UNIVERSIDAD SEÑOR DE SIPÁN

FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TESIS

DETECCIÓN AUTOMÁTICA DE ATAQUES DE INYECCIÓN MEDIANTE APRENDIZAJE AUTOMÁTICO EN BASES DE DATOS NOSQL

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

Autor (es):

Bach. Paico Chileno Daniel

ORCID: https://orcid.org/0000-0002-5734-3915

Bach. Valdera Contreras Jhon Harry

ORCID: https://orcid.org/0000-0003-4785-9253

Asesor:

Mg. Bravo Ruiz Jaime Arturo

ORCID: https://orcid.org/0000-0003-1929-3969

Línea de Investigación:

Infraestructura, Tecnología y Medio Ambiente

Pimentel – Perú 2021

RESUMEN

En los últimos años los gestores de datos NoSQL son la solución perfecta para la disponibilidad de la información en tiempo real, tecnologías como Big Data e Internet de las cosas están revolucionando la sociedad, para ello, empresas como Facebook, Google, Amazon entre otros, han tenido la necesidad de crear sus propias bases de datos NoSQL para satisfacer mejor a sus clientes, dependiendo el problema a resolver las bases de datos NoSQL se orientan en diferentes categorías tales como, documentos, clave – valor, columnas, grafos y memoria. La popularidad de SQL tiene como consecuencia la incomprensión de la existencia de ataques de inyección en almacenes de datos NoSQL, empresas como la fundación The Open Application Security Project (OWASP) sigue posicionando a los ataques de inyección entre los riesgos más peligrosos en aplicaciones web , es por ello que tanto la plataforma PARSE SERVER y la API de autenticación de Fortnite sufrieron ataques de inyección por parte de un dominio mal configurado y de uso de expresiones regulares (regex), por lo tanto, los aspectos de seguridad en NoSQL son críticos, pues lamentablemente este en sus inicios no era la preocupación principal de los diseñadores de sistemas, a consecuencia de ello, NoSQL tiene problemas de rendimiento en las soluciones defensivas existentes de los ataques de inyección. Por esta razón, se presenta un estudio para detectar ataques de inyección mediante aprendizaje automático en bases de datos NoSQL, para ello se determinaron los mejores algoritmos de clasificación en base al criterio de exactitud asi como en los casos donde se acontece ataques de inyección, asi mismo, se construyó el conjunto de datos el cual se orienta a un patrón que sigue la estructura de la Notación de Objetos de Java Script (JSON), el cual consta de 19 atributos, 2 etiquetas y 509 ejemplos que le pertenecen a inicio y cierre de documento, clave, valor, separador y operador. Los resultados obtenidos demuestran que los algoritmos de clasificación de Red Neuronal, Perceptrón Multicapa, Árbol de decisión, Random Forest, Support Vector Machine y K-vecinos más próximos cumplieron los objetivos de la investigación, permitieron una clasificación correcta de los datos con una precisión muy alta de 98.7%, 98.7%, 96.3%, 89.6%, 84.8% y 84.8% respectivamente, y en un periodo de tiempo muy corto. Los modelos de clasificación propuestos en este estudio pueden detectar ataques de inyección en bases de datos NoSQL orientado a documentos, como estudio futuro se plantea crear un Dataset basado en un solo patrón orientado a cualquier tipo de base de datos NoSQL.

PALABARAS CLAVE: Algoritmos de Clasificación, Aprendizaje automático, Ataques de Inyección, Data Set, JSON, NoSQL.

ABSTRAC

In recent years NoSQL data managers are the perfect solution for the availability of information in real time, technologies such as Big Data and Internet of things are revolutionizing society, for this, companies like Facebook, Google, Amazon among others, have had the need to create their own NoSQL databases to better satisfy their customers, depending on the problem to solve NoSQL databases are oriented in different categories such as, documents, key - value, columns, graphs and memory. The popularity of SQL has resulted in a lack of understanding of the existence of injection attacks on NoSQL data stores. Companies such as The Open Application Security Project (OWASP) Foundation continue to position injection attacks among the most dangerous risks in web applications, That's why both the PARSE SERVER platform and the Fortnite authentication API suffered injection attacks from a misconfigured domain and the use of regular expressions (regex). Therefore, the security aspects in NoSQL are critical, because unfortunately this was not the main concern of system designers in the beginning. For this reason, a study is presented to detect injection attacks by means of automatic learning in NoSQL databases, for which the best classification algorithms were determined based on the accuracy criterion as well as in the cases where injection attacks occur. Likewise, the data set was constructed which is oriented to a pattern that follows the structure of the Java Script Object Notation (JSON), which consists of 19 attributes, 2 tags and 509 examples that belong to the start and end of the document, key, value, separator and operator. The results obtained show that the classification algorithms of Neural Network, Multilayer Perceptron, Decision Tree, Random Forest, Support Vector Machine and K-neighbors fulfilled the objectives of the research, allowed a correct classification of the data with a very high accuracy of 98.7%, 98.7%, 96.3%, 89.6%, 84.8% and 84.8% respectively, and in a very short period of time. The classification models proposed in this study can detect document oriented NoSQL database injection attacks. As a future study it is proposed to create a Dataset based on a single pattern oriented to any type of NoSQL database.

KEYWORD: Classification Algorithms, Machine Learning, Injection Attacks, Data Set, JSON, NoSQL.