيانات NSL−KDD"، *مجلة الرافدين لعلوم الحاسوب والرياضيات*، ISSN 1815-4816، 2013.
2. كرم محمد مهدي صالح الرشيدلي، " تصميم وتنفيذ نظام كشف التطفل باستخدام التقنيات الذكائية"، رسالة ماجستير هندسة برمجيات، جامعة الموصل، 2011.
3. S. Kaplantzis, N. Mani, "A study on classification techniques for network intrusion detection". In P. Prapinmonkolkarn, & T. Angkaew (Eds.), *Proceedings of the IASTED International Conference on Networks and Communication Systems*, Vol. 1, pp. 352 - 357, 2006.
4. L. M. Ibrahim, D. T. Basheer, M. S. Mahmod, "A Comparison Study For Intrusion DATABASE (KDD99, NSL-KDD) Based on Self Organization Map (SOM) Artificial Neural Network", *Journal of Engineering Science and Technology*, vol. 8, no. 1, pp. 107-119, 2013.
5. Wassim Jaziri, Ed., *Local Search Techniques: Focus on Tabu Search*, Vienna, Austria, In-Teh is Croatian branch of I-Tech Education and Publishing KG, 2008.
6. F. Glover, M. LAGUNA, R. MARTÍ, *Principles of Tabu Search*, Boulder,Colorado (USA) ,University of Colorado, 2006.
7. S. Bydoń,"Supervisory fuzzy controller for linear control system", M.S. thesis MSc., Univ. of Mining and Metallurgy, Poland, 2001.
8 .M. A. Yosif ," Use of a Clustering Method and Fuzzy Logic for Classification of Tissue Images", *Tikrit Journal of Pure Sciences*, vol. 16, no. 3, pp. 268-277 , 2011.
9. P. Winter, E. Hermann, M. Zeilinger," Inductive Intrusion Detection in Flow based Network Data using One class Support Vector Machine", *2011 4th IFIP International Conference on New Technologies, Mobility and Security*, 2011.
10. D. Dasgupta, D. Ferebee, "Enhancing Computer Security With Smart Technology [Book Review]", *IEEE Computational Intelligence Magazine*, vol. 3, no.2, pp. 70-71, april 2008.
11. M. Gyanchandani, R. N. Yadav, J. L. Rana, "Intrusion Detection using C4.5: Performance Enhancement by Classifier Combination", ACEEE Int. J. on Signal & Image Processing, vol. 1, no. 3, Dec. 2010.
12. KDD Cup 1999 DataSet, *kdd.ics.uci.edu*, University of California, Irvine, 1999, [Online]. Available: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99, [Accessed: October 28, 1999].
13. J. Gupta, J. Singh, "Detecting Anomaly Based Network Intrusion Using Feature Extraction and Classification Techniques", *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, pp. 1453-1456, May/Jun 2017.
14. W. Hu, Wei Hu, S. Maybank, "AdaBoost-Based Algorithm for Network Intrusion Detection IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 38 , no. 2, pp. 577-583, April 2008.
15. S. Devaraju, S. Ramakrishnan, "Performance Comprarison for Intrusion Detection System Using Neural Network with KDD DATASET", *ICTACT Journal on Soft Computing*, vol. 4, no. 03, pp. 743-752, APRIL 2014.