Registered in Germany Europa Revue Journal
ISSN: 1862-6564 Vol. 6 No. 10 (2026) Published October 14, 2026

MODELS AND ALGORITHMS FOR DETECTING OBFUSCATED SQL INJECTION ATTACKS IN API REQUESTS BASED ON MULTI-SCALE HYBRID NEURAL NETWORKS AND ATTENTION MECHANISM

Gayibov Islom

Department of Information Security, Urgench State University

Abstract

In modern cloud-native and microservice information systems, data exchange is predominantly executed via REST and GraphQL Application Programming Interfaces (APIs) using structured JSON and XML payloads. However, SQL injection (SQLi) attacks embedded within nested API requests and disguised through multi-stage obfuscation evade traditional regular-expression-based Web Application Firewalls (WAFs). This paper presents a structural decomposition model for hierarchical API requests, an information-theoretic Shannon entropy estimator for obfuscation detection, and a hybrid deep learning architecture integrating Multi-Scale 1D Convolutional Neural Networks (MS-1D-CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and a Bahdanau Attention mechanism. Character-level embedding is utilized to eliminate out-of-vocabulary (OOV) vulnerabilities, while class imbalance is mitigated via Focal Loss. Furthermore, the asymptotic computational complexity of the inline filtration algorithm is formally proved, and a rigorous comparative synthesis against peer-reviewed benchmarks on the CSIC 2010 HTTP dataset is provided.

Keywords

API security, SQL injection (SQLi), obfuscation, Shannon entropy, character-level embedding, Multi-Scale 1D-CNN, Bi-LSTM, Attention mechanism, CSIC 2010.

Annotatsiya: Zamonaviy bulutli va mikroxizmatli axborot tizimlarida ma’lumotlar almashinuvi REST va GraphQL amaliy dasturlash interfeyslari (API) orqali JSON hamda XML tuzilmalari shaklida amalga oshiriladi. Biroq, murakkab iyerarxik API so‘rovlari ichiga yashirilgan va ko‘p bosqichli obfuskatsiya (URL/Hex/Unicode kodlash, SQL sharhlari bilan bo‘laklash) orqali niqoblangan SQL-in’eksiya (SQLi) hujumlari an’anaviy regulyar ifodalarga asoslangan veb-xavfsizlik devorlari (WAF) tomonidan samarali aniqlanmaydi. Ushbu maqolada API so‘rovlarining daraxtsimon

tuzilmasini dekompozitsiya qilish, Shennon axborot entropiyasi orqali kodlangan anomaliyalarni baholash hamda ko‘p masshtabli konvolyutsion tarmoq (Multi-Scale 1D-CNN), ikki yo‘nalishli uzoq-qisqa muddatli xotira (Bi-LSTM) va diqqat (Attention) mexanizmini birlashtiruvchi gibrid chuqur o‘rganish modeli ishlab chiqilgan. Modelda lug‘atdan tashqari (Out-of-Vocabulary) mutatsiyalarni yo‘qotmaslik uchun belgilar darajasidagi embedding (Character-level Embedding) qo‘llanilgan, sinflar nomutanosibligi muammosi esa vaznli fokal yo‘qotish funksiyasi (Focal Loss) orqali hal etilgan. Maqolada algoritmning vaqt va xotira bo‘yicha asimptotik murakkabligi (O-tahlil) nazariy asoslangan hamda CSIC 2010 HTTP xalqaro etalon ma’lumotlar to‘plamida nufuzli ilmiy tadqiqotlar tomonidan qayd etilgan natijalar bilan qiyosiy tahlil keltirilgan.

Kalit so‘zlar: API xavfsizligi, SQL-in’eksiya (SQLi), obfuskatsiya, Shennon entropiyasi, Character-level Embedding, Multi-Scale 1D-CNN, Bi-LSTM, Attention mexanizmi, CSIC 2010.

Axborot-kommunikatsiya texnologiyalarining zamonaviy rivojlanish bosqichida monolit veb-ilovalar o‘rnini taqsimlangan mikroxizmatlar (microservices) va bulutli arxitekturalar egalladi. Ushbu tizimlarda mijoz ilovalari (mobil dasturlar, SPA veb-interfeyslar, IoT shlyuzlar) hamda server ma’lumotlar bazalari o‘rtasidagi muloqot amaliy dasturlash interfeyslari — REST (Representational State Transfer) va GraphQL API orqali amalga oshiriladi.

Xalqaro OWASP (Open Web Application Security Project) tashkilotining rasmiy hisobotlarida in’eksiya zaifliklari (OWASP Top 10:2021 — A03:2021 Injection hamda OWASP API Security Top 10:2023 — API8:2023 / API10:2023) veb va API ekotizimlarining eng xavfli zaifliklari qatorida qayd etilgan [11]. Xususan, SQL-in’eksiya (SQLi) hujumlari orqali tajovuzkorlar API parametrlariga zararli relyatsion so‘rov konstruksiyalarini kiritib, autentifikatsiyani chetlab o‘tish (Authentication Bypass), ma’lumotlar bazasi sxemasini aniqlash va maxfiy axborotni ruxsatsiz sizdirish (Data Exfiltration) imkoniyatiga ega bo‘ladi.

O‘zbekiston Respublikasining 2022-yil 15-apreldagi O‘RQ-764-son «Kiberxavfsizlik to‘g‘risida»gi Qonunida davlat organlari va muhim axborot infratuzilmasi obyektlarida

axborot tizimlari hamda ma’lumotlar bazalarini ruxsatsiz ta’sirlardan himoyalashning zamonaviy mexanizmlarini joriy etish qat’iy belgilangan [1].

Amaliyotda qo‘llaniladigan an’anaviy veb-xavfsizlik devorlari (masalan, ModSecurity OWASP Core Rule Set — CRS) kiruvchi HTTP so‘rovlarini deterministik imzolar (signatures) va regulyar ifodalar (Regular Expressions — Regex) yordamida tekshiradi. Biroq, ilmiy tadqiqotlar shuni ko‘rsatadiki, deterministik WAF tizimlari quyidagi uchta fundamental ilmiy-texnik ziddiyatga duch keladi:

1. Tajovuzkorlar SQL kalit so‘zlarini o‘n oltilik (Hex), URL yoki Unicode formatlarida kodlash, kalit so‘zlar orasiga C-uslubidagi sharhlarni qo‘shish (UNI/**/ON SEL/**/ECT), mantiqiy ekvivalent almashtirishlar (1=1 o‘rniga ASCII('a')=97 yoki 0x50=0x50) orqali regulyar ifodalarni chetlab o‘tadi [2].
2. Strukturaviy murakkablik (JSON/XML Nesting): API so‘rovlarida ma’lumotlar ko‘p qavatli ichma-ich joylashgan (nested) JSON obyektlari va massivlari ichida uzatiladi. Oddiy matn qidiruviga asoslangan filtrlar JSON kalitlari va murakkab matnli qiymatlarni noto‘g‘ri talqin qilib, yuqori soxta trevogalarga (False Positive Rate — FPR) sabab bo‘ladi.
3. Kontekstni tushunmaslik: Regulyar ifodalar so‘rovning semantik kontekstini tahlil qila olmaydi. Masalan, foydalanuvchi sharhida yoki manzil maydonida qonuniy ravishda uchraydigan select, update, from yoki apostrof (') belgilari an’anaviy WAF tomonidan hujum sifatida bloklanadi.

Mazkur ziddiyatni hal etish uchun API so‘rovlarini belgilar darajasida chuqur sintaktik va kontekstual tahlil qiluvchi, obfuskatsiya mutatsiyalariga chidamli neyron tarmoq modellarini yaratish talab etiladi.

Adabiyotlar tahlili va metodologik qiyoslash.

Jahon ilmiy adabiyotlarida SQL-in’eksiya va veb-anomaliyalarni aniqlash usullari evolyutsiyasi uchta asosiy yo‘nalishga bo‘linadi: statik-dinamik sintaktik tahlil, klassik mashinaviy o‘rganish hamda chuqur neyron tarmoqlar. Statik va dinamik dasturiy tahlil usullari. Halfond, Viegas va Orso [2] o‘zlarining fundamental tadqiqotida SQL-in’eksiya hujumlarining 7 ta asosiy taksonomik sinfini (Tautologies, Illegal/Logically Incorrect Queries, Union Query, Piggy-Backed Queries, Stored Procedures,

Inference/Blind SQLi, Alternate Encodings) tasniflab, ularni dastur kodida abstrakt sintaktik daraxtlar (AST) yordamida aniqlashni taklif etgan. Biroq, mikroxizmatlar arxitekturasida API shlyuz (API Gateway) ichki dastur kodiga va ma'lumotlar bazasi sxemasiga to‘g‘ridan-to‘g‘ri ega bo‘lmagani sababli, AST-asosidagi tahlilni tarmoq chekkasida (Edge) qo‘llash imkonsizdir. Klassik mashinaviy o‘rganish (ML) modellari**.** Torrano-Gimenez va hammualliflar [4] Ispaniya Milliy Tadqiqot Kengashida (CSIC) veb-xavfsizlik tizimlarini sinash uchun CSIC 2010 HTTP Dataset etalon bazasini ishlab chiqdilar va statistik anomaliyalarni aniqlash modelini taklif qildilar. Keyingi tadqiqotlarda Campazas-Vega va boshqalar [10] chiziqli va ansambl klassifikatorlarini (Logistic Regression, Random Forest, SVM) o‘rganib chiqdilar. Klassik ML usullarining asosiy kamchiligi — xususiyatlarni qo‘lda ajratish (manual feature engineering) va so‘zlarni TF-IDF (Term Frequency-Inverse Document Frequency) orqali tokenizatsiya qilishga bog‘liqligidir. Hujumchi bitta belgini kodlasa yoki bo‘shliqni sharh bilan almashtirsa, TF-IDF lug‘ati yangi tokenni tanimaydi (Out-of-Vocabulary — OOV muammosi). Chuqur o‘rganish (Deep Learning) arxitekturalari. Matnli ketma-ketliklardan xususiyatlarni avtomatik ajratishda Kim [5] ko‘p o‘lchamli filtrli konvolyutsion neyron tarmoqni (CNN) taklif etdi, Zhang, Zhao va LeCun [6] esa belgilar darajasidagi konvolyutsion tarmoqlar (Character-level CNN) morfologik buzilishlar va kodlangan matnlarni lug‘atsiz tahlil qila olishini isbotladi. Ushbu g‘oyani veb-xavfsizlikka tatbiq etgan holda, Xie va boshqalar [8] Elastic-Pooling CNN (EP-CNN) modelini ishlab chiqib, so‘rovlarni kesmasdan (without truncation) matritsaga o‘tkazish orqali yuqori aniqlikka erishdilar. Biroq, sof CNN modellari faqat lokal n-grammalarni (yonma-yon turgan belgilar guruhini) tutadi va so‘rovning boshi bilan oxiri o‘rtasidagi uzoq masofali mantiqiy bog‘liqlikni yo‘qotadi. Ketma-ketlik xotirasini saqlash uchun Hochreiter va Schmidhuber [3] tomonidan yaratilgan LSTM (Long Short-Term Memory) rekurrent tarmoqlari qo‘llanila boshlandi. Gandhi va boshqalar [9] CNN-BiLSTM gibrid arxitekturasini SQLi tahlili uchun taklif etib, lokal va global xususiyatlar kombinatsiyasi alohida olingan modellardan ustun ekanligini ko‘rsatdi. Shunga qaramay, mavjud CNN-BiLSTM modellarida bir xil o‘lchamli konvolyutsiya yadrosi ishlatilgani sababli turli uzunlikdagi obfuskatsiya qoliplari to‘liq qamrab

olinmaydi, shuningdek, zararli yuklama qismiga selektiv e'tibor beruvchi Attention (Diqqat) mexanizmi [7] va entropiya tahlili integratsiya qilinmagan.

Metodologiya va taklif etilayotgan matematik model

API so'rovlarini strukturaviy dekompozitsiya qilish va entropiya modeli. An'anaviy URL-qatorlardan farqli o'laroq, API so'rovi R murakkab iyerarxik tuzilmaga ega:

$$R = \langle \mathrm{M},\ \mathrm{U},\ \mathrm{H_{hdr}},\ \mathrm{B_{json}} \rangle \quad (1)$$

bu yerda $\boldsymbol{M} \in \{\mathbf{GET, POST, PUT, DELETE, PATCH}\}$ — HTTP metodi, $\boldsymbol{U}$ — URI yo'li va query parametrlari, $\boldsymbol{H_{hdr}}$ — HTTP sarlavhalar to'plami, $\boldsymbol{B_{json}}$ — daraxtsimon JSON yoki XML so'rov tanasi.

JSON obyektini $\boldsymbol{G} = (\boldsymbol{V_G}, \boldsymbol{E_G})$ yo'naltirilgan daraxt sifatida ifodalaymiz, bunda barg tugunlar (leaf nodes) foydalanuvchi tomonidan kiritilgan qiymatlarni saqlaydi. Tahlil oldidan daraxtning faqat qiymat barglari $\boldsymbol{\mathcal{P}} = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \dots, \boldsymbol{p_m}\}$ ajratib olinadi. Hujumchilar ko'p qavatli kodlashdan (masalan, Double URL Encoding: $\%2527 \rightarrow \%27 \rightarrow\ '$) foydalangani sababli, har bir parametr $\boldsymbol{p_i}$ ga rekursiv dekodlash operatori $\boldsymbol{\mathcal{D}^{(r)}}$ qo'llaniladi:

$$s_i = \mathcal{D}^{(r)}(p_i) \quad \text{to} \quad \mathcal{D}^{(r)}(p_i) = \mathcal{D}^{(r-1)}(p_i) \quad (2)$$

Normal API parametrlari tabiiy tilning barqaror taqsimotiga bo'ysunadi. Hex, Base64 yoki murakkab simvollar aralashmasidan iborat SQL-in'eksiya yuklamalari esa belgilar taqsimotining tartibsizligini (entropiyasini) keskin oshiradi. Kiruvchi parametr satri $\boldsymbol{S}$ dagi belgilar chastotasi bo'yicha Shennon entropiyasi $\boldsymbol{H(S)}$ quyidagicha hisoblanadi:

$$H(S) = -\sum_{k=1}^{M} p(s_k) \log_2 p\,(s_k) \quad (3)$$

bu yerda $\boldsymbol{p(s_k)}$ — $\boldsymbol{S}$ satrida $\boldsymbol{s_k}$ belgisining uchrash ehtimolligi, $\boldsymbol{M}$ — satrdagi noyob belgilar soni. Entropiya ko'rsatkichi $\boldsymbol{H(S)}$ hamda maxsus SQL operator belgilariningulushi $\boldsymbol{\rho_{sql}(S)}$ yordamchi statistik deskriptor sifatida yakuniy klassifikatorga uzatiladi.

Belgilar darajasidagi Embedding (Character-Level Embedding). Kanonik normallashtirilgan so'rov satri S uzunligi L ga teng belgilar ketma-ketligiga keltiriladi:

$$S = (c_1, c_2, \dots, c_L), \quad c_t \in \mathcal{V} \quad (4)$$

bu yerda $\boldsymbol{\mathcal{V}}$ — bosma ASCII belgilari alifbosi. Har bir belgi o‘qitiladigan proyeksiya matritsasi $\boldsymbol{W_e} \in \boldsymbol{R^{|\mathcal{V}|\times d}}$ orqali $\boldsymbol{d}$ −o‘lchamli uzluksiz vektor fazoga o‘tkaziladi [6]:

$$E = [e_1, e_2, \dots, e_L]^T \in R^{L\times d} \quad (5)$$

Ko‘p masshtabli (Multi-Scale) 1D-CNN qatlami. SQL-in’eksiya konstruksiyalari turli uzunlikdagi sintaktik birliklardan iborat: qisqa operatorlar (OR, --, /* — 2–3 belgi), o‘rta kalit so‘zlar (UNION, WHERE, SLEEP — 5 belgi) va uzun funksiyalar (SELECT, CONCAT, HAVING — 6–7 belgi). Parallel ravishda uch xil yadro o‘lchamiga ega $k \in \{3, 5, 7\}$ bir o‘lchamli konvolyutsiya qatlamlari qo‘llaniladi [5]. Har bir k-yadro uchun F ta filtrdan iborat $\boldsymbol{W^{(k)}} \in \boldsymbol{R^{k\times d\times F}}$
tenzori orqali xususiyatlar xaritasi hisoblanadi:

$$c_i^{(k)} = \mathrm{ReLU}\left(\sum_{u=0}^{k-1}\sum_{v=1}^{d} W_{u,v}^{(k)} \cdot E_{i+u,v} + b^{(k)}\right) \quad (6)$$

Ketma-ketlik uzunligini qisqartirish va lokal siljishlarga invariantlikni ta’minlash uchun qadami s_p bo‘lgan lokal Max-Pooling amali bajariladi hamda uchala masshtab natijalari bitta boyitilgan xususiyatlar matritsasiga birlashtiriladi:

$$\hat{c}_j^{(k)} = \max_{0\le r<s_p} c_{j\cdot s_p+r}^{(k)}, \quad X_{cnn} = \left[\hat{C}^{(3)} \,\|\, \hat{C}^{(5)} \,\|\, \hat{C}^{(7)}\right] \in \mathbb{R}^{L'\times 3F} \quad (7)$$

bu yerda $\boldsymbol{L' = \lfloor L/s_p \rfloor}$ — siqilgan ketma-ketlik uzunligi.

Ikki yo‘nalishli LSTM (Bi-LSTM) va Bahdanau Attention mexanizmi. Siqilgan lokal xususiyatlar ketma-ketligi $X_{cnn} = (x_1, x_2, \dots, x_{L'})$ uzoq masofali sintaktik qonuniyatlarni o‘rganish uchun Bi-LSTM qatlamiga uzatiladi [3, 12]:

$$\begin{aligned} f_t &= \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ i_t &= \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \\ \tilde{C}_t &= \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \quad (8)$$

To‘g‘ri va teskari LSTM qatlamlari birlashtirilib, ikki yo‘nalishli yashirin holat olinadi:

$$h_t = \left[\overrightarrow{h}_t \,\|\, \overleftarrow{h}_t\right] \in \mathbb{R}^{2u} \quad (9)$$

Zararli segmentga maksimal og‘irlik berish uchun Bahdanau va boshqalar [7] taklif etgan additiv diqqat (Attention) qatlami kiritiladi:

$$u_t = \tanh(W_a h_t + b_a) \quad (10)$$

$$\alpha_t = \frac{\exp(u_t^T v_a)}{\sum_{j=1}^{L'} \exp(u_j^T v_a)}, \quad \sum_{t=1}^{L'} \alpha_t = 1 \quad (11)$$

$$z_{att} = \sum_{t=1}^{L'} \alpha_t h_t \quad (12)$$

bu yerda $\boldsymbol{\alpha_t} \in [\mathbf{0}, \mathbf{1}]$— $\boldsymbol{t}$ -segmentning xavflilik darajasini ko‘rsatuvchi diqqat og‘irligi, $\boldsymbol{v_a}$ — o‘qitiladigan kontekst vektori.

Diqqat vektori $\boldsymbol{z_{att}}$ga (3)-formuladagi Shennon entropiyasi $\boldsymbol{H(S)}$ va maxsus belgilar zichligi $\boldsymbol{\rho_{sql}(S)}$ qo‘shilib, to‘liq bog‘langan qatlam orqali hujum ehtimolligi hisoblanadi:

$$\hat{y} = P(\text{SQLi} \mid R) = \sigma\left(W_{out}\left[z_{att} \,\|\, H(S) \,\|\, \rho_{sql}(S)\right] + b_{out}\right) \quad (13)$$

Modelni o‘qitishda sinflar nomutanosibligini bartaraf etish uchun $\boldsymbol{L_2}$-regulyarizatsiyali Focal Loss funksiyasidan foydalaniladi:

$$\mathcal{L}_{\mathcal{FL}}(y, \hat{y}) = -\alpha_c y(1-\hat{y})^{\gamma} \log(\hat{y}) - (1-\alpha_c)(1-y)y^{\gamma} \widehat{\log(1-\hat{y})} + \frac{\lambda}{2}\|\Theta\|_2^2 \quad (14)$$

Algoritmik amalga oshirish va murakkablik tahlili

Quyida API shlyuzida (Reverse Proxy / Middleware) real vaqt rejimida ishlovchi algoritm keltirilgan:

1. **Qadam (Strukturaviy dekompozitsiya):** Kiruvchi API so‘rovi $\boldsymbol{R} = \langle \boldsymbol{M}, \boldsymbol{U}, \boldsymbol{H}_\{\boldsymbol{hdr}\}, \boldsymbol{B}_\{\boldsymbol{json}\} \rangle$ qabul qilinadi va daraxtning qiymat barglari $\boldsymbol{\mathcal{P}_{list}}$ ajratib olinadi.

2. **Qadam (Dekodlash va entropiya):** Har bir $\boldsymbol{p_i} \in \boldsymbol{\mathcal{P}_{list}}$ uchun (2)-formula asosida rekursiv dekodlash $\boldsymbol{s_i} = \boldsymbol{\mathcal{D}^{(r)}(p_i)}$ bajariladi hamda (3)-formula bo‘yicha Shennon entropiyasi $\boldsymbol{H(s_i)}$ va SQL belgilar zichligi $\boldsymbol{\rho_{sql}}$ hisoblanadi.

3. **Qadam (Vektorlashtirish):** (5)-formula orqali belgilar darajasidagi embedding matritsasi $\boldsymbol{E} \in \boldsymbol{R^{L\times d}}$ shakllantiriladi.

4. **Qadam (Ko‘p masshtabli sintaktik tahlil):** $\boldsymbol{k} \in \{\mathbf{3},\mathbf{5},\mathbf{7}\}$ yadrolar bo‘yicha (6)–(7)-formulalar yordamida konvolyutsiya va Max-Pooling bajarilib, $\boldsymbol{X_{cnn}}$ matritsasi hosil qilinadi.

5. **Qadam (Kontekstual va Attention tahlili):** (8)–(12)-formulalar asosida Bi-LSTM holatlari va Bahdanau Attention vektori $\boldsymbol{z_{att}}$ hamda $\boldsymbol{\alpha_t}$ og‘irliklari hisoblanadi.

6. **Qadam (Qaror qabul qilish):** (13)-formula bo‘yicha xavf ehtimoli $\boldsymbol{P_{risk}}$ aniqlanadi. Agar $\boldsymbol{P_{risk}} \geq \boldsymbol{\tau_{th}}$ bo‘lsa, so‘rov bloklanadi (HTTP 403 Forbidden) va SIEM tizimiga xabar yuboriladi; aks holda so‘rov backend mikroxizmatiga o‘tkaziladi.

Asimptotik hisoblash murakkabligi:

- Rekursiv dekodlash va Shennon entropiyasi: satr uzunligi $\boldsymbol{L}$ ga nisbatan chiziqli — $\boldsymbol{O(L)}$.
- Multi-Scale 1D-CNN qatlami: yadro o‘lchamlari $\boldsymbol{k} \in \{\mathbf{3},\mathbf{5},\mathbf{7}\}$, kirish kanali d va filtrlar soni $\boldsymbol{F}$ uchun murakkablik:

$$T_{CNN} = O\left(L \cdot d \cdot F \cdot \sum_{k\in\{3,5,7\}} k\right) = O(15 \cdot L \cdot d \cdot F) \quad (15)$$

- Bi-LSTM va Attention qatlami: Max-Pooling ($\boldsymbol{s_p} = \mathbf{4}$) qatlami ketma-ketlik uzunligini $\boldsymbol{L' = L/4}$ marta qisqartirishi hisobiga murakkablik quyidagicha ifodalanadi:

$$T_{BiLSTM+Att} = O\left(\frac{L}{s_p} \cdot (3F \cdot u + u^2)\right) \quad (16)$$

Umumiy vaqt murakkabligi kiruvchi satr uzunligiga nisbatan qat’iy chiziqli $\boldsymbol{O(L)}$ tartibda qoladi.

Tahlil va qiyosiy natijalar.

Veb va API hujumlarini aniqlash modellarini tekshirish uchun Ispaniya Milliy Tadqiqot Kengashining Axborot xavfsizligi instituti tomonidan yaratilgan HTTP Dataset CSIC

2010 xalqaro etalonidan foydalaniladi [4]. Ushbu to‘plam elektron tijorat veb-ilovasiga yuborilgan real HTTP so‘rovlaridan iborat bo‘lib, 36 000 ta normal o‘qitish so‘rovlari, 36 000 ta normal test so‘rovlari hamda 25 000 dan ortiq anomal va hujumkor so‘rovlarni (SQLi, XSS, CRLF, Buffer Overflow) o‘z ichiga oladi**.**

Quyidagi 1-jadvalda SQL-in’eksiya va HTTP anomaliyalarini aniqlash bo‘yicha xalqaro ilmiy manbalarda (IEEE, Elsevier) chop etilgan bazaviy va gibrid modellarning tasdiqlangan ko‘rsatkichlari hamda taklif etilayotgan arxitekturaning qiyosiy tahlili keltirilgan.

1-jadval. Xalqaro ilmiy tadqiqotlarda SQL-in’eksiya va veb-hujumlarni aniqlash modellarining qiyosiy tahlili

Tadqiqot / Mualliflar	**Qo‘llanilgan arxitektura va usul**	**Ma’lumotlar bazasi (Dataset)**	**Aniqlik / Samaradorlik ko‘rsatkichlari**	**Asosiy yutuq va arxitekturaviy cheklovlar**
Torrano-Gimenez va boshq. [4]	Statistik anomaliya modeli va qoidalar tahlili	CSIC 2010 HTTP Dataset	Bazaviy filtrlash; obfuskatsiyada yuqori FPR	Etalon dataset yaratilgan; ammo statik qoidalar murakkab kodlashda soxta xatolik beradi.
Xie va boshq. [8]	Elastic-Pooling CNN (EP-CNN, belgilar matritsasi)	Real Web App Logs + SQLi to‘plami	Accuracy: >98,5%	So‘rovni kesmasdan tahlil qiladi; biroq rekurrent xotira (LSTM) qatlami yo‘q.
Gandhi va boshq. [9]	Gibrid CNN-BiLSTM modeli	SQL Injection Payload Dataset	Accuracy: ≈ 98%	Lokal (CNN) va ketma-ketlik (Bi-LSTM) xususiyatlarini birlashtirgan; lekin Attention va entropiya yo‘q.
Campazas-Vega va boshq. [10]	Klassik ML modellari va GridSearchCV optimizatsiyasi	Muvozanatlangan SQLi ma’lumotlari	Detection Rate:> 97%, FAR: < 0,07%	Yuqori tezlikka ega; ammo L7 API payload ichidagi chuqur semantik obfuskatsiyani ochmaydi.
Taklif etilayotgan model	Rekursiv Dekodlash + Shennon Entropiyasi + MS-1DCNN + Bi-LSTM + Attention	CSIC 2010 + JSON API SQLi Benchmark	Gibrid CNN-BiLSTM bazasida (≥ 98%) hisoblash qadamlarini s_p marta qisqartirish	Ko‘p masshtabli konvolyutsiya, entropiya baholash va Bahdanau Attention orqali obfuskatsiyaga chidamlilikni ta’minlaydi.

Jadvaldagi ilmiy natijalar tahlili shuni ko‘rsatadiki, so‘rovni belgilar darajasida konvolyutsion matritsaga o‘tkazish klassik tokenizatsiya (Word2Vec / TF-IDF) usullarining asosiy zaifligi bo‘lgan lug‘atdan tashqari (OOV) muammoni to‘liq bartaraf etadi [6, 8]. CNN qatlamidan keyin Bi-LSTM qatlamini ulash esa uzoq masofali mantiqiy SQL operatorlarini aniqlash aniqligini sezilarli oshiradi [9]. Taklif etilayotgan model ushbu ikki arxitekturani rivojlantirib, unga ko‘p masshtabli filtrlar ($\boldsymbol{k} \in \{\mathbf{3}, \mathbf{5}, \mathbf{7}\}$), Shennon entropiyasi deskriptori hamda Bahdanau Attention [7] qatlamini

qo‘shadi. Natijada model nafaqat hujumni aniqlaydi, balki α_t diqqat koeffitsiyentlari orqali JSON parametrining aynan qaysi qismida in’eksiya joylashganligini ko‘rsatib beradi.

Xulosa.

Mazkur ilmiy tadqiqotda zamonaviy veb-ilovalarning API interfeyslariga yo‘naltirilgan niqoblangan SQL-in’eksiya hujumlarini aniqlashning matematik modeli va algoritmi ishlab chiqildi. Ierarxik JSON API so‘rovlarining qiymat barglarini ajratib olish, ko‘p qavatli kodlashlarni rekursiv dekodlash ($\mathcal{D}^{(r)}$) hamda Shennon axborot entropiyasi $H(S)$ yordamida anomal kodlash zichligini miqdoriy baholash modeli taklif etildi. Turli uzunlikdagi SQL sintaktik qoliplarini parallel ajratuvchi $k \in \{3, 5, 7\}$ o‘lchamli konvolyutsiya qatlamlari, ikki yo‘nalishli rekurrent xotira (Bi-LSTM) va Bahdanau Attention mexanizmi yagona gibrid arxitekturaga integratsiya qilindi. Lokal Max-Pooling qatlami yordamida Bi-LSTM qatlamining hisoblash murakkabligi ketma-ketlik uzunligiga nisbatan s_p barobarga qisqartirilishi asimptotik $O(L)$ tahlil orqali asoslandi hamda Attention og‘irliklari (α_t) orqali xavfsizlik hodisalarini vizual interpretatsiya qilish mexanizmi yaratildi.

Foydalanilgan adabiyotlar ro‘yxati.

1. O‘zbekiston Respublikasining «Kiberxavfsizlik to‘g‘risida»gi Qonuni, 15.04.2022-yildagi O‘RQ-764-son // Qonunchilik ma’lumotlari milliy bazasi, 16.04.2022-y., 03/22/764/0313-son.
2. Halfond W. G., Viegas J., Orso A. A classification of SQL-injection attacks and countermeasures // Proceedings of the IEEE International Symposium on Secure Software Engineering (ISSSE). – Arlington, VA, USA, 2006. – P. 13–15.
3. Hochreiter S., Schmidhuber J. Long short-term memory // Neural Computation. – 1997. – Vol. 9, No. 8. – P. 1735–1780. – DOI: 10.1162/neco.1997.9.8.1735.
4. Torrano-Gimenez C., Perez-Villegas A., Alvarez G. An anomaly-based approach for intrusion detection in web traffic // Journal of Information Assurance and Security. – 2010. – Vol. 5, No. 4. – P. 446–454.

5. Kim Y. Convolutional neural networks for sentence classification // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). – Doha, Qatar, 2014. – P. 1746–1751. – DOI: 10.3115/v1/D14-1181.

6. Zhang X., Zhao J., LeCun Y. Character-level convolutional networks for text classification // Advances in Neural Information Processing Systems (NeurIPS). – 2015. – Vol. 28. – P. 649–657.

7. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate // Proceedings of the 3rd International Conference on Learning Representations (ICLR). – San Diego, CA, USA, 2015. – DOI: 10.48550/arXiv.1409.0473.

8. Xie X., Ren C., Fu Y., Xu J., Guo J. SQL Injection Detection for Web Applications Based on Elastic-Pooling CNN // IEEE Access. – 2019. – Vol. 7. – P. 151475–151481. – DOI: 10.1109/ACCESS.2019.2947527.

9. Gandhi N., Patel J., Sisodiya R., Doshi N., Mishra S. A CNN-BiLSTM based Approach for Detection of SQL Injection Attacks // 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). – Dubai, UAE: IEEE, 2021. – P. 378–383. – DOI: 10.1109/ICCIKE51210.2021.9410675.

10. Campazas-Vega A., Crespo-Martínez I. S., Guerrero-Higueras Á. M., et al. SQL injection attack detection in network flow data // Computers & Security (Elsevier). – 2023. – Vol. 127. – Art. no. 103093. – DOI: 10.1016/j.cose.2023.103093.

11. OWASP Foundation. OWASP API Security Top 10 2023 // Open Web Application Security Project, 2023. – URL: https://owasp.org/API-Security/editions/2023/en/0x11-t10/

12. Gers F. A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM // Neural Computation. – 2000. – Vol. 12, No. 10. – P. 2451–2471. – DOI: 10.1162/089976600300015015.