Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

Digital Forensics Analysis for SQL Injection Attacks

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Master in
Computer Science**

by

Ammar Adel Ahmed

Supervised by

Asst. Prof. Dr. Najla Badie Ibraheem

2023 A.D. **1445 A.H.**

ABSTRACT

Digital forensics known as computer or cyber forensics refers to the process of collecting, preserving, analyzing, and presenting digital evidence from electronic devices and computer systems for legal or investigative purposes. This specialized field of forensic science involves the application of various techniques and methodologies to identify, recover, and examine digital data to uncover potential evidence of crimes, security breaches, or other malicious activities..

The primary objective of this thesis is to conduct a comprehensive digital forensic analysis specifically targeting SQL injection attacks, which are a prevalent type of web attack. Thesis methodology based on Digital Forensic Research Workshop (DFRWS) as investigation model to accomplish this, the research leverages advanced tools and methodologies. During the data collection phase, two prominent digital forensic tools, namely BHE (Browser History Examiner) and Disk Drill, are utilized. These tools have earned recognition for their capability to extract relevant information from various digital sources, including web browsers and disk drives.

In the analysis stage, an array of AI algorithms is employed. These algorithms encompass support vector machines (SVM), Naive Bayes (NB), logistic regression (LR), random forests (RF), decision trees (DT), convolutional neural networks (CNN), and bidirectional encoder representations from transformers (BERT). A carefully curated dataset formatted in SQL v3, containing a total of 30,873 records, serves as the training data for these algorithms. The achieved accuracy rates of the algorithms are noteworthy, with SVM achieving 94%, NB achieving 79%, LR and RF achieving 93%, DT achieving 92%, CNN achieving 96%, and the BERT model showcasing exceptional accuracy of 99.9%.

The evidence and results are saved and protected using the hash function SHA256.The thesis culminates in the generation of digital forensic reports that provide concise and comprehensive summaries of key findings. These reports focus on critical information such as the attacker's URL, the day of the month, the most frequent attacks per week, and the exact time of day. To effectively present and interpret these reports, Tableau tools are employed.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

التحليل الجنائي الرقمي لهجمات حقن (SQL)

رسالة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة الماجستير في
علوم الحاسوب

من قبل

عمار عادل احمد حامد

بإشراف
ا.م.د. نجلاء بديع ابراهيم

الخلاصة

يشير التحليل الجنائي الرقمي إلى عملية جمع الأدلة الرقمية وحفظها وتحليلها وتقديمها من الأجهزة الإلكترونية وأنظمة الكمبيوتر لأغراض قانونية أو تحقيقية. يتضمن هذا المجال المتخصص في علوم التحليل الجنائي تطبيق تقنيات ومنهجيات مختلفة لتحديد البيانات الرقمية واستعادتها وفحصها للكشف عن الأدلة المحتملة للجرائم أو الخروقات الأمنية أو الأنشطة الضارة الأخرى.

الهدف الأساسي من هذه الرسالة هو إجراء تحليل جنائي رقمي شامل يستهدف على وجه التحديد هجمات حقنSQL ، وهي نوع شائع من هجمات الويب. تعتمد منهجية الرسالة على ورشة عمل أبحاث التحليل الجنائي الرقمي (DFRWS) كنموذج تحقيق لتحقيق ذلك، ويستفيد البحث من الأدوات والمنهجيات المتقدمة. خلال مرحلة جمع البيانات، يتم استخدام أداتين رقميتين بارزتين للتحليل الجنائي ، وهما (Browser History Examiner) BHE و Disk Drill. وقد اكتسبت هذه الأدوات التقدير لقدرتها على استخراج المعلومات ذات الصلة من مصادر رقمية مختلفة، بما في ذلك متصفحات الويب ومحركات الأقراص.

في مرحلة التحليل، يتم استخدام مجموعة من خوارزميات الذكاء الاصطناعي. تشمل هذه الخوارزميات أجهزة المتجهات الداعمة(SVM) ، و(Naive Bayes (NB، والانحدار اللوجستي(LR) ، والغابات العشوائية(RF) ، وأشجار القرار (DT) ، والشبكات العصبية التلافيفية(CNN) ، وتمثيلات التشفير ثنائية الاتجاه من المحولات . .(BERT) تعمل مجموعة البيانات المنسقة بعناية والمنسقة فيSQLv3 ، والتي تحتوي على إجمالي 30873 سجلًا، بمثابة بيانات تدريب لهذه الخوارزميات. معدلات الدقة التي تم تحقيقها للخوارزميات جديرة بالملاحظة، حيث حقق%94 SVM ، و NBحقق 79%، و LRو RF حققا 93%، و DTحققا 92%، و CNNحققا 96%، ونموذج BERT يعرض دقة استثنائية تبلغ %.99.9

يتم حفظ الأدلة والنتائج وحمايتها باستخدام وظيفة التجزئة .SHA256 ومن خلال استخدام وظيفة التجزئة، يصبح أي تعديل غير مصرح به أو تلاعب بالأدلة قابلاً للاكتشاف، وبالتالي تعزيز مصداقية ومقبولية الأدلة في الإجراءات القانونية. وتتوج الرسالة بإنشاء تقارير التحاليل الجنائية الرقمية التي تقدم ملخصات موجزة وشاملة للنتائج الرئيسية. تركز هذه التقارير على المعلومات المهمة مثل عنوان URL للمهاجم، واليوم من الشهر، والهجمات الأكثر تكرارًا في الأسبوع، والوقت المحدد من اليوم. ولعرض هذه التقارير وتفسيرها بشكل فعال، يتم استخدام أدوات Tableau.