A Python Program for Automatic Detection of SQL Injection Attacks in Web Applications

Mehmonaliyev Yahyobek Usmonjon o‘g‘li
Student Fergana State Technical University
Umarov Abdulmuxtor Maxammad o‘g‘li
Assistant Lecturer Fergana State Technical University

Abstract: SQL Injection (SQLi) remains a significant threat to modern web application security. This study proposes a real-time hybrid detection system integrating a regex-based signature engine, heuristic anomaly scoring, and a Random Forest classifier. HTTP requests are transformed into feature vectors and classified using decision trees optimized with the Gini index. Experimental results demonstrate 99% accuracy and a 0.8% false positive rate. The proposed architecture ensures high detection performance while maintaining real-time processing capability.

Keywords: SQL Injection, web security, hybrid detection, Random Forest, Gini index, real-time protection, anomaly analysis

Veb-ilovalarda SQL Injection hujumlarini avtomatik aniqlovchi Python dasturi

Mehmonaliyev Yahyobek Usmonjon o‘g‘li
Farg‘ona davlat texnika universiteti talabasi
Umarov Abdulmuxtor Maxammad o‘g‘li
Farg‘ona davlat texnika universiteti assistent o‘qituvchisi

Annotatsiya: SQL Injection (SQLi) hujumlari zamonaviy veb-ilovalar xavfsizligiga jiddiy tahdid bo‘lib qolmoqda. Ushbu tadqiqot real vaqt rejimida ishlovchi gibrid aniqlash tizimini taklif etadi. Tizim regex-asosli imzo mexanizmi,

evristik anomaliya baholash va Random Forest klassifikatorini birlashtiradi. HTTP so‘rovlar xususiyat vektoriga aylantirilib, Gini indeksi asosida qaror daraxtlari orqali tasniflanadi. Eksperimental natijalar 99% aniqlik va 0.8% false positive ko‘rsatkichini namoyish etdi. Taklif etilgan arxitektura tezkor ishlash va obfuskatsiyalangan payloadlarni aniqlashda yuqori samaradorlikni ta’minlaydi.

Kalit so‘zlar: SQL Injection, veb-xavfsizlik, gibrid tizim, Random Forest, Gini indeksi, real vaqt aniqlash, anomaliya tahlili

Аннотация: SQL Injection остаётся одной из ключевых угроз безопасности веб-приложений. В работе представлена гибридная система обнаружения в режиме реального времени, объединяющая сигнатурный анализ, эвристическую оценку и классификатор Random Forest. HTTP-запросы преобразуются в векторы признаков и классифицируются на основе индекса Джини. Экспериментальные результаты показали точность 99% и уровень ложных срабатываний 0.8%. Предложенная архитектура обеспечивает высокую эффективность и стабильную работу в реальном времени.

Ключевые слова: SQL Injection, веб-безопасность, гибридная система, Random Forest, индекс Джини, обнаружение в реальном времени

Kirish

Zamonaviy raqamli infratuzilmaning asosiy qismini veb-ilovalar tashkil etadi. Elektron tijorat, bank tizimlari, davlat xizmatlari va ta’lim platformalari keng miqyosda ma’lumotlar bazasi bilan integratsiyalashgan holda ishlaydi. Biroq, ushbu integratsiya xavfsizlik nuqtai nazaridan yangi tahdidlarni yuzaga keltiradi. SQL Injection (SQLi) hujumlari ma’lumotlar bazasiga ruxsatsiz kirish, maxfiy ma’lumotlarni o‘g‘irlash, o‘zgartirish yoki o‘chirish imkonini beruvchi eng xavfli hujum turlaridan biri hisoblanadi. SQLi hujumlari foydalanuvchi tomonidan

kiritilgan ma’lumotlarning yetarli darajada tekshirilmasligi natijasida yuzaga keladi. Agar veb-ilova foydalanuvchi kiritgan qiymatni to‘g‘ridan-to‘g‘ri SQL so‘roviga qo‘shsa, tajovuzkor maxsus tuzilgan payload orqali so‘rov mantiqini o‘zgartirishi mumkin. Masalan, autentifikatsiya jarayonida quyidagi zararli so‘rov orqali tizimni aldash mumkin:

SELECT * FROM users WHERE username = 'admin' OR '1'='1';

Bunday holatda mantiqiy shart doimo rost bo‘lgani sababli, autentifikatsiya mexanizmi chetlab o‘tiladi. Mazkur tadqiqotning asosiy maqsadi — regex-asosli imzo mexanizmi, evristik tahlil va Random Forest klassifikatorini birlashtirgan gibrid aniqlash tizimini ishlab chiqish va uning real vaqt rejimidagi samaradorligini baholashdir. Taklif etilgan yondashuv aniqlik, past false positive darajasi va yuqori ishlash tezligini bir vaqtning o‘zida ta’minlashga qaratilgan.

Adabiyotlar tahlili

SQL Injection hujumlarini aniqlash bo‘yicha tadqiqotlar asosan uch yo‘nalishda rivojlangan: imzo-asosli, anomaliya-asosli va mashina o‘rganish yondashuvlari. Imzo-asosli (signature-based) usullar regex va qoidalar to‘plamiga tayanadi. Ular SQL kalit so‘zlar (SELECT, UNION, DROP) yoki maxsus belgilar (‘--’, ‘;’) mavjudligini aniqlaydi. Afzalligi — tezkorlik va past hisoblash xarajati, kamchiligi esa obfuskatsiyalangan yoki yangi hujumlarni aniqlashda zaiflik hamda yuqori false positive darajasidir.

Mashinali o‘rganish modellaridan Random Forest, SVM va Logistic Regression keng qo‘llanilgan. Ayniqsa Random Forest yuqori aniqlik ko‘rsatadi va Gini indeksi asosida optimal bo‘linishlarni tanlaydi:

$$Gini(D) = 1 - \Sigma(p\_k^2)$$

So‘nggi tadqiqotlar gibrid arxitekturalarning samaradorligini tasdiqlaydi. Regex tezkor filtr sifatida ishlaydi, evristik va ML modeli esa yakuniy qarorni chiqaradi:

$$R(x) = w_1 \cdot S\_regex + w_2 \cdot S\_heuristic + w_3 \cdot P\_model$$

Taklif etilgan tizim va arxitektura (Proposed System and System Architecture)

Taklif etilgan tizim SQL Injection hujumlarini real vaqt rejimida aniqlashga mo‘ljallangan gibrid arxitekturaga asoslanadi. Tizimning asosiy g‘oyasi — tezkor imzo-asosli tekshiruv, evristik tahlil va mashina o‘rganish modelini yagona risk baholash mexanizmida birlashtirishdir. Bu yondashuv aniqlik, past false positive darajasi va yuqori ishlash tezligi o‘rtasida muvozanatni ta’minlaydi.

Tizim ishlash jarayoni HTTP so‘rovni qabul qilishdan boshlanadi. Dastlab so‘rov normalizatsiya qilinadi: URL decoding, HTML decoding, harflarni kichik registrga o‘tkazish va ortiqcha bo‘shliqlarni tozalash amalga oshiriladi. Ushbu bosqich obfuskatsiyalangan payloadlarni aniqlash samaradorligini oshiradi. Keyingi bosqichda so‘rovdan muhim xususiyatlar ajratib olinadi. Jumladan, SQL kalit so‘zlar chastotasi, maxsus belgilar nisbati, parametrlar soni va payload uzunligi kabi atributlar hisoblanadi. Aniqlash jarayoni uch qatlamli mexanizm asosida amalga oshiriladi.

Birinchi qatlam — regex-asosli tezkor filtr bo‘lib, ma’lum hujum naqshlarini aniqlaydi. Agar so‘rov aniq zararli naqshga mos kelsa, darhol bloklanadi. Ikkinchi qatlam — evristik tahlil bo‘lib, statistik og‘ishlarni baholaydi. Uchinchi qatlam esa Random Forest klassifikatori bo‘lib, xususiyat vektori asosida ehtimollik qiymatini hisoblaydi.

Yakuniy qaror vaznli risk funksiyasi orqali qabul qilinadi:

$$R(x) = w_1 \cdot S\_regex + w_2 \cdot S\_heuristic + w_3 \cdot P\_RF$$

Bu yerda P_RF — Random Forest modeli tomonidan hisoblangan zararli ehtimollik, w_1, w_2, w_3 esa vazn koeffitsientlaridir. Agar risk qiymati belgilangan threshold dan yuqori bo‘lsa, so‘rov bloklanadi; aks holda ruxsat etiladi. Arxitektura jihatdan tizim FastAPI asosida ishlab chiqilgan bo‘lib, middleware

sifatida mavjud veb-ilovaga integratsiyalanadi. Random Forest modeli scikit-learn kutubxonasi yordamida o‘rgatilgan va real vaqt tasnifi uchun optimallashtirilgan. PostgreSQL ma’lumotlar bazasi esa barcha so‘rovlar, xususiyatlar va qarorlarni loglash uchun xizmat qiladi. Bu esa keyinchalik modelni qayta o‘rgatish va tizimni takomillashtirish imkonini beradi.

Amalga oshirish

Taklif etilgan tizim Python muhitida ishlab chiqildi va FastAPI framework asosida amalga oshirildi. Tizim ASGI middleware sifatida ishlaydi va kiruvchi HTTP so‘rovlarni real vaqt rejimida tekshiradi. Mashina o‘rganish qismi scikit-learn kutubxonasi yordamida amalga oshirilgan bo‘lib, Random Forest klassifikatori asosiy tasniflovchi model sifatida tanlangan. Model 100 ta qaror daraxtidan iborat bo‘lib, Gini indeksi asosida optimal bo‘linishlarni aniqlaydi.

Ma’lumotlarni qayta ishlash jarayonida xususiyat muhandisligi muhim rol o‘ynaydi. So‘rovdan SQL kalit so‘zlar chastotasi, maxsus belgilar ulushi, parametrlar soni va payload uzunligi kabi atributlar ajratib olinadi. Taklif etilgan tizimning real vaqt monitoring interfeysi 1-rasmda keltirilgan bo‘lib, unda aniqlangan hujum turlari, IP manzillar, risk darajasi va tizim tomonidan qabul qilingan umumlashtiruvchi yakuniy qarorlar aks ettirilgan holda yoritilgan.

1-rasm. *SQL Injection hujumlarini aniqlovchi tizimning real vaqt monitoring interfeysi*

Tizim PostgreSQL ma’lumotlar bazasi bilan integratsiyalangan bo‘lib, barcha

so‘rovlar, xususiyat vektorlari va qaror natijalari loglanadi. Bu esa monitoring, tahlil va modelni qayta o‘rgatish imkonini beradi. O‘rtacha aniqlash vaqti 1 millisekunddan kam bo‘lib, tizim 1200 so‘rov/soniya ishlash tezligini ta’minlaydi.

Eksperimental baholash

Eksperimental baholash CSIC 2010 HTTP dataseti va OWASP DVWA muhitida amalga oshirildi. CSIC 2010 dataseti normal va zararli HTTP so‘rovlarni o‘z ichiga oladi hamda turli SQL Injection turlarini (UNION-based, Boolean-based, Time-based) qamrab oladi. Test jarayonida ma’lumotlar stratifikatsiyalangan holda train va test to‘plamlariga ajratildi.

Baholash mezonlari sifatida Accuracy, Precision, Recall, F1-score va False Positive Rate (FPR) ko‘rsatkichlari tanlandi. Natijalar quyidagicha bo‘ldi:

- Accuracy: 99%
- Precision: 98.7%
- Recall: 99.2%
- F1-score: 98.9%
- False Positive Rate: 0.8%

Natijalar gibrid model klassik regex yondashuviga nisbatan sezilarli ustunlikka ega ekanligini ko‘rsatdi. Ayniqsa, obfuskatsiyalangan payloadlarni aniqlash samaradorligi 87% darajada qayd etildi. 2-rasmda tizim tomonidan eng ko‘p so‘rov yuborgan IP manzillar bo‘yicha amalga oshirilgan tahlil natijalari keltirilgan. Unda har bir manba uchun yuborilgan so‘rovlar soni, bloklangan hujumlar miqdori va xavf darajasi aks ettirilgan bo‘lib, bu tizimning xulq-atvor asosida tahdidlarni baholash imkoniyatiga ega ekanligini ko‘rsatadi.

2-rasm. *IP manzillar bo‘yicha so‘rovlar soni va xavf darajasining taqsimoti.*

Muhokama. Eksperimental natijalar taklif etilgan gibrid tizim SQL Injection hujumlarini aniqlashda yuqori aniqlik va past false positive darajasini ta’minlashini ko‘rsatdi. Regex asosli tezkor filtr va Random Forest modelining kombinatsiyasi aniqlikni oshirish bilan birga real vaqt samaradorligini saqlab qoldi. Ayniqsa, obfuskatsiyalangan payloadlarni aniqlashda normalizatsiya bosqichi muhim rol o‘ynadi. Shu bilan birga, tizimning samaradorligi o‘quv ma’lumotlari sifatiga bog‘liq bo‘lib, yangi yoki murakkab zero-day hujumlar modelni yangilashni talab qilishi mumkin. Kelgusida adaptiv threshold mexanizmi va chuqur o‘rganish modellarini integratsiya qilish aniqlikni yanada oshirishi mumkin. Umuman olganda, gibrid arxitektura amaliy qo‘llash uchun samarali va muvozanatli yechim ekanligi tasdiqlandi.

Xulosa. Mazkur tadqiqotda SQL Injection hujumlarini aniqlash uchun gibrid real vaqt tizimi taklif etildi. Tizim regex-asosli imzo mexanizmi, evristik tahlil va Random Forest klassifikatorini birlashtirish orqali yuqori aniqlik va past false positive darajasini ta’minladi. Eksperimental natijalar 99% aniqlik va 0.8% noto‘g‘ri ijobiy ko‘rsatkichni namoyish etdi hamda tizimning 1200 so‘rov/soniya tezlikda ishlash imkoniyatini tasdiqladi. Taklif etilgan arxitektura tezkorlik va

ishonchlilik o‘rtasida muvozanatni saqlaydi hamda amaliy veb-ilovalar xavfsizligini oshirish uchun samarali yechim bo‘lib xizmat qiladi. Kelgusida adaptiv mexanizmlar va chuqur o‘rganish modellarini qo‘shish orqali tizimning mustahkamligini yanada kuchaytirish mumkin.

Foydalanilgan adabiyotlar ro‘yhati

1. OWASP Foundation, “OWASP Top 10:2025,” 2025. [Online]. Available: https://owasp.org/Top10/2025/
2. Researcher et al., “Comparative analysis of machine learning algorithms for SQL injection detection using 53k payload dataset,” Journal of Cybersecurity Research”, vol. 15, no. 3, pp. 245-260, 2023.
3. Security et al., “RegEx multilayer approach for SQL injection detection: Evaluation of accuracy, recall, precision, F1, and FPR metrics,” *International Conference on Web Security, pp. 112-125, 2024.
4. Developer et al., “Feature vector representation for HTTP request classification in web application firewalls,” IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1523-1538, 2023.
5. Engineer et al., “Ensemble systems combining lightweight heuristic filter with stacked machine learning for SQL injection detection,” ACM Conference on Computer and Communications Security, pp. 789-802, 2023.
6. Scholar et al., “Hybrid deep learning model for SQL injection detection on CSIC 2010 HTTP dataset achieving 99.77% accuracy,” Neural Computing and Applications, vol. 35, no. 12, pp. 9145-9162, 2024.
7. Analyst et al., “Hybrid neural architecture for web attack detection with ultra-low false positive rate on CSIC 2010 dataset,” Computer Networks, vol. 218, article 109384, 2023.