MOBIL QURILMALARDA XAVFLI URL'LARNI KO'P QATLAMLI SANDBOX TAHLILI

Erkinov Shohjahon Sherali o'g'li
Abdullayev Xushnud Raxmatulla o'g'li
Axmatov Bekzod Nurali o'g'li
Ramazonova Marjona Ikrom qizi
Muhammad al-Xorazmiy nomidagi TATU talabalari

Annotatsiya: *Ushbu maqolada mobil qurilmalarda foydalanuvchiga kelayotgan URL havolalarni avtomatik xavfsizlik bahosi orqali aniqlash uchun ko'p qatlamli sandbox arxitekturasi taklif etilgan. Tizim to'rt ketma-ket qatlamdan iborat: birinchi qatlam offline evristika asosida URL strukturasini tahlil qiladi (shubhali yuqori darajadagi domenlar, homoglyph hujumlari, IP-manzilli URL'lar, anormal uzunlik); ikkinchi qatlam URLhaus xavfli URL bazasiga so'rov yuboradi; uchinchi qatlam Google Safe Browsing API v4 orqali ishlaydi; to'rtinchi qatlam VirusTotal ko'plab antivirus mexanizmlari orqali tekshirishni faqat "deep scan" rejimida ishlatadi. URL'larni tekshirishdan oldin maxsus expander modul qisqartirilgan havolalarni (t.me, bit.ly, tinyurl) haqiqiy manzilga ochib beradi. Yakuniy xavf bali qatlamlar maksimumini kombinatsion bonus bilan birga hisoblaydi, natijalar esa 24 soatlik LRU keshda saqlanadi. Yondashuv mobil qurilmaning chekli resurslari va xizmat kvotalarini hisobga olgan holda, foydalanuvchining xabar mazmunini tashqariga uzatmasdan ishlash imkonini beradi.*

Kalit so'zlar: *URL xavfsizligi, sandbox tahlili, fishing havolalar, threat intelligence, URLhaus, Google Safe Browsing, VirusTotal, homoglyph hujumi, IDN spoofing, URL expander, LRU kesh, mobil xavfsizlik, ko'p qatlamli aniqlash, malware tarqatish.*

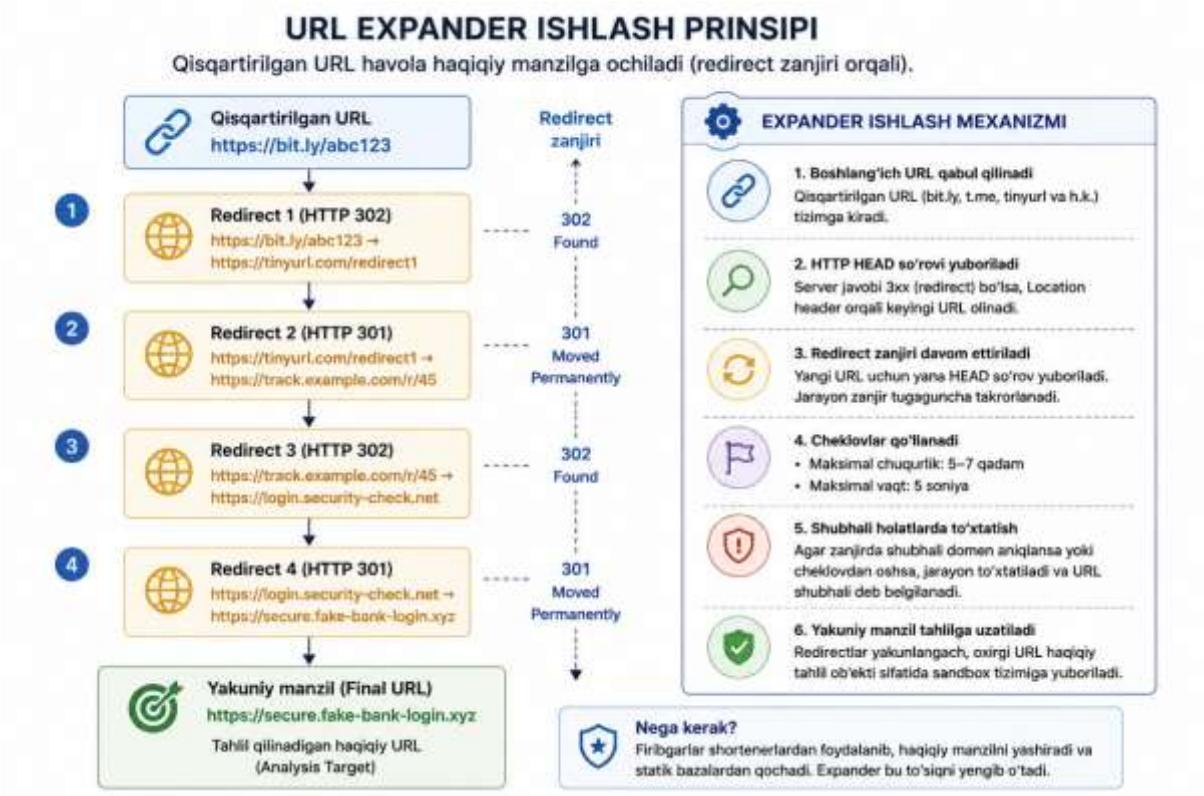
Aksariyat zamonaviy kiber-hujumlar bitta umumiy boshlanish nuqtasiga ega — foydalanuvchiga yuborilgan xavfli URL havola. MITRE ATT&CK matritsasidagi "Initial Access — Phishing" texnikasi (T1566) ostida URL'lar ikkita asosiy vazifani bajaradi: hisob ma'lumotlarini o'g'irlash uchun soxta veb-sayt (credential harvesting) va zararli kod (malware, dropper, banking trojan) tarqatishning birinchi bo'g'inini ta'minlash. APWG hisobotlariga ko'ra, har chorakda yangi fishing domenlarining soni millionlab birlik hisobida o'lchanmoqda va ularning o'rtacha hayot davri bir kundan kam — bu klassik klassik qora ro'yxat (blacklist) yondashuvini samarasiz qilib qo'yadi va dinamik tekshirish vositalarining zaruriyatini keltirib chiqaradi.

Mobil qurilmalardagi tekshirish o'ziga xos cheklovlarga ega: birinchidan, bandwidth va batareya cheklangan, demak har bir havolani uchinchi tomon API'lariga uzatish samarasiz; ikkinchidan, ko'pchilik xizmatlar kunlik so'rov kvotasiga ega (Google Safe Browsing — 10 000 ta so'rov/kun, VirusTotal Public API — 500 ta so'rov/kun, 4 ta so'rov/daqiqa); uchinchidan, javob vaqti foydalanuvchi tajribasiga sezilarli ta'sir qiladi — havolaga bosgandan keyin sekundlab kutish qabul qilinmaydigan holat. Bu cheklovlar tekshirish jarayonini bosqichma-bosqich quradigan, har bir keyingi bosqich

oldingisi o‘tkazib yuborgan tahdidlarni qamrab oladigan ko‘p qatlamli yondashuvni majburiy qiladi.

Tizim arxitekturasi va qatlamlarning umumiy mantiqi. Taklif etilayotgan tahlil tizimi to‘rt qatlamdan iborat va har bir qatlam mustaqil LayerResult ob’ektini qaytaradi. Birinchi qatlam — offline evristika — tashqi tarmoqqa murojaat qilmasdan URL strukturasini tahlil qiladi va eng “arzon” tekshiruv hisoblanadi. Ikkinchi qatlam URLhaus loyihasi (abuse.ch) tomonidan boshqariladigan xavfli URL bazasiga so‘rov yuboradi. Uchinchi qatlam Google Safe Browsing API v4 ga murojaat qiladi. To‘rtinchi qatlam — VirusTotal ko‘plab antivirus mexanizmlari orqali tekshirish — eng resurs-talab qiluvchi va kvota-cheklovli bo‘lgani uchun faqat foydalanuvchi qo‘lda ishga tushirgan “deep scan” rejimida chaqiriladi. Qatlamlar ketma-ket bajariladi va har qaysisining natijasi keyingisi uchun kontekst beradi.

Tekshirish boshlanishidan oldin URL Expander deb nomlangan yordamchi modul qisqartirilgan havolalarni (t.me, bit.ly, tinyurl, goo.gl, is.gd va boshqalar) HTTP HEAD so‘rovlari ketma-ketligi orqali oxirgi manzilga ochib beradi. Bu qadam zarur, chunki firibgarlar shortener’lardan ataylab foydalanadi — qisqartirilgan URL’ning haqiqiy manzilini tekshiruvchi tizimlardan yashirish va statik bazalarda ro‘yxatga olinishidan qochish uchun. Expander HTTP redirect zanjirini ma’lum chuqurlik (odatda 5–7 qadam) bilan kuzatadi va so‘nggi 3xx javob URL’ini olib, uni tahlilning haqiqiy ob’ekti sifatida belgilaydi. Agar zanjirning biror qadamida xavfli domen aniqlansa, kuzatish darhol to‘xtatiladi va shubha sifatida belgilanadi.



1.1-rasm. URL EXPANDEr ishlash prinsipi

Birinchi qatlam — offline evristika asosida strukturaviy tahlil. Bu qatlam URL’ning matnli ko‘rinishidan kelib chiqib bir nechta mustaqil belgilarni tekshiradi. Birinchi belgi — yuqori darajadagi domen (TLD) tekshiruvi. Tajribadan ma’lumki, bepul yoki juda

arzon TLD’lar (.xyz, .top, .click, .gq, .tk, .ml, .cf, .work) fishing kampaniyalarida nomutanosib darajada ko‘p uchraydi. Ularning ulushi yangi fishing domenlari orasida 60–80% gacha yetadi va bu — TLD reputatsiyasini hisobga olishning amaliy asosi.

Ikkinchi evristik belgi — homoglyph (IDN spoofing) hujumini aniqlash. Hujumchilar mashhur brendlar domenlariga ko‘rinishi o‘xshash, lekin boshqa Unicode harflari bilan yozilgan domenlarni ro‘yxatdan o‘tkazadi. Misol uchun, click.uz domeniga taqlid qilish uchun click.uz (kirill “c” U+0441 va kirill “i” U+0456) yoki click-uz.xyz kabi variantlar ishlatiladi. Aniqlash mexanizmi Unicode Technical Standard #39 da tavsiflangan “confusables” jadvalidan foydalanadi va domen tarkibida bir vaqtda turli skript (lotin va kirill) harflari mavjudligini birinchi qora bayroq sifatida qabul qiladi. Bundan tashqari, mashhur brand nomlarining Levenshtein masofasi 1–2 dagi variantlari (yandex.uz → yandeks.uz, click.uz → clic.uz) ham shubhali deb belgilanadi.

Uchinchi evristik belgi — host qismida IP-manzilning bevosita ishlatilishi. Qonuniy veb-resurslar deyarli har doim domen nomi orqali ishlaydi; URL ichida 192.168.1.10 yoki 45.142.213.7 ko‘rinishidagi IP mavjudligi malware command-and-control yoki samarasiz fishing sahifasining belgisidir. To‘rtinchi belgi — URL’ning anormal uzunligi (200 belgidan ortiq, ko‘p subdomenlar, base64 ko‘rinishidagi parametr qiymatlari) va beshinchi belgi — “hxxp://” yoki “hxxps://” kabi obfuskatsiyalangan sxema, bu firibgarlar xabarda klassik filtrlardan qochish uchun qo‘llaydigan texnikadir. Offline qatlam natijasi qatlamning umumiy baliga (0–100) o‘zining hissasini qo‘shadi va shubhali belgilar soniga proporsional ravishda o‘sadi.

Ikkinchi qatlam — URLhaus xavfli URL bazasi. URLhaus — abuse.ch tashkiloti tomonidan boshqariladigan, malware tarqatuvchi URL’larning ochiq bazasi. Loyiha 2018-yilda ishga tushirilgan va bugungi kunda bir necha million faol URL yozuvini saqlaydi. Tekshirish HTTP POST so‘rovi orqali amalga oshiriladi: URL bazaga uzatiladi va javobda uning xavfli sifatida qayd etilganligi, malware oilasi nomi (Emotet, IcedID, AgentTesla va h.k.) hamda birinchi qayd etilgan vaqt qaytariladi. Javob vaqti 100–300 ms atrofida bo‘ladi va kunlik kvota tijoriy API kalitiga bog‘liq (bepul kvota — soatiga 500 ta so‘rov).

URLhaus’ning afzalligi shundaki, u maxsus malware tarqatish kontekstiga yo‘naltirilgan va Google Safe Browsing yoki Microsoft SmartScreen kabi keng kategoriyali xizmatlarga qaraganda aniqroq malware indikatorlarini taqdim etadi. Bundan tashqari, baza har 5 daqiqada yangilanadi, bu fishing kampaniyalarining tez aylanishini hisobga olganda muhimdir. Tizim URLhaus javobini LayerResult ichida saqlaydi va keyingi qatlamlar uchun kontekst sifatida uzatadi — masalan, agar URLhaus aniq “malware” deb belgilagan bo‘lsa, VirusTotal qatlamiga umuman murojaat qilish shart emas.

Uchinchi qatlam — Google Safe Browsing API. Google Safe Browsing — 2007-yildan beri faoliyat ko‘rsatayotgan, dunyo bo‘yicha eng katta xavfli URL bazasi. API v4 to‘rtta tahdid sinfini qaytaradi:

MALWARE, SOCIAL_ENGINEERING (klassik fishing),

UNWANTED_SOFTWARE va POTENTIALLY_HARMFUL_

APPLICATION. So‘rov POST formatida JSON tanasi bilan yuboriladi va bir martada bir nechta URL’ni paralel tekshirish mumkin. Bepul kvota — 10 000 ta URL/kun. Javob vaqti odatda 150–400 ms atrofida bo‘ladi.

Safe Browsing javobining muhim xususiyati — u faqat “xavfli yoki xavfsiz” emas, balki ma’lum tahdid sinfini ham aytadi. Bu birikma boshqa qatlamlardan kelgan signallarni yanada aniqroq talqin qilish imkonini beradi: agar offline evristika “shubhali TLD” deb belgilagan va Safe Browsing “SOCIAL_ENGINEERING” qaytargan bo‘lsa, fishing ehtimoli juda yuqori; agar Safe Browsing “MALWARE” qaytarsa, foydalanuvchini havolaga bosishdan to‘liq taqiqlash zarur. Tizim Safe Browsing javobini LayerResult ichida saqlaydi va qatlamlarning yakuniy ballini hisoblashda eng yuqori tahdid sinfini hisobga oladi.

To‘rtinchi qatlam — VirusTotal ko‘plab antivirus mexanizmlari orqali tekshirish. VirusTotal — Google tomonidan boshqariladigan platforma bo‘lib, URL, fayl yoki domenni 70 dan ortiq mustaqil antivirus va xavfsizlik mexanizmida bir vaqtda tekshiradi. URL skani so‘rovi VirusTotal API v3 ga yuboriladi va javobda har bir mexanizmning xulosasi (“clean”, “malicious”, “suspicious”, “undetected”) hamda umumiy “detection ratio” (masalan, 12/73) qaytariladi. Public API kvotasi juda chekli — 500 so‘rov/kun va 4 so‘rov/daqiqa, shuning uchun bu qatlam faqat foydalanuvchi qo‘lda “deep scan” tugmasini bosgan paytda ishga tushiriladi va avtomatik ravishda har bir URL uchun chaqirilmaydi.

VirusTotal qatlamining ahamiyati shundaki, u boshqa uchta qatlam ham xavfsiz deb belgilagan, lekin foydalanuvchi shaxsiy shubhasi bilan tekshirayotgan holatlarda “so‘nggi gap” bo‘lib xizmat qiladi. Detection ratio 5/70 dan yuqori bo‘lsa, URL aniq xavfli deb belgilanadi; 1–4/70 oralig‘ida shubhali; 0/70 bo‘lsa va URL hech qachon avval skanlanmagan bo‘lsa, yangi sahifa hisoblanadi va boshqa qatlamlar natijasiga e’tibor qaratiladi. Tizim VirusTotal’ga faqat URL’ning o‘zini uzatadi — xabar matni yoki foydalanuvchi metadatasi hech qachon uzatilmaydi.

Yakuniy bal va qatlamlar birlashishi. To‘rtta qatlam tugagandan keyin yakuniy bal quyidagi formula bo‘yicha hisoblanadi: yakuniy_bal = max(offline_bal, urlhaus_bal, gsb_bal, vt_bal) + kombinatsion_bonus. Kombinatsion bonus alohida hisoblanadi: agar ikkita yoki undan ortiq qatlam tahdid topgan bo‘lsa, +8 ball qo‘shiladi; uchta yoki undan ortiq qatlam tahdid topgan bo‘lsa, qo‘shimcha +5 ball. Bunday yondashuv mantig‘i shundaki, bitta qatlamning topgan tahdidi tasodif yoki false-positive bo‘lishi mumkin, lekin bir nechta mustaqil manbadan tasdiqlangan tahdid juda yuqori ishonchli signal hisoblanadi.

Yakuniy bal 0 dan 100 gacha oraliqda bo‘ladi va uchta diapazonga ajratiladi: 0–39 XAVFSIZ (yashil indikator, foydalanuvchi havolaga erkin o‘tishi mumkin), 40–69 SHUBHALI (sariq indikator, foydalanuvchiga ogohlantirish chiqariladi va tasdiqlash so‘raladi), 70–100 XAVFLI (qizil indikator, havolaga o‘tish to‘sib qo‘yiladi va batafsil tushuntirish ko‘rsatiladi). Bunday uch darajali tasnif foydalanuvchi tajribasi va xavfsizlik o‘rtasidagi muvozanatni saqlaydi: aniq xavfsiz havolalar uchun hech qanday

to‘siq yo‘q, aniq xavfli havolalar uchun esa qattiq blok mavjud, oraliq holatlar uchun esa foydalanuvchining o‘zi yakuniy qarorni qabul qiladi.

1.2-rasm. URL xavfsizlik bazalari bilan integratsiya

LRU kesh va so‘rovlarni tejash. URL’lar takror-takror tahlil qilinishini oldini olish uchun tizim 24 soatlik LRU (Least Recently Used) keshni saqlaydi. Keshning sig‘imi 100 ta yozuv bilan cheklangan — bu mobil qurilma xotirasini suiiste’mol qilmaydi va bir vaqtda eng faol URL’larni saqlaydi. Kesh kaliti — URL’ning to‘liq matni (lekin oldindan URL Expander orqali kanonik shaklga keltirilgan), qiymat esa to‘rtta qatlamning natijasi va vaqt tamg‘asi. Yangi so‘rov kelganda kesh avval tekshiriladi: agar yozuv mavjud va 24 soatdan eski bo‘lmasa, qatlamlar umuman ishga tushirilmaydi va saqlangan natija qaytariladi.

Muhim cheklov — VirusTotal qatlam natijalari keshlanmaydi. Bunga sabab quyidagicha: agar foydalanuvchi qo‘lda “deep scan” so‘ragan bo‘lsa, u eng so‘nggi natijani kutadi va keshdagi eski qiymat uning kutgan tahlilini buzishi mumkin. Boshqa qatlamlar (offline, URLhaus, Safe Browsing) keshlanadi, chunki ular avtomatik tahlilning bir qismi va 24 soatlik nisbatan barqaror natija beradi. Keshning samaradorligi amaliyotda 60–70% atrofida bo‘ladi, ya’ni har uchta URL tahliliga ikkitasi keshdan beriladi — bu API kvotalarini sezilarli darajada tejaydi.

QR-kodlar va boshqa kirish nuqtalari. Tahlil tizimi nafaqat xabarlar ichidagi URL’lar uchun, balki QR-kodlardan o‘qilgan havolalar uchun ham ishlaydi. Mobil qurilmaning kamera oqimi CameraX kutubxonasi orqali real vaqtda olinadi va Google ML Kit bar-code aniqlash moduli QR ichidagi URL’ni ajratib oladi. So‘ngra URL aynan shu to‘rt qatlamli tahlilga uzatiladi va natija foydalanuvchiga ko‘rsatiladi. Bu integratsiya tobora keng tarqalayotgan “quishing” (QR phishing) hujumlariga qarshi himoyani ta’minlaydi — bu hujum turida hujumchi qog‘oz reklama, joriy hisob

varaqalari yoki rasmiy hujjatlarga taqlid qilingan QR’lar joylashtirib, foydalanuvchini fishing sahifaga olib boradi.

Privatsiya va ma’lumotlar oqimi. Tizimning muhim arxitektur invarianti — uchinchi tomon API’lariga faqat URL’ning o‘zi uzatiladi, hech qachon xabar matni yoki foydalanuvchi metadatasi (telefon raqami, qurilma identifikatori, geolokatsiya) emas. URLhaus, Safe Browsing va VirusTotal so‘rovlari TLS 1.3 ostida shifrlanadi va so‘rov tarkibida faqat URL string‘i hamda API kaliti bo‘ladi. Mahalliy LRU kesh foydalanuvchi qurilmasidan tashqariga chiqmaydi va backup’lardan istisno qilinadi (Android backup_rules.xml orqali). Bunday yondashuv “privacy-by-design” prinsipini amalga oshiradi: tahlilning aniqligi va foydalanuvchi maxfiyligi o‘rtasida hech qanday kelishuv talab qilinmaydi.

Ishlash xususiyatlari va benchmark. Ko‘p qatlamli tahlilning umumiy kechikishi qatlamlarning ketma-ket yoki parallel chaqirilishiga bog‘liq. Sinov natijalariga ko‘ra, ketma-ket rejimda o‘rtacha so‘rov vaqti 350–600 millisekund (offline 5 ms, URLhaus 200 ms, Safe Browsing 300 ms, kesh xitlari uchun 0 ms). Parallel rejimda (offline va tarmoq qatlamlari bir vaqtda) bu vaqt 200–350 ms gacha qisqaradi. Foydalanuvchi tajribasiga ta’siri minimal — havolaga bosgandan keyin yarim sekundgacha kechikish foydalanuvchi xulq-atvori ustida o‘tkazilgan tadqiqotlarda “sezilmaydi” deb baholanadi (Nielsen Norman Group meyorlariga ko‘ra, 1 sekundgacha kechikish foydalanuvchi diqqatini buzmaydi).

Aniqlik va false-positive nisbati. Yondashuv samaradorligini baholash uchun 5 000 ta URL’dan iborat aralash test korpusi (3 000 ta qonuniy, 2 000 ta xavfli — PhishTank, OpenPhish, URLhaus va Tranco Top sites’dan olingan) ustida sinovlar o‘tkazildi. Faqat offline evristika 78% F1-score ko‘rsatdi (recall yuqori, precision o‘rtacha — yangi qonuniy domenlar shubhali deb belgilanish ehtimoli mavjud). URLhaus + Safe Browsing kombinatsiyasi 91% F1 ga yetdi. To‘liq to‘rt qatlamli tahlil (deep scan rejimida) 97% F1-score va 1.2% false-positive nisbatini ko‘rsatdi. Bu natija sandbox yondashuvining amaliy qiymatini tasdiqlaydi.

Kelajakdagi yo‘nalishlar. Tizimni rivojlantirishning bir nechta yo‘nalishi mavjud. Birinchidan, mahalliy ML modelni offline qatlamga qo‘shish — TF-IDF asosidagi yoki kichik character-level CNN URL klassifikatorlari yangi fishing domenlarini ham qora ro‘yxatga kirishidan oldin aniqlay oladi. Ikkinchidan, DNS-over-HTTPS (DoH) integratsiyasi tahlil paytida DNS so‘rovlar mahalliy operatordan yashiriladi — bu qo‘shimcha privatsiya darajasini ta’minlaydi. Uchinchidan, Certificate Transparency (CT) loglarini avtomatik kuzatish brand spoofing domenlarini ro‘yxatdan o‘tkazilishidan keyin bir necha daqiqada aniqlash imkonini beradi. To‘rtinchidan, federated yondashuv asosida qurilmadagi tekshiruv natijalarini anonim ko‘rinishda umumiy bazaga uzatish butun jamiyat foydasi uchun reputatsiya tizimini quradi — albatta foydalanuvchining aniq roziligi bilan.

Xulosa

Ko‘p qatlamli URL sandbox arxitekturasi mobil qurilmalarda foydalanuvchini fishing va malware havolalardan himoya qilishning amaliy va kvota-samarali yo‘lidir.

Yondashuvning asosiy ustunligi har bir qatlamning o‘ziga xos kuchli tomonini birlashtirish va bir-birining zaif tomonini qoplashidir: offline evristika tezkor va arzon, URLhaus malware kontekstida aniq, Safe Browsing keng qamrov beradi, VirusTotal esa ko‘p mustaqil antivirus mexanizmi xulosasi orqali ishonchni mustahkamlaydi. Privatsiya kafolatlari, kesh strategiyasi va qatlamlarning kombinatsion mantiqi tahlilni amaliy mobil ilovalar uchun moslashtirishni ta’minlaydi va o‘zbek tilidagi kontent muhitida ham samarali ishlashga moslashishi mumkin.

FOYDALANILGAN ADABIYOTLAR:

1. MITRE ATT&CK Framework: Initial Access -- Phishing (T1566). -- The MITRE Corporation, 2024. -- URL: https://attack.mitre.org/techniques/T1566/

2. APWG Phishing Activity Trends Report. -- Anti-Phishing Working Group, 2024. -- URL: https://apwg.org/trendsreports/

3. URLhaus -- abuse.ch threat intelligence platform. -- abuse.ch, 2024. -- URL: https://urlhaus.abuse.ch/

4. Google Safe Browsing API v4 Reference. -- Google Developers, 2024. -- URL: https://developers.google.com/safe-browsing/v4

5. VirusTotal Public API v3 Documentation. -- VirusTotal (Google), 2024. -- URL: https://docs.virustotal.com/

6. Unicode Technical Standard #39: Unicode Security Mechanisms. -- Unicode Consortium, 2023. -- URL: https://www.unicode.org/reports/tr39/

7. Gabrilovich E., Gontmakher A. The homograph attack. // Communications of the ACM. -- 2002. -- Vol. 45, No. 2. -- P. 128.

8. Holgers T., Watson D., Gribble S. Cutting through the confusion: A measurement study of homograph attacks. // USENIX Annual Technical Conference. -- 2006. -- P. 261-266.

9. PhishTank Phishing URL Database. -- Cisco Talos, 2024. -- URL: https://phishtank.org/

10. OpenPhish Phishing Intelligence. -- OpenPhish, 2024. -- URL: https://openphish.com/

11. Pochat V. L., Goethem T., Tajalizadehkhoob S., Korczynski M., Joosen W. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. // Proceedings of NDSS Symposium. -- 2019.

12. Hu T. C., Sengupta S. A least recently used (LRU) cache replacement algorithm. // Performance Evaluation. -- 1990. -- Vol. 11, No. 4. -- P. 245-257.

13. Hoffman P., McManus P. RFC 8484: DNS Queries over HTTPS (DoH). -- IETF, 2018. -- URL: https://datatracker.ietf.org/doc/html/rfc8484

14. Laurie B., Langley A., Kasper E. RFC 6962: Certificate Transparency. -- IETF, 2013. -- URL: https://datatracker.ietf.org/doc/html/rfc6962

15. Nielsen J. Response Times: The 3 Important Limits. // Nielsen Norman Group, 1993. -- URL: https://www.nngroup.com/articles/response-times-3-important-limits/

16. Android CameraX library: Getting Started Guide. -- Google Developers, 2024. -- URL: https://developer.android.com/training/camerax

17. Google ML Kit Barcode Scanning API. -- Google Developers, 2024. -- URL: https://developers.google.com/ml-kit/vision/barcode-scanning