تصميم نموذج جدار ناري لحماية تطبيقات الوب اعتماداً على الخوارزمية الجينية

م. أمير أحمد*
د. محمد بسام الكردي**

الملخص

منذ انتشار الشبكات الحاسوبية والتكنولوجيا وظهور الوب، زاد الاعتماد على الإنترنت بشكل كبير لما يوفره من تسهيلات جمة للمستخدمين، فازداد معه عدد الخدمات الإلكترونية المقدمة وحجمها وتعقيدها لتغطي مساحة كبيرة في جميع القطاعات، ولكن ترافق مع هذه الزيادة زيادة في الهجمات القائمة على تطبيقات الوب والذي يعتبر في وقتنا الحالي غير آمن أبداً.

وتبعا للتطور الكبير الذي يشهده هذا المجال فلقد زاد معه تطور الهجمات التي تهدف لاختراق حسابات المستخدمين وسرقة البيانات الحساسة التي يمكن استخدامها للإضرار بمستخدمي هذه التطبيقات، لذلك كان لابد من دراسة الآليات الهجومية والدفاعية وتوظيفها في تطوير نظام قادر على حماية هذه التطبيقات من أنواع مختلفة من الهجمات.

نقدم في هذا البحث تطبيق جدار ناري معتمد على نوع من أنواع الخوارزميات التطورية (الخوارزمية الجينية) للحماية من ثلاث أنواع هجمات حقن النصوص البرمجية وهي (Cross-site scripting(xss), sql injection, blind sql injection) حيث نقوم بتوليد طلبات هجومية اعتماداً على الخوارزمية الجينية ومن ثم نقوم بتزويد قاعدة البيانات الخاصة بالجدار الناري بهذه الطلبات، حيث أظهرت النتائج كفاءة النظام المقترح للتحقق من نوعية الطلبات الواردة لتطبيق ويب بدقة (99.2%) مما سيساهم في حماية التطبيقات من الاختراق بهذه الأنواع المذكورة سابقا.

الكلمات مفتاحية: أمن الشبكات – الذكاء الصنعي والأمن السيبراني – الخوارزمية التطورية – نماذج الهجمات الإلكترونية والتهديدات – الخوارزمية الجينية – جدار الحماية لتطبيق الوب.

* ماجستير علوم وب، الجامعة الافتراضية السورية.
** دكتوراه في الذكاء الصنعي، أستاذ في الجامعة الافتراضية السورية.

Designing a firewall model to protect web applications based on the genetic algorithm

Eng: Ameer Ahmad[*]
Dr: Mohammad Bassam Kurdy[**]

Abstract

Since the spread of computer networks and technology and the emergence of the Web, the reliance on the Internet has increased significantly because of the great facilities it provides to users. With it, the number, size, and complexity of the electronic services provided increased to cover a large area in all sectors, but this increase was accompanied by an increase in attacks based on Web applications, which are considered Nowadays it is not safe at all.

According to the great development witnessed in this field, the development of attacks aimed at penetrating user accounts and stealing sensitive data that can be used to harm the users of these applications has increased, so it was necessary to study the offensive and defensive mechanisms and employ them in developing a system capable of protecting these applications from different types of attacks.

In this research, we present a firewall application based on a type of evolutionary algorithm (genetic algorithm) to protect against three types of scripting attacks (cross-site scripting (XSS), SQL injection, blind SQL injection), where we generate offensive requests based on Genetic algorithm and then we provide the database of the firewall with these requests, where the results showed the efficiency of the proposed system to check the quality of requests received for a web application with accuracy (99.2%), which will contribute to the protection of applications from penetration of these types mentioned previously.

Keywords: Network security, Artificial Intelligence and Cybersecurity, Evolutionary algorithm, Cyber attacks and Threat models, genetic algorithm, Web Application Firewall.

[*]Master in MWS, Syrian Virtual University.
[**]Ph.D. in Artificial Intelligence, Professor in Syrian Virtual University.

المقدمة

في ظل هذا التطور التكنولوجي الهائل والتقدم الكبير في مجال أنظمة المعلومات الرقمية وانتشار ثقافة الانترنت والتوسع في شبكة الوب العالمية وزيادة الاعتماد على المواقع الالكترونية لتحقيق خدمات وفوائد تعود بالنفع على الأفراد والمنظمات المستفيدة من هذه الخدمات، كان لابد من حماية هؤلاء المستخدمين وضمان أمن وخصوصية بياناتهم.

ومؤخراً كان هنالك زيادة في عدد وتنوع الهجمات الالكترونية التي تستهدف المواقع الإلكترونية لاختراق حسابات المستخدمين وسرقة البيانات الحساسة التي تضر بمزود الخدمة والمستخدم على حد سواء، ومما يجعل من الصعوبة بمكان الكشف عن هذه الهجمات وحماية بيانات المستخدمين من قبل المحللين الأمنيين والعاملين في مجال الأمن السيبراني [35,31,34].

للتعامل مع هذه المشكلة توصل البحث العلمي لفكرة "Threat Intelligence" والتي تعني ذكاءات التهديد وتشير إلى مجموعة من البيانات التي تم جمعها، وتقييمها وتطبيقها فيما يتعلق بالتهديدات الأمنية، والجهات الفاعلة في التهديد والاستغلال والبرمجيات الخبيثة ونقاط الضعف ومؤشرات الاختراق[1–20,4-17].

وتزامن ذلك مع ظهور ذكاءات التهديد السيبراني Cyber Threat Intelligence (CTI) لمساعدة باحثي الأمن في التعرف على مؤشرات الهجمات الإلكترونية، واستخراج المعلومات حول طرق الهجوم، وبالتالي الاستجابة للهجوم بدقة وفي الوقت المناسب، فعندما يتم جمع قدر كبير من البيانات من حلول المراقبة الأمنية أو إنشاؤها بواسطة حلول مراقبة أمنية مختلفة، فإن تقنيات تحليل البيانات الضخمة الذكية ضرورية لتجريد وتفسير واستخراج المعرفة من البيانات التي تم جمعها [2–3].

ويستخدم المهاجمون عدة طرق لمهاجمة المستخدم الضحية، هدفها الوصول الى البيانات الحساسة التي تخصه مثل المعلومات المالية، أو الوصول الى جهازه والتحكم فيه لتطبيق المزيد من الهجمات على الموقع.

ومن هذه الطرق نشر البرمجيات الخبيثة والفايروسات، أو قفل وتشفير بيانات المستخدم الضحية كما في فايروس الفدية [5].

وهنالك عدة أنواع شائعة للهجمات السيبرانية أيضا منها البرامج الضارة Malware، تصيد المعلومات Phishing، الرجل في الوسط Man in the Middle Attacks، حجب الخدمة Denial of Service، حجب الخدمة الموزع Distributed Denial of Service، حقن Sql Injection، الهجوم دون انتظار Zero day، الاتصال النفقي عبر DNS (DNS Tunneling) [5,18–6-7].

وعلى الرغم من اختلاف الطرق المتبعة في مختلف الهجمات إلا أنها تتشارك بالجوهر الذي يعتمد على دورة حياة مشتركة إلى حد ما، بدءاً من استطلاع الضحية إلى تنفيذ الأنشطة الضارة على الجهاز أو الشبكة التي يستخدمها الضحية[8,33].

ولقد أظهرت الدراسات الحديثة تطور هذه الهجمات بشكل مستمر وزيادة فعاليتها ضد المواقع الإلكترونية المختلفة وعدم كفاية جدران الحماية لردع هذه الهجمات والحد من آثارها الضارة[9,19,22]، لذلك

كان لابد من البحث عن آلية فعالة للتنبؤ بهذه الهجمات، ودراسة الآليات والتقانات الهجومية والدفاعية في الوب بهدف توظيفها في تطوير نظام قادر على الحماية وكشف الهجمات الشبكية في الوب، وتحديداً المنطوية تحت إطار الذكاء الصنعي وبناء جدار ناري خاص بتطبيقات وب اعتماداً على الخوارزمية الجينية، واستكشاف تأثير زيادة اتساع الهجمات على قابلية التوسع في الخوارزمية، بالإضافة الى استخدام دفاع ديناميكي قائم على الخوارزمية الجينية، وتأثير التطور المشترك للجزء الديناميكي (مهاجم – مدافع) في ذلك.

الدراسة المرجعية

ان الانتقال إلى السحابة يزيد المرونة للعديد من المؤسسات حول العالم، ولكنه أيضاً يزيد من المخاطر التشغيلية والتعقيدات الخاصة بالحفاظ على أمن البيانات وسير العمل[17-10,16-11].

وبالاتجاه نحو البيانات الرقمية، أصبح من المجدي أكثر للمهاجمين أن تحاول الانخراط في سرقة البيانات أو تعطيل الخدمات عبر الإنترنت لتحقيق مكاسب مادية أو مالية، أو من أجل التجسس أو لتعطيل خدمة ما.

[7] يستكشف هذا المرجع طريقة جديدة للتعامل مع إنشاء التهديدات واكتشافها في الشبكة، وتهدف إلى إثبات أنه من خلال تعريف واستخدام الجينات الصحيحة، يمكن أن تؤدي الخوارزمية الجينية أداءً أفضل من النظام القائم على القواعد الثابتة والمستخدم لإنشاء مجموعة واسعة من الهجمات والدفاع ضدها.

يركز الحل المقترح في هذا المرجع على ثلاثة أنواع من الهجمات: استخراج البيانات Data Exfiltration، واختطاف الخادم Server Hijack، وحجب الخدمة Denial of Service.

أثبت هذا المرجع أنه باستخدام خوارزمية بسيطة للتطور النحوي يمكن التغلب على جدار حماية بسيط لتطبيق الوب، وقدم استدلالاً لإنشاء استراتيجية دفاع ديناميكية لمعالجة العديد من المخاطر والتحديات كما أثبت أنه بإمكان الخوارزمية الجينية التعامل مع زيادة اتساع الهجمات بمزيد من التميز كما أنها تعطي نتائج أفضل، ولقد واجه المرجع صعوبات في تطبيقه على نطاق واسع، فالتطبيق والطرق التي تم تطويرها تحاكي وتنمذج الطلبات والاستراتيجيات الحقيقية للهجمات العدائية في الأمن السيبراني فقط، كما واجه صعوبات تتجلى في الوقت اللازم لتشغيل كل تجربه مما يصعب عملية التكرار من الناحية الحسابية.

[8] يركز هذا المرجع على الأمن السيبراني فهو مجال سريع التطور ويكاد يحتل الصدارة في أبحاث المهاجمين والمدافعين على حد سواء خلال العقد المنصرم، فعدد التهديدات يرتفع بشكل كبير ويسعى المهاجمين دوماً لإيجاد الثغرات والبقاء في مواكبة التقنيات المتطورة بشكل مستمر، كما يستكشف إمكانات الذكاء الصنعي في تحسين دفاعات الأمن السيبراني من خلال تحديد نقاط القوة والضعف فيه.

أثبت هذا المرجع أن الذكاء الصنعي تقنية لا غنى عنها في الأمن السيبراني وذلك مع تزايد سرعة وتعقيد الهجمات، كما قدم مراجعه شاملة للتهديدات

الإلكترونية والحلول، كما ناقش الطرق التي يمكن بها إطلاق الهجمات الإلكترونية واكتشافها ومكافحتها، كما واجه المرجع تحديات أهمها السباق القائم بين المهاجمين لاكتشاف الثغرات واستغلالها من جهة والمدافعين لاكتشاف الثغرات وترميمها بالحلول المناسبة أيضاً من جهة أخرى، كما شكلت البنية التحتية الداعمة للحوسبة السحابية عائقاً من حيث التكلفة الحسابية التي تسببها خوارزميات الذكاء الصنعي ومشاكل السرعة الناجمة عن التعقيد العالي للخوارزميات والتطورات المستمرة التي تشهدها البنية التحتية بشكل عام .

[30] يقترح هذا المرجع خوارزمية جديدة لنظام كشف الاختراق network intrusion detection system (NIDS) باستخدام مجموعة مميزات محسنة تم اختيارها مباشرة بواسطة الخوارزمية الجينية المعتمدة على البحث الشامل والتجميع الضبابي للوسائل fuzzy C-means clustering (FCM). حيث قام الباحث بوضع مجموعة المميزات العميقة المستخرجة بواسطة نموذج CNN المحدد في مصنف التعبئة BG للتحقق فيما بعد من صحة الأداء باستخدام CV ذات ال 5 مراحل.

أثبتت النتائج الأداء الموثوق للغاية الذي تم تحقيقه من خلال إجراء التحقق الصحيح الخماسي المراحل للخوارزمية المقترحة الذي تضمن تطبيقا جيدا في بيئة شبكة الكمبيوتر العملية NIDS. العيب الأساسي لهذا النظام هو الزمن الكبير اللازم لكشف الاختراق.

[13] تحاكي الخوارزمية الجينية التطور البيولوجي الذي تتكيف فيه الكائنات الحية مع بيئتها من خلال تكرار الطفرات والتصالب والاختيار لأجيال عديدة، ويتم استخدامها بشكل شائع لإيجاد الحلول المثلى وتطبيقها على العديد من المشكلات.

قام الباحث في هذا المرجع بتوسيع الخوارزمية الجينية لتشمل مجال تقييم الأمان من خلال اختيار العديد من الجينات الرئيسية والتي تستخدم بشكل أساسي في هجمات Cross-site scripting (XSS)، كما قام الباحث بالتحقق من الإنشاء التلقائي لطلبات الحقن والتي تشكل نقاط ضعف في تطبيقات الويب، كما قدم الإجرائية المتبعة ونتيجة التحقق.

أثبتت النتائج لهذا المرجع فعالية الخوارزمية المقترحة في إنشاء طلبات حقن جديدة كليا لم تكن معروفة مسبقا، وبإمكانها تجاوز جدران الحماية.

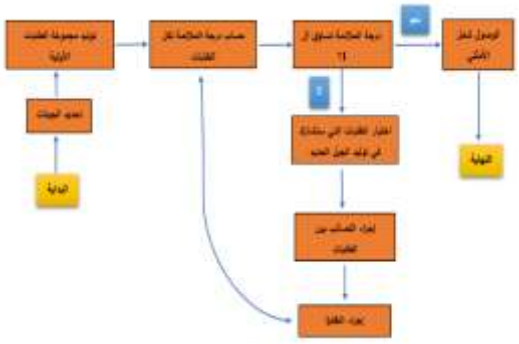
ولقد عانت الأبحاث السابقة في هذا المجال من قيود و مشاكل كثيرة تتجلى في عدم قدرتها على وضع نظام أمني لحماية تطبيقات الوب بحيث يتناسب مع الزيادة والتنوع والتطور في الهجمات فضلاً عن عدم قدرة الأنظمة على تحديث قواعدها بشكل يتناسب طرداً مع هذه الزيادة لحماية تطبيقات الوب وكشف الهجمات والحد من خطورتها وعدم وجود قاعدة بيانات تشمل الأنواع المتعددة من الطلبات الهجومية، مما شكل دافعا في الكشف عن تأثير استخدام الخوارزمية الجينية ضمن نظام أمني قابل للتوسع من حيث تحديث القواعد بشكل يتناسب مع تطور الهجمات على أمان تطبيقات وب ونجاعة استبدال الأنظمة القائمة على القواعد الثابتة بهذا النظام حيث تم تغطية أكثر المميزات تكراراً في الطلبات الهجومية واعتمادها كجينات ضمن النظام المقترح، كما قمنا بدراسة

الآليات والتقانات الهجومية والدفاعية في الوب بهدف توظيفها في نظام قادر على الحماية وكشف الهجمات الشبكية في الوب، وتحديداً المنطوية تحت إطار الذكاء الصنعي ودراسة إمكانية بناء جدار ناري خاص بتطبيقات وب اعتماداً على الخوارزمية الجينية وقياس أدائه مقارنة بالأنظمة الدفاعية الأخرى، وكذلك دراسة إمكانية استخدام دفاع ديناميكي قائم على الخوارزمية الجينية، وتأثير التطور المشترك للجزء الديناميكي (مهاجم – مدافع) في حماية تطبيق وب .

مواد البحث وطرائقه

إن ترميز المسألة بشكل صحيح يؤثر بشكل مباشر على عمل الخوارزمية الجينية، فالخوارزمية الجينية تتبع الطرق المستوحاة من الطبيعة في التطور فهي تقوم بتعريف مجموعة حلول أولية عشوائية وتعمل على تطويرها وهكذا حتى نصل للحل الأمثل المناسب[12-15].

وفيما يلي مخطط النموذج المعتمد على الخوارزمية الجينية:



الشكل(1) النموذج المقترح الذي يعتمد على الخوارزمية الجينية

العناصر الأساسية للخوارزمية الجينية وما يعبر عنها في مسألة توليد SQL-Injection-Attack :

تعتمد الخوارزمية الجينية على مبدأ التطور من جيل لآخر حيث أن عناصرها الأساسية هي الجيل (generation) وهو فعليا مكون من مجموعة حلول (population) وكل حل يعبر عن فرد له صبغيات (chromosome) و يتكون الفرد الواحد من مجموعة من الجينات (genes) تعبر عن المميزات التي تظهر الاختلاف بين فرد و آخر[21-23].

عملية إنشاء الحل الأولي Initialization:

وسنقوم الآن بإسقاط هذه المصطلحات التطورية على مسألة SQL-Injection-Attack-Request على النحو التالي:

- مجموعة الحلول Population: مجموعة الطلبات المرسلة إلى تطبيق وب.
- الصبغيات Chromosome: الطلب المرسل إلى تطبيق وب.
- الجينات Gene: تعبر عن المميزة الواحدة ضمن الطلب و عدد مرات ظهور هذه المميزة ضمن الطلب الواحد[25].

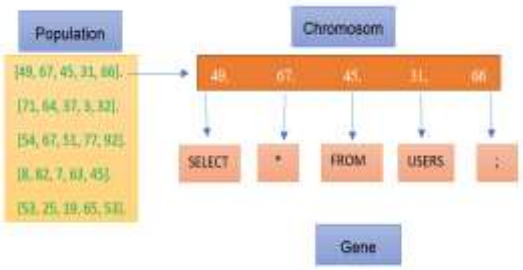
وفيما يلي نوضح طريقة إسقاط المصطلحات التطورية على مسألة SQL-Injection-Attack كما في الشكل التالي:

كل سطر من الاسطر التي هي ضمن قاعدة البيانات يعبر عن صبغي Chromosome (طلب).

وكل خلية ضمن السطر تعبر عن جين Gene (مميزة).

ومجموعة الأسطر مجتمعه تشكل مجموعة الأفراد Population (الطلبات جميعها المخزنة ضمن قاعدة البيانات) [24,27].

في الحل المقترح يتألف الطلب المراد انشاؤه Chromosome من خمس جينات كالتالي وكل جين يحمل رقم يعبر عنه وال Population تتألف من عدد من الطلبات كما هو موضح في الشكل التالي:



الشكل (2) آلية توليد الطلبات في النموذج المقترح

طبعا إذا قمنا باعتماد العشوائية المطلقة أثناء عملية إنشاء قاعدة البيانات فسينتج لدينا حل أولي يحوي على قصور شديد في إمكانية تحديد اذا كان هذا الطلب جيدا أم لا.

لذلك قمنا بالتوسع في المميزات الهامة لكل طلب لإمكانية اكتشاف الطلبات المطورة والتي من الممكن أن يقوم المهاجم باستخدامها ضد أي تطبيق وب وفي أي وقت يريده.

كما من اللافت الانتباه إلى أنه عند توسعنا في المميزات الهامة لكل طلب كان لابد أن ينتج لدينا تابع ملائمة ذو قيمة كبيرة جدا، ولكن ذلك لن يؤثر على فعالية الخوارزمية وأدائها بل سيحسن بشكل كبير إمكانية اكتشاف أي طلب جديد وسيعمل على تغطية أكبر قدر ممكن من الطلبات الخبيثة.

تابع الملائمة:

الطلب المقدم لتطبيق وب يتم إنشاءه وفق قيود معينه يجب أن يطبقها ويتناسب معها، تابع الملاءمة Fitness Function هو التابع الذي يقوم بتقييم الحل وإعطاءه درجة ملائمة معينه بحسب تحقيقه لمجموعة القيود المطلوبة.

وتنقسم هذه القيود إلى نوعين مرنة وصلبة:

- القيود الصلبة Hard Constraints: هي التي لا يمكن تعديلها أبداً ولا بد من التقيد بها وتطبيقها عند بناء الطلب، وعند الإخلال بها يصبح الطلب غير صالح وغير قابل للتنفيذ.

كأن يوجد طلب بدون Method أو بدون Path أو بدون Body .

- القيود المرنة Soft Constraints: هي القيود التي يمكن تعديلها (أي لا يشترط وجودها ضمن الطلب) ويمكن للطلب المخل بأحد هذه القيود أن يكون قابلاً للتنفيذ لكنه يكون قد ابتعد عن الأمثلية.

كأن يوجد طلب يحوي على Badwords أو Nullbyte أو Shapoo أو الخ..

يأخذ تابع الملاءمة بعين الاعتبار كلا من القيود الصلبة والمرنة، حيث يقوم باختبار جميع الشروط على الطلبات وعند الإخلال بأحد الشروط يقوم بإضافة ثابت التثقيل الخاص بهذه الشرط.

وبافتراض القيود الصلبة هي:

$HC= (hc_1, hc_2, hc_3)$

والقيود المرنة هي:

$SC= (cs_1, cs_2, \dots . cs_z)$

يقوم تابع الملائمة باحتساب درجة الملائمة للطلب الممثل كشعاع Vector على النحو التالي:

$$(1)\ V(r) = \sum_{i=1}^{3} hc_i v_i + \sum_{i=3+1}^{z} cs_i w_i$$

$$Fitness\ value(\mathrm{r}) = 1 \div (1 + \sum_{n=1}^{m}(Vr_m)) \quad (2)$$

تعبر المعادلة (1) عن قيمة تحقيق الطلب (r)للقيود، وكلما اقتربت القيمة للصفر كلما كان الطلب غير محقق للقيود بنسبة أكبر، وتكون القيمة مساوية للصفر عندما يكون الطلب غير محقق لكافة القيود.

hc_i: تعبر عن القيد الصلب (i) وتكون قيمته (1) إذا حقق الطلب القيد وقيمته (0) إذا كان الطلب غير محقق للقيد.

cs_i: تعبر عن القيد المرن (i) وتكون قيمته (1) إذا حقق الطلب القيد وقيمته (0 فقط) إذا كان الطلب غير محقق للقيد.

v_i: تعبر عن ثابت تثقيل القيد الصلب (i) والقصد هنا درجة أهمية القيد أي تزيد قيمة الثابت بزيادة أهمية القيد المخل فيه.

w_i: تعبر عن ثابت تثقيل القيد المرن (i).

وتعبر المعادلة (2) عن قيمة ملائمة مجموعة المميزات الهامة في الطلب، والتي يتم فيها احتساب مجموع قيم ملاءمة كل المميزات.

حيث تتراوح قيمة تابع الملائمة بين الصفر والواحد، أي أنه كلما كانت قيمة الملائمة أكبر كلما كان الحل أنسب حتى تصل قيمة الكفاءة للواحد أي تم الوصول للحل الأمثل الذي يحقق كل الشروط الموجودة.

عملية الاختيار:

يقصد بعملية الاختيار أو الانتقاء في الخوارزمية الجينية بأن نقوم باختيار مجموعة حلول جزئية من الجيل N لتدخل بتوليد الجيل الجديد N+1 [26].

يعتمد الاختيار بشكل أساسي على درجة الملائمة المعطاة لكل فرد من أفراد الجيل وبحسب درجة الملاءمة وطريقة الاختيار يتكون لدينا مجموعة صغيرة من الأفراد الأكفأ بحسب المفهوم التطوري وبإكمال عمليات الخوارزمية الجينية على هذه المجموعة المنتقاة يتولد لدينا جيل جديد من الأفراد.

لدينا مجموعه من الحلول ولكل حل درجة ملائمة قد حصل علها من تطبيق تابع الملائمة عليه.

سنقوم باعتماد طريقة الانتقاء العشوائي القائم على التقييم، لأن هذه الطريقة تعتبر من أكفأ الطرائق في الاختيار فهي تحقق التوازن في عملية البحث في فضاء الحلول بين عاملي الاستكشاف والتعمق.

سنقوم باختيار مجموعه من الحلول بشكل عشوائي وانتقاء الأفضل منها وبهذه الطريقة نكون قد حققنا عامل الاستكشاف من خلال العشوائية في اختيار المجموعة، وحققنا عامل التعمق من خلال اختيار الأفضل من هذه المجموعة.

عملية اختيار هذين الطلبات بطريقة الانتقاء العشوائي القائم على التقييم على النحو التالي:

– نقوم باختيار عدد معين n من مجموعة الطلبات بشكل عشوائي.

– نقوم باختيار الحل الأكثر ملائمة من الطلبات n استنادا لتابع الملائمة.

حيث أن العدد n هو عدد الطلبات المنتقاة في طريقة الانتقاء العشوائي القائم على التقييم، و يجب ان يتوازن اختيار العدد بحسب حجم مجموعة الحلول Population .

– إذا كان العدد قليل جداً فان ذلك يؤثر سلباً على نتيجة التطور بحيث يصبح الاقتراب من الحل أبطأ ولكن العدد القليل بنفس الوقت مفيد لتعزيز عامل الاستكشاف Exploration في عملية الوصول للحل الأمثل.

–وإن كان العدد كبير فإننا سنختار دائما الأفضل من مجموعة الأفراد مما يدعم عامل التعمق للوصول للحل ال Exploration [36,32].

وفيما يلي توضيح الآلية المتبعة لانتقاء الجداول التي تدخل في توليد الجيل القادم:

1– ليكن i متحول يساوي الصفر.

2– نقوم باختيار الطلب R1 بطريقة الانتقاء العشوائي القائم على التقييم.

3– نقوم باختيار الطلب R2 بطريقة الانتقاء العشوائي القائم على التقييم.

4– نقوم بتوليد طلب يدخل في أفراد الجيل اللاحق بإجراء عملية التداخل بين الطلبين R1,R2 .

5– I=i+1 .

6– نقوم بتكرار الخطوات من 2 ل5 حتى تصبح قيمة i تساوي حجم مجموعة الأفراد المطلوبة.

عملية التصالب Crossover:

عملية التصالب هنا مبنية على عملية تداخل فردين من بين الافراد الذين تم اختيارهم في مرحلة الانتقاء وهذين الفردين (الحلين) نريد أن نستخرج منهما فرد جديد أي طلب جديد [28,29].

وبناء على الطلبين اللذان تم اختيارهما مسبقا، نقوم بقراءة جيناتهما، ثم نحدث تداخل بين هذه الجينات، علماً أن كل طلب يحوي على عدد من المميزات ونعبر عن عدد المميزات ب N ميزة في الجيل الواحد.

وعند احداث التداخل العشوائي بين هذين الطلبين تنتج عن هذه العملية بيانات نضعها في طلب جديد وهو الجيل الجديد المنتج من الطلبين السابقين بعملية التداخل العشوائي.

ولكي نوضح فكرة التصالب بين طلبين سنعتمد على المثال التالي:

أقاد الـ الـ الي:

SELECT \ getRequestString("username");

UserName FROM Users WHERE UserId = 105 OR 1=1;

أقاد الـ الـ ي :

UserName : getRequestString("username");

SELECT \ FROM Users WHERE UserId = 105 OR 1=1;

الشكل(3) آلية التصالب في النموذج المقترح

كما لاحظنا في المثال السابق تم اقتطاع الطلبين في موقع الجين الثالث وتم اضافة الجزء الثاني من الطلب الثاني الى الجزء الأول من الطلب الأول ودمجهما في طلب جديد، كما تمت إضافة الجزء الثاني من الطلب الأول الى الجزء الأول من الطلب الثاني ودمجهما في طلب جديد.

الطفرة Mutation:

لدينا طلب من الطلبات نريد أن نحدث عليه طفره من أجل إحداث الاستكشاف في فضاء الحلول الذي نبحث فيه عن الطلب الأمثل، نقوم بإحداث الطفرة

على إحدى خواص الجين أي في حالتنا هي إحدى الجينات ال 5 للطلب، ونستبدل قيمتها بإحدى القيم التي أنشأت بشكل عشوائي، وبذلك نكون قد أحدثنا طفرة في هذا الطلب بتغيير قيمة إحدى مميزاته وهذا ما سيزيد من عامل الاستكشاف Exploration، كما في الشكل(4)، حيث تم اجراء طفره على الجين الثاني في الطلب الأول فتم تغيير قيمته من (:) الى (=)، كما تم اجراء طفره على الجين الثاني في الطلب الثاني وتم تغيير قيمته من (\) الى (*) مثلما هو موضح في الشكل التالي.

الشكل (4) آلية الطفرة في النموذج المقترح

تصميم الجدار الناري Web Application Firewall:

نظرا للوقت الكبير نسبياً واللازم لتطبيق تابع الملائمة على طلب الوب الوارد لتطبيق وب معين في الزمن الحقيقي ونظراً لكثرة الطلبات الواردة، تم تصميم الجدار الناري اعتماداً على بيئة python3.10.2، يحوي على Search Box وزر البحث الذي نقوم من خلاله بإدخال طلب لا على التعيين للبحث عنه في الداتا بيز المؤلفة من ثلاث ملفات اكسل (ملف لكل نوع من أنواع الطلبات الهجومية المدروسة والمنشأة بواسطة الخوارزمية الجينية)، فاذا رأى الطلب نفسه أو شيء مشابه له بنسبة 75 بالمئة، فيعيد نوعه الموضح في اسم الملف الذي يحتويه وانه بنسبة 90% عباره عن invalid request كما يلي:

90% Request attack- tybe – invalid request

نقصد ب tybe نوع الطلب الموضح في اسم الملف الذي يحتويه، وفيما عدا ذلك فانه يعيد:

90% valid request

النتائج والمناقشة

بتنفيذ النظام المقترح المعتمد على الخوارزمية الجينية ذات ال(10000) جيل والمعتمدة على حوالي (200) جين وعدد الطلبات في كل جيل حوالي 100 طلب وكل طلب يتألف من 5 جينات، حيث تم إنشاء الجينات المستخدمة في هذا التحقق من الأكواد المستخدمة في(burp suite برنامج مفتوح المصدر) ومن دراستنا لأكثر الطلبات الهجومية شيوعاً، ومن ثم تحليل هذه الأكواد واستخراج الجينات الأكثر أهمية (أي الأكثر تكرارا في الطلبات الهجومية)، كما تم التحقق من التعبير النحوي للطلب بواسطة مكتبة tidy 5.4.0 حيث تقوم بتحديد عدد الأخطاء والتحذيرات، ولقد قمنا باستخدام مكتبة selenium 3.4.3 مع أداة Chrome web driver على التوازي لقياس قابلية تنفيذ الطلب المولد ضمن المتصفح، ومن ثم قمنا باستخدام مكتبة CRS 2.2.9 لاستبعاد الطلبات التي تحوي كم كبير من الأخطاء النحوية واختيار الطلبات التي تحصل على درجة ملائمة عالية، أما بالنسبة لعملية التصالب بين طلبين مختارين فلقد قمنا بتحديد نقطتين على جينات كل طلب ثم قمنا بتبادل الجينات بين النقطتين المحددتين

لإنشاء طلبين جديدين كما هو موضح سابقاً مع مراعاة أن يكون الوقت اللازم لإحداث الطفرة أقصر ما يمكن وذلك لضمان سرعة النموذج، ومن ثم قمنا باعتماد معدل طفرة 25 % بالنسبة لكل جيل، أما شرط التوقف فكان تبعاً لعدد الأجيال بحيث لا يتجاوز 10000 جيل.

وتوضح الجداول التالية بعض الأمثلة للطلبات الهجومية التي حصلنا عليها بواسطة الخوارزمية الجينية مع تحديد نوعها وهل هي جديدة أم معروفة مسبقاً:

الجدول (1) sql injection attack

Request No	Sql Injection Request	New
٢٣٩٥٢	Drop table username; --'	No
٢٤١٩٤	admin' and 1=0 union all select 'admin'	No
٢٤٤٣٩	Select case when substring ='#';	No
٢٤٩٧٧	Select if (x=y,sleep(30),'a');	yes
٢٥٥٩٦	Select * into @'admin';	yes
٢٥٧٠١	As inject where x=0#;	yes
٢٦١٧٢	Union all select @user();	yes
٢٦٣٥١	And Sleep (10)=';	yes
٢٧٠٩٤	waitfor delay '00:00:30' or sleep(5);	yes

الجدول (2) xss Request attack

Request No	XSS Request	New
٢٣٧٤	\u0061lert``;</script></tr>java<tbody/>	No
٥٠٨٦	data=type="image" <video><source >button onerror=alert();	No
٩٧٦٥	/u0061lert``;</script>height=
accept-charset=2	yes
١٤٧٢٦	<script src=%(jscript)>	yes
٢١٩٥١	<type='image' src=%(scriptlet)>	yes

الجدول (3) blind sql injection attack

Request No	blind sql injection Request	New
٢٧٦١٣	SELECT pg_sleep(5);	No
٢٨٢٩١	or SELECT pg_sleep(5);	yes
٢٨٧٤٥	b'/my%20documents/' and 1	yes

ولقد تم قياس دقة النظام المقترح بتجريب 100 طلب لاكتشاف الطلبات الهجومية وتحديد نوعها، منها 25 طلب عادي و25 من النوع xss Request attack و25 من النوع sql injection attack و25 من النوع blind sql injection attack وكانت النتائج كما هو موضح في الجدول التالي:

الجدول (4) قياس دقة النظام المقترح

الـ ع	HTTP Request	xss Request attack	sql injection attack	blind sql injection attack
الدقة	%٩٩,١	%٩٨,٨	%٩٩,٧	%٩٩,٢

وبالاعتماد على النتائج الواردة في الجدول السابق تم قياس دقة النظام ككل بحساب المتوسط الحسابي للنتائج أعلاه كما يلي:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{99.1 + 98.8 + 99.7 + 99.2}{4} = 99.2$$

فكانت دقة النظام المقترح في دراستنا هي (99.2).

كما تمت مقارنة النظام المقترح مع ثلاث أنظمة أخرى في نفس المجال البحثي وهم الأنظمة المقترحة في المراجع (14 و13 و12) من حيث المنهجية المقترحة وأنماط الهجمات المدروسة واستخدام قواعد التحقق الافتراضية وعدد الهجمات المولدة وزمن التنفيذ وزمن الاستجابة والدقة كما يلي:

الجدول (5) مقارنة النظام المقترح مع المراجع الأخرى

عند مقارنة نظامنا المقترح بالنظام المقترح في المرجع [12] نلاحظ تفوق نظامنا من حيث أنماط الهجمات المدروسة وعدد الهجمات المولدة وزمن التنفيذ وزمن الاستجابة، وعند مقارنته أيضاً بالنظام المقترح في المرجع [13] نلاحظ تفوق نظامنا من حيث أنماط الهجمات المدروسة وعدد الهجمات المولدة وزمن التنفيذ، كما لاحظنا تفوقه على النظام المقترح في المرجع [14] من حيث الدقة وتباينه أيضاً معه من حيث أنماط الهجمات المدروسة.

الخلاصة

تم في هذا البحث بناء تطبيق جدار ناري لاكتشاف طلبات الويب (xss, sql injection, blind sql injection) الخبيثة وتحديد نوعها اعتماداً على الخوارزمية الجينية، واستكشاف تأثير زيادة اتساع الهجمات على قابلية التوسع في الخوارزمية (وذلك عند زيادة عدد الجينات فكلما كان اختيارنا لها أنسب كلما استطعنا تقليل عدد الأجيال اللازم لاكتشاف طلبات جديدة)، بالإضافة الى استخدام دفاع ديناميكي قائم على الخوارزمية الجينية، وتأثير التطور المشترك للجزء الديناميكي (مهاجم – مدافع) في ذلك، ولقد أعطى التطبيق دقة (99.2%) بتحديد ماهية الطلبات الجديدة، وهكذا نجد أن النظام المقترح سيساهم في حماية تطبيقات الوب من الهجمات المذكورة سابقا، وسيزيد من ثقة المستخدمين بهذه التطبيقات، كما من الممكن في المستقبل القريب التوسع في إمكانية الكشف عن أنواع جديده من الهجمات، وتحديد جينات بعناية أكثر لتقليل عدد الأجيال اللازمة للاكتشاف.

[13] Takaesu, Isao.(2017). Automatic Generation of Injection Codes using Genetic Algorithm. MBSD.

[14] Applebaum, Simon.(2021). Signature-based and Machine-Learning-based Web Application Firewalls: A Short Survey. University of Liverpool. United Kingdom.

[15] Nguyen, H.T., Franke, K. and Petrovic, S., (2010) Towards a generic feature-selection measure for intrusion detection. In 2010 20th International Conference on Pattern Recognition (pp. 1529-1532). IEEE.

[16] Appelt, D., Nguyen, C.D., Panichella, A. and Briand, L.C., (2018) A machine-learning-driven evolutionary approach for testing Web Application Firewalls. IEEE Transactions on Reliability, 67(3), pp.733-757.

[17] Appelt, D., Nguyen, C.D. and Briand, L., (2015) Behind an application firewall, are we safe from SQL injection attacks? In 2015 IEEE 8th international conference on software testing, verification and validation (ICST) (pp. 1- 10). IEEE.

[18] Prabhudesai, P., Bhalerao, A.A. and Prabhudesai, R., (2019) Web Application Firewall: Artificial Intelligence Arc. International Research Journal of Engineering and Technology (IRJET).

[19] Torrano-Giménez, C., Perez-Villegas, A. and Alvarez, G., (2009) A self-learning anomaly-based web application firewall. In Computational Intelligence in Security for Information Systems (pp. 85-92). Springer, Berlin, Heidelberg.

[20] Nguyen, H.T., Torrano-Giménez, C., Alvarez, G., Petrović, S. and Franke, K., (2011) Application of the generic feature selection measure in detection of web attacks. In Computational Intelligence in Security for Information Systems (pp. 25-32). Springer, Berlin, Heidelberg.

[21] Folini, C. and Ristić, I., (2018) ModSecurity Handbook. The Complete Guide to the Popular Open-source Web Application Firewall. Second edition. Feisty Duck. London.

[22] Manaseer, Saher.(2018). Centralized Web Application Firewall Security System. Amman, Jordan.

[23] Thang, N.M., 2020. Improving Efficiency of Web Application Firewall to Detect Code Injection Attacks with Random Forest Method and Analysis Attributes HTTP Request. Programming and Computer Software, 46(5), pp.351-361.

Reference

[1] Singh, Simon. (2003). The Science of Secrecy from Ancient Egypt to Quantum Cryptography, Association for Computing Machinery. New York, United States.

[2] Kaplan, Andreas. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons. India.

[3] Michael Hogan, Elaine Newton. (2015). Supplemental Information for the Interagency Report on Strategic U.S. Government Engagement in International Standardization to Achieve U.S. Objectives for Cybersecurity. NIST.

[4] Ashlock, Daniel. (2006). Evolutionary Computation for Modeling and Optimization, Springer-Verlag. New York.

[5] Banka, Almaldeen. (2019). The Risks of Cyber Attacks and their Economic Impacts: The Case of the GCC Countries. Egypt.

[6] Ali Dehghantanha, Mauro Conti, Tooska Dargahi. (2018). Cyber Threat Intelligence, George Mason University. USA.

[7] Djeffal, S. (2020). Adaptive Defense Against Adverserial Artificial Intelligence at the Edge of the Cloud using Evolutionary Algorithms, Massachusetts Institute of Technology. USA.

[8] Zeadally, S., Adi, E., Baig, Z. and Khan, I. (2020). Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity, IEEE. USA.

[9] Tekerek, A. and Bay, O. (2019). Design And Implementation Of An Artificial Intelligence-Based Web Application Firewall Model, ResearchGate. Turkey.

[10] Abdul Jawad, N. and Kurdy, M. (2019). Stock Market Price Prediction System Using Neural Networks And Genetic Algorithm, Journal of Theoretical and Applied Information Technology. Syria.

[11] Garcia, D. Lugo, A. Hemberg, E. O'Reilly, U. (2017). Investigating Coevolutionary Archive Based Genetric Algorithms on Cyber Defense Networks, Cambridge University. United Kingdom.

[12] Appelt , Dennis.(2018). A Machine Learning-Driven Evolutionary Approach for Testing Web Application Firewalls, ResearchGate.

[24] Moosa, A., (2010) Artificial neural network-based Web Application Firewall for SQL injection. International Journal of Computer and Information Engineering, 4(4), pp.610-619.

[25] Thomas-Reynolds, Dainya and Butakov, Sergey, (2020) Factors Affecting the Performance of Web Application Firewall. WISP 2020 Proceedings.

[26] Vartouni, A.M., Teshnehlab, M. and Kashi, S.S., (2019) Leveraging deep neural networks for anomaly-based Web Application Firewall. IET Information Security, 13(4), pp.352-361.

[27] Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems, 189, 105124.

[28] Prabhudesai, P., Bhalerao, A.A. and Prabhudesai, R., (2019) Web Application Firewall: Artificial Intelligence Arc. International Research Journal of Engineering and Technology (IRJET).

[29] Raikar, D., et al. (October 2012). Preventing SQL Injection Attacks Using Combinatorial Approach. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(8), 46- 52.

[30] Kiseon, Kim. (2020). Genetic convolutional neural network for intrusion detection systems. Researchgate.

[31] The State of Web Application Security (2011), http://www.barracudanetworks.com/ns/downloads/White_Papers/Barracuda_Web_App_Firewall_WP_Cenzic_Exec_Summary.

[32] AQTRONIX, "AQTRONIX WebKnight - Open Source Web Application Firewall (WAF) for IIS,".(2016)[online].Available:https://www.aqtronix.com/?PageID=99.

[33] Trustwave, "ModSecurity: Open Source Web Application Firewall,"(2016). [online]. Available: https://modsecurity.org.

[34] S. Prandl, M. Lazarescu, D. Pham, '"A Study of Web Application Firewall Solutions,".(2015). In: Proceedings of 11th International Conference on Information Systems Security ,Kolkata, India.

[35] I. Ristic, '"ModSecurity Handbook,".(2010). United Kingdom: Feisty Duck.

[36] D. Pałka, M. Zachara, '"Learning Web Application Firewall - Benefits and Caveats," In: Proceedings of the IFIP WG 8.4/8.9 International Cross Domain Conference and Workshop on Availability.(2011). Reliability and Security for Business, Enterprise and Health Information Systems, Vienna, Austria.