VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ MAGISTRO STUDIJŲ PROGRAMA

Naudotojo apsauga nuo duomenų viliojimo atakų

Users' Protection Against Phishing Attacks

Magistro baigiamasis darbas

Atliko: Andrius Tumšys

Darbo vadovas: doc. dr. Kristina Lapin

Recenzentas: doc. dr. Vytautas Čyras

Vilnius – 2026

Turinys

Santrauka

Šiame magistro darbe nagrinėjama duomenų viliojimo (angl. *phishing*) atakų problema ir jų prevencijos metodai, orientuoti į galutinio naudotojo mokymą ir atsparios naudotojo sąsajos projektavimo gaires. Tyrimo metu išanalizuota 12 esamų apsaugos įrankių ir nustatyta, kad didžioji jų dalis apsiriboja techniniais apsaugos sprendimais — trūksta aiškinamojo dizaino ir mokomojo komponento, nors būtent šių savybių derinys literatūroje identifikuojamas kaip veiksmingiausias ilgalaikiam naudotojų atsparumui formuoti. Sprendžiant šią problemą buvo kuriamas hibridinis metodas, derinantis proaktyvią apsaugą su kontekstiniais edukaciniais elementais. Atsižvelgiant į tai, suformuluotos 33 projektavimo gairės su taikymo rekomendacijomis, apimančios techninį aptikimą, aiškinamąjį dizainą bei elgsenos elementus, ir suskirstytos pagal tikslinę auditoriją ir kuriamo įrankio tipą.

Remiantis šiomis gairėmis sukurta *Chrome* naršyklės plėtinio architektūra bei praktinis įgyvendinimas. Plėtinys analizuoja internetinius puslapius realiu laiku, o nustatytas grėsmes komunikuoja pasitelkdamas vizualinius indikatorius, iššokančius pranešimus ir šoninę sekciją su detaliu mokomuoju paaiškinimu. Įskiepis testuotas su skirtingų IT kompetencijų grupių naudotojais naudojant trifazį testavimo modelį. Rezultatai parodė, kad įrankis reikšmingai padidino grėsmių atpažinimo tikslumą ir sutrumpino sprendimo priėmimo laiką, o didžiausią naudą patyrė mažesnę IT patirtį turintys dalyviai. Taip pat patvirtinta, jog savalaikis mokymas užtikrina žinių išlaikymą — net ir be plėtinio pagalbos atpažinimo tikslumas išliko aukštas. Tyrimo rezultatai patvirtina hibridinio metodo, derinančio proaktyvų blokavimą su kontekstiniais edukaciniais įspėjimais, veiksmingumą.

Raktiniai žodžiai: duomenų viliojimas, naršyklės plėtinys, aiškinamasis dizainas, kibernetinis saugumas, naudotojų mokymas, naudotojo sąsaja.

Summary

This master's thesis examines the problem of phishing attacks and their prevention methods, focusing on end-user education and resilient user interface design guidelines. The research analyzed 12 existing security tools, revealing that the majority rely solely on technical protection measures, lacking explanatory design and educational components — the combination of which is identified in the literature as most effective for building long-term user resilience. To address this gap, a hybrid approach was developed, combining proactive protection with contextual educational elements. Consequently, 33 design guidelines were formulated along with application recommendations, encompassing technical detection, explainable design, and behavioral psychology elements, organized by target audience and tool type.

Using these guidelines, a *Chrome* browser extension was developed. The extension analyzes web pages in real-time and communicates identified threats using visual indicators, pop-up messages, and a side panel with detailed educational explanations. The extension was tested with participants of varying IT expertise levels using a three-phase testing model. The results demonstrated that the tool significantly increased threat detection accuracy and reduced decision-making time, with the greatest benefit observed among participants with less IT experience. Furthermore, the testing confirmed that just-in-time education ensures knowledge retention, as participants maintained a high accuracy rate even without the extension's assistance. The findings confirm the effectiveness of a hybrid approach combining proactive blocking with contextual educational warnings in cybersecurity.

Keywords: phishing, browser extension, explanatory design, cybersecurity, user education, user interface.

Įvadas

Tyrimo objektas

Aiškinamieji ir duomenų viliojimui atsparūs sąsajų projektavimo metodai, skirti naudotojų sukčiavimo aptikimo įgūdžiams gerinti, ir priemonės, skirtos automatizuoti sukčiavimo svetainių tikrinimą.

Tyrimo hipotezė

Aiškinamųjų naudotojo sąsajų integravimas su duomenų viliojimui atspariais projektavimo metodais gerokai sumažina sėkmingų duomenų viliojimo atakų tikimybę.

Tyrimo aktualumas

Duomenų viliojimo atakos tebėra rimta kibernetinio saugumo problema, dėl kurios visame pasaulyje patiriama finansinių nuostolių ir gadinama įmonių reputacija. Nors techninės saugumo priemonės, tokios kaip nepageidaujamų laiškų filtrai ir ugniasienės, patobulėjo, pagrindinė silpnoji vieta dažnai yra žmonės [DFM+22; Hon12]. Užpuolikai naudojasi žmonių patiklumu, smalsumu ir pasimetimu, kad apgaule priverstų naudotojus atskleisti jautrią informaciją [Mun19]. Todėl akivaizdu, kad reikia metodų, kurie padėtų naudotojams atpažinti šias grėsmes ir jų išvengti.

Naujausiuose tyrimuose siūlomi du pagrindiniai būdai šiai problemai spręsti. Pirmasis — tai naudotojų mokymas ir jų gebėjimo atpažinti įtartinus pranešimus ir svetaines gerinimas. Tai apima aiškinamojo dizaino elementų, pavyzdžiui, paprastų ir aiškių įspėjimų ir naudingų patarimų naudojimą naudotojo sąsajoje, paaiškinančių, kodėl kažkas gali būti nesaugu [AAD20]. Patarimas gali parodyti, kodėl tam tikra nuoroda yra įtartina, išryškindamas neįprastus domenų pavadinimus arba neatitinkantį prekės ženklą. Šie aiškinamieji elementai, aiškiai pateikiantys įspėjimus ir paaiškinimus, padeda naudotojams patiems atpažinti bandymus sukčiauti ir ugdyti ilgalaikius aptikimo įgūdžius.

Antrasis metodas apima sukčiavimui atsparios naudotojo sąsajos dizainą. Paprasti patobulinimai, pavyzdžiui, vizualiai išryškintas tikrasis svetainės domenas, vienodų prekės ženklų naudojimas arba aiškių tikrinimo piktogramų rodymas gali padėti naudotojams greitai atskirti tikrą svetainę nuo netikros [MKK08]. Derinant šiuos sukčiavimui atsparaus dizaino modelius su aiškinamosiomis funkcijomis, sukuriamas veiksmingas metodas: vienas priemonių rinkinys tiesiogiai užkerta kelią suklaidinimui, o kitas moko naudotojus suprasti, kodėl šios priemonės yra svarbios.

Taip pat yra įvairių įrankių ir paslaugų, padedančių naudotojams išvengti duomenų viliojimo atakų. Kai kurios jų remiasi dideliais viešai prieinamais žinomų sukčiavimo svetainių sąrašais ([Ope25; Phi25]), blokuoja prieigą arba rodo perspėjimus prieš naudotojui įvedant asmeninę informaciją. Tačiau daugelis šių priemonių nepaaiškina, kodėl svetainė blokuojama, arba aktyviai nepadeda naudotojams mokytis aptikimo įgūdžių. Naudotojai gali pradėti aklai pasitikėti šiomis priemonėmis, užuot supratę, kaip apsisaugoti.

Šiame tyrime bus nagrinėjamos dabartinės sukčiavimo aptikimo paslaugos — kaip jos veikia, kokių privalumų ir kokių trūkumų turi — ir ieškoma būdų, kaip jas suderinti su aiškinamuoju ir sukčiavimui atspariu naudotojo sąsajos dizainu. Naudodamiesi žinomomis sukčiavimo svetainių duomenų bazėmis ir bandydami naujus sąsajos metodus, galėsime įvertinti, ar naudotojai geriau atpažįsta įtartinas svetaines. Siekiama ne tik apsaugoti naudotojus, bet ir laikui bėgant padaryti juos sąmoningesnius ir labiau pasitikinčius savimi. Aiškinamasis dizainas — tai būdas pateikti informaciją taip, kad ją būtų lengva suprasti. Naudojamas aiškus tekstas, vaizdai ir struktūra, kad padėtų žmonėms išmokti sudėtingesnius dalykus [Gar10]. Galiausiai, integravus aiškinamąjį dizainą, sukčiavimui atsparius šablonus ir automatinio tikrinimo priemones, galima sukurti sąsajas, kurios padės naudotojams patiems tapti geriausia apsauga nuo duomenų viliojimo atakų.

Tikslas

Tyrimo tikslas yra naudotojų apsaugos nuo duomenų viliojimo atakų problemos analizė ir metodas, skirtas projektuoti sąsajoms, skirtoms apsaugoti naudotojus nuo duomenų viliojimo atakų.

Uždaviniai

1. Išanalizuoti esamas mokslines publikacijas apie duomenų viliojimo atakas, naudotojų mokymą, naudotojo sąsajos dizaino principus, aiškinamuosius ir duomenų viliojimui atsparaus dizaino modelius, siekiant suprasti šiuo metu taikomus principus ir jų spragas.
2. Išanalizuoti esamas paslaugas, padedančias naudotojams aptikti duomenų viliojimo atakas, jų privalumus ir trūkumus.
3. Ištirti naudotojo sąsajos sprendimų veiksmingumą, kuriant ir testuojant alternatyviuosius maketus.
4. Sukurti naudotojo sąsajos projektavimo metodą, į kurį būtų įtrauktos aiškinamosios ir nuo duomenų viliojimo apsaugančios savybės.
5. Išbandyti metodą kuriant naršyklės įskiepį, kuris padėtų naudotojams atpažinti duomenų viliojimo atakas.

Tyrimo metodai

Šis tyrimas apima analitinius ir empirinius tyrimo metodus. Analizuosime mokslinę literatūrą ir kitus šaltinius, kad suprastume nagrinėjamą sritį, nustatysime svarbius aiškinamųjų ir nuo duomenų viliojimo atakų atsparių dizaino metodų aspektus, ieškosime panašių tyrimų, nagrinėsime jų įgyvendinimą ir rezultatus. Be to, bus renkamos dabartinės paslaugos, padedančios naudotojams atpažinti galimas duomenų viliojimo atakas, taip pat jų privalumai ir trūkumai. Atliekant eksperimentus bus naudojamos viešai prieinamos duomenų viliojimo svetainių duomenų bazės, kad būtų įvertintas skirtingų naudotojo sąsajos dizainų veiksmingumas. Be to, bus parengtos aiškinamosios ir sukčiavimui atsparios naudotojo sąsajos gairės ir jų naudojimo rekomendacijos,

kurios bus naudojamos kuriant naršyklės įskiepį, padedantį naudotojams aptikti galimas duomenų viliojimo atakas.

Siekiami rezultatai

Šiame tyrime bus nustatytos esamos paslaugos, padedančios naudotojams išvengti duomenų viliojimo atakų, jų privalumai ir trūkumai. Bus parengtos aiškinamojo ir duomenų viliojimui atsparaus projektavimo gairės ir jų panaudojimo rekomendacijos ir jomis remiantis sukurtas naršyklės įskiepis, padedantis naudotojams aptikti galimas duomenų viliojimo atakas.

1. Duomenų viliojimo atakų literatūros apžvalga

1.1. Duomenų viliojimo atakų iššūkiai kibernetiniam saugumui

Duomenų viliojimo atakos yra paplitusi kibernetinė grėsmė, kurios mastas ir poveikis vis didėja. Naujausiose ataskaitose šis sukčiavimo būdas yra laikomas labiausiai paplitusiu kibernetiniu nusikaltimu ir pagrindine saugumo pažeidimų priežastimi [BBK+25]. 2020 metais duomenų viliojimo atakos buvo dažniausiai pranešamas kibernetinis nusikaltimas, dėl kurio buvo padaryta 43 % duomenų saugumo pažeidimų, kurie sukėlė didelius finansinius nuostolius įmonių elektroninio pašto pažeidimų atvejais. Technologinės apsaugos priemonės, tokios kaip elektroninių laiškų filtrai, blokavimo sąrašai ir kita, tobulėja, tačiau užpuolikai vis dažniau geba jas apeiti taikydamiesi į žmogiškuosius veiksnius, kurie išlieka silpniausia apsaugos sistemų grandimi [BBK+25]. Nors ir aptikimo priemonės užfiksuoja didžiąją dalį duomenų viliojimo atakų, naudotojai vis tiek susiduria su šios atakos žinutėmis ir būtent naudotojo veiksmai nulemia atakos sėkmę [BBK+25]. Tyrimai rodo, jog žmonės dažnai sunkiai atpažįsta gerai parengtus duomenų viliojimo tinklapius. Tai reiškia, kad vien automatinių aptikimo sistemų nepakanka [RTN25]. Todėl labai svarbu sutelkti dėmesį į naudotojus orientuotą apsaugą, nes duomenų viliojimo atakos naudojasi žmonių patiklumu ir klaida, o ne sistemos pažeidžiamumu [GST+18].

1.2. Duomenų viliojimo atakų tipai

Duomenų viliojimo atakas galima skirstyti pagal naudojamą jų perdavimo priemonę ar techniką. Jos skirstomos į kelias pagrindines kategorijas, tokias kaip socialinės inžinerijos pagrindu vykdomas (apgaulė, žmogaus pasitikėjimo išnaudojimas) bei techninės klastos *(angl. technical subterfuge)* atakas, kurios išnaudoja techninį pažeidžiamumą [AHN+21]. Dažniausiai literatūroje pasitaikančių duomenų viliojimo atakų tipai:

1. Duomenų viliojimas elektroniniais laiškais *(angl. deceptive phishing)*;
2. Tikslingai nukreiptos duomenų viliojimo atakos *(angl. spear phishing)*;
3. Tikslingai nukreiptos atakos į svarbius asmenis ar įstaigas *(angl. whaling)*;
4. Duomenų viliojimas klonavimo pagalba *(angl. clone phishing)*;
5. Duomenų viliojimas SMS žinutėmis *(angl. SMiShing or SMS phishing)*;
6. Duomenų viliojimas balso pranešimais ar skambučiais *(angl. vishing or voice phishing)*;
7. Duomenų viliojimas naudojantis socialiniais tinklais *(angl. soshing or Social media phishing)*;
8. Duomenų viliojimas kenkėjiškomis programomis *(angl. malware-based phishing)*;
9. Peradresavimo duomenų viliojimo atakos *(angl. DNS-based phishing)*;
10. Duomenų viliojimas įterpiant turinį *(angl. content-injection phishing)*;
11. Duomenų viliojimas tarp;ininko metodu *(angl. man-in-the-middle phishing)*;
12. Duomenų viliojimas paieškų sistemose *(angl. search engine phishing)*.

1.2.1. Duomenų viliojimas elektroniniais laiškais

Duomenų viliojimas elektroniniais laiškais yra labiausiai paplitusi atakos forma. Sukčiai masiškai siunčia suklastotus elektroninius laiškus, apsimesdami patikimais asmenimis ar organizacijomis ir skatina aukas perduoti jautrią informaciją [AHN+21]. Laiškuose dažnai raginama neatidėliojant imtis įvairių veiksmų, pavyzdžiui patvirtinti paskyrą ar atnaujinti informaciją, tokiu būdu verčiant naudotoją spausti kenkėjiškas nuorodas ar prisegtukus [AHN+21]. Įtikinant auką laiško autentiškumu, naudojantis oficialiais įmonių logotipais ir atpažįstamomis frazėmis, sukčius nukreipia naudotoją į netikrą svetainę, kurioje prašoma suvesti įvairius prisijungimo, banko kortelių duomenis ar kitą asmeninę informaciją [AHN+21]. Šis modelis dar vadinamas apgaulingu duomenų viliojimu *(angl. deceptive phishing)*, pabrėžiant socialinės inžinerijos, o ne techninio pažeidžiamumo naudojimą [AHN+21].

1.2.2. Tikslingai nukreiptos duomenų viliojimo atakos

Tikslingai nukreiptos duomenų viliojimo atakos yra elektroninių laiškų atakos atmaina, kuri yra vykdoma prieš konkrečius asmenis ar nedidelę jų grupę. Tokie laiškai yra personalizuoti, jiems naudojama informacija apie auką, pavyzdžiui vardai, įmonė kurioje jis dirba ar neseniai vykdyti projektai ir veiklos. Tokiu būdu ataka atrodo dar įtikinamesnė [AHN+21]. Sukčiai iš anksto renka duomenis apie auką iš įvairių viešų šaltinių ar persekiojimo, taip didindami tikimybę, kad adresatas paspaus kenksmingą nuorodą ar perduos informaciją [RRJ+20]. Kadangi pranešimai yra asmeniniai, jie rečiau kelia įtarimą ir pasižymi didesniu sėkmingos atakos rodikliu. Tokiai atakai vykdyti reikia planavimo etapo, kurio metu sukčiai skiria laiko asmeninės informacijos paieškai prieš pradedant duomenų viliojimo ataką [AHN+21].

1.2.3. Tikslingai nukreiptos atakos į svarbius asmenis ar įstaigas

Tikslingai nukreiptos atakos į svarbius asmenis ar įstaigas yra itin selektyvus tikslingo duomenų viliojimo variantas, skirtas aukšto rango vadovams ar svarbiems asmenims *(angl. big fish)*, pavyzdžiui, generaliniams direktoriams, finansų vadovams ar valstybės pareigūnams [AHN+21]. Šios atakos remiasi ypatingu suasmeninimu ir įvairia vidine informacija, neretai surinkta iš ne viešai prieinamų šaltinių, apsimetant kolegomis ar patikimais asmenimis aukštose pareigose [RRJ+20]. Kadangi taikiniai gali prieiti prie itin vertingos informacijos ar turi finansinių įgaliojimų, sėkmingos tokio tipo atakos gali sukelti didelę žalą. Laiškai būna kuriami itin kruopščiai, atsižvelgiant į toną bei kontekstą, dažnai imituoja svarbius verslo susirašinėjimus ar teismo šaukimus, kad nesukeltų įtarimo. Tokio tipo atakose yra pasitelkiamos specifinės žinios, kurios didina pasitikėjimą ir išnaudoja aukos privilegijų galimybės [RRJ+20].

1.2.4. Duomenų viliojimas klonavimo pagalba

Duomenų viliojimas klonavimo pagalba — tai specializuota elektroninių laiškų atakų technika, kai tikri pranešimai kuriuos auka yra gavę ir anksčiau yra nukopijuojami ir suklastojami [AHN+21]. Sukčiai naudodami tikrą laišką, dažnai su prisegtuku ar nuoroda, sukuria beveik identišką jo kopiją,

tačiau pakeičia nuorodą ar prisegtuką į kenkėjišką taip pat padirbdamas siuntėjo adresą [AHN+21]. Gavėjas atpažįsta laišką, nes jis atrodo kaip neseniai matyta žinutė ar pokalbio pratęsimas, todėl yra mažiau įtartina. Taip yra išnaudojamas pirminio bendravimo pasitikėjimas ir auka yra skatinama paspausti kenkėjišką nuorodą, tokiu būdu išviliojant informaciją ar įdiegiant programą [AHN+21].

1.2.5. Duomenų viliojimas SMS žinutėmis

Duomenų viliojimo SMS žinutėmis atakų metų perdavimui naudojamos trumposios SMS žinutės. Sukčiai siunčia suklastotas SMS, apsimesdami patikimais asmenimis ar organizacija, pavyzdžiui bankais, siuntų tarnybomis ar valdžios institucijomis, tokiu būdu siekdamas priversti auką paspausti kenkėjišką nuorodą ar atskleisti asmeninę informaciją [TCR24]. Šios žinutės turi itin didelį jų atidarymo ir greito perskaitymo rodiklį, todėl aukos neretai užklumpamos netikėtai [TCR24]. Žinutėse dažnai pabrėžiamas skubumas ar grėsmė, pavyzdžiui problema su paskyrą ar prizo laimėjimo pranešimas, ir pateikiama nuoroda ar telefono numeris. Naudotojai paspaudę nuorodą yra nukreipiami į duomenų viliojimo svetainę arba jiems yra atsiunčiama kenkėjiška mobili programėlė [TCR24]. Tyrimai rodo, kad išmaniųjų telefonų išplitimas paverčia duomenų viliojimą SMS žinutėmis augančia ir didele grėsme, nes yra paprasčiau pavogti duomenis ar įdiegti kenkėjišką programinę įrangą [TCR24].

1.2.6. Duomenų viliojimas balso pranešimais ar skambučiais

Duomenų viliojimas balso pranešimais ar skambučiais yra atakų forma, kai pasitelkiamas garsinis, o ne rašytinis bendravimo būdas. Sukčiai skambučiu ar balso pranešimais apsimeta patikimais asmenimis, pavyzdžiui banko darbuotojais, techninės pagalbos specialistais ar valdžios atstovais ir siekia įtikinti auką atskleisti informaciją ar atlikti veiksmus naudingus užpuolikui [AHN+21]. Dažnai tai yra suklastoti sukčiavimo įspėjimai, kurių metu auka įtikinama patvirtinti asmeninius duomenis arba skambučiai su policijos ar mokesčių grėsmėmis, kurie skatina paniką. Sukčiai klastoja ar maskuoja skambinančio asmens informaciją, kad pateiktų patikimą numerį [AHN+21]. Bendravimas žodžiu apsunkina aukoms galimybę nustatyti apgaulingus skambučius lyginant su elektroniniais laiškais ar SMS žinutėmis, kuriuose galima ilgiau ieškoti įvairių užuominų apie apgaulę. Šis atakos būdas remiasi gyvu pokalbiu, kuriame yra vykdomas spaudimas bei psichologinės manipuliacijos, kurios padeda išvilioti slaptažodžius, vienkartinius PIN ar kreditinių kortelių duomenis [AHN+21].

1.2.7. Duomenų viliojimas naudojantis socialiniais tinklais

Duomenų viliojimas naudojantis socialiniais tinklais — tai atakos, kurios pasinaudoja socialinių tinklų pranešimais duomenims išvilioti. Sukčiai naudojasi įvairiomis platformomis, tokiomis kaip „Facebook“, „Instagram“, „X“/„Twitter“, „Linkedin“ ir kt., kad apsimestų patikimais asmenimis ar organizacijomis. Jie įsilaužia ar suklastoja paskyras, kurių tikslas yra tiekti kenksmingą turinį ar nuorodas, tokiu būdu išviliojant informaciją iš kitų platformos naudotojų [AHN+21]. Dažniausiai pasitaikančios grėsmės paskyrų užgrobimai ir sukčiavimas apsimetant, pavyzdžiui, sukčius gali perimti naudotojo draugo paskyrą ir išsiųsti kenkėjiškas nuorodas jo draugų sąrašui arba

sukurti netikrą klientų aptarnavimo paskyrą, siekdamas apgauti prekės ženklo klientus. Šios atakos dažnai nėra apsaugotos tradicinių elektroninių laiškų filtrų ar organizacijos saugumo įrankių, todėl jas aptikti sudėtingiau [AHN+21]. Tyrimai fiksuoja milijonus bandymų nukreipti naudotojus į netikrus socialinių tinklų puslapius, ypač paplitę netikri „Facebook“ prisijungimo langai [AHN+21]. Platformose naudotojai dažnai pasitiki gautais pranešimais, todėl jos yra labai patrauklios sukčiams.

1.2.8. Duomenų viliojimas kenkėjiškomis programomis

Duomenų viliojimo kenkėjiškomis programomis atakų pirminis tikslas – užkrėsti aukos sistemą kenkėjiška programa, o ne tiesiogiai išvilioti prisijungimo ar kitą informaciją. Sukčiai platina laiškus ar nuorodas, kuriomis pradeda kenkėjiškos programinės įrangos įdiegimą aukos įrenginyje [AHN+21]. Įdiegta programa slaptai renka jautrią informaciją, pavyzdžiui klavišų paspaudimus, saugomus slaptažodžius ar daro ekrano nuotraukas ir perduoda užpuolikui [AHN+21]. Tokiu būdu duomenų viliojimo ataka perauga į hibridinę ataką, kur duomenų viliojimo ataka tampa kenkėjiškų programų pristatymo priemone. Tyrimuose pažymima, kad tai yra reikšminga grėsmė, nulemianti masines prisijungimo duomenų ir finansinės informacijos vagystes [AHN+21].

1.2.9. Peradresavimo duomenų viliojimo atakos

Peradresavimo duomenų viliojimo atakose naudojama technika, kuria naudotojai yra nukreipiami į netikras svetaines manipuliuojant domenų vardų sistema, o ne siunčiant konkretų pranešimą. Užpuolikas pakeičia domenų vardų nustatymą užkrėsdamas naudotojo įrenginį ar puldamas domenų vardų serverį taip nukreipdamas srautą į apgaulingą svetainę [AHN+21]. Naudotojas, įvedęs teisingą nuorodą, pavyzdžiui, banko svetainės adresą, nepastebimai patenka į suklastotą svetainę, kur yra pasisavinami jo prisijungimo duomenys. Šioms atakoms nėra būtina, kad naudotojas paspaustų įtartiną nuorodą, jose išnaudojamos interneto infrastruktūros spragos [AHN+21].

1.2.10. Duomenų viliojimas įterpiant turinį

Duomenų viliojimas įterpiant turinį — tai ataka, kai sukčius įsilaužia į tikrą svetainę ir joje patalpina kenkėjišką turinį ar programinį kodą. Vietoj nukreipimo į netikrą svetainę, užpuolikas išnaudoja svetainės saugumo spragas ir įterpia netikras formas, iššokančius langus ar klaidinančias žinutes tinklapyje [AHN+21]. Pavyzdžiui, pažeistoje svetainėje galima sukurti prisijungimo langą, kuris prašo naudotojo pakartotinai įvesti slaptažodį. Kadangi svetainė yra patikima, naudotojas yra lengvai apgaunamas. Įterptas turinys gali slapta nukreipti naudotoją į kitas svetaines ar rinkti duomenis, todėl aptikti jas sudėtinga.

1.2.11. Duomenų viliojimas tarpininko metodu

Duomenų viliojimas tarpininko metodu — tai ataka, kai užpuolikas įsiterpia tarp naudotojo ir tikros svetainės ar sistemos, perimdamas ir keisdamas jų tarpusavio komunikaciją. [AHN+21]. Asmuo naudojasi tikra svetaine, tačiau visas užklausų srautas keliauja per užpuoliko tarpinį serverį.

Tai leidžia perimti prisijungimo duomenis, slapukus ar kitą informaciją [AHN+21]. Naudotojas paspaudžia nuorodą, kuri jungiasi prie tarpininko serverio, o šis perduoda tikrą puslapį aukai. Įvedus prisijungimo duomenis, užpuolikas gali juos perimti ir valdyti sesiją. Ši ataka įgyvendinama nuorodos ar domenų klastojimu ir peradresavimo nuodijimu tinkluose. [AHN+21]. Jau nuo 2004 m. Ollmann nurodė šias atakas kaip pavojingą grėsmę, galinčią apeiti modernius saugumo modulius, jei naudotojas yra suklaidinamas pasitikėti suklastotu sertifikatu ar domenu [AHN+21].

1.2.12. Duomenų viliojimas paieškų sistemose

Duomenų viliojimas paieškų sistemose vyksta tuomet, kai užpuolikai kuria įvairias kenkėjiškas svetaines, kurios yra dirbtinai rodomos paieškos rezultatų viršuje, pritraukdami naudotojus, ieškančius įvairių produktų ar paslaugų [AHN+21]. Vietoje tiesioginio kontakto elektroniniu paštu ar žinutėmis, sukčiai vilioja aukas, kurios aptinka suklastotas svetaines per paieškos sistemas. Užpuolikai pasitelkia populiarius raktinius žodžius ir apgaulingas paieškos sistemų optimizavimo technikas, kad jų puslapiai atrodytų patikimesni [AHN+21]. Sukuriamos padirbto bankų prisijungimo ar elektroninių parduotuvių svetainės ir jos yra reklamuojamos paieškos sistemose, kad būtų matomos tarp pirmųjų rezultatų. Daugelis naudotojų pasitiki aukšta paieškos pozicija ir nesusimąsto, kad svetainė gali būtų netikra. Atakos išnaudoja paieškos algoritmų vertinimo procesus, kad duomenų viliojimo svetainės atrodytų patikimesnės [AHN+21]. Tai parodo, kad duomenų viliojimas nėra ribojamas tiesioginio bendravimo, jis taip pat gali būti vykdomas per paieškos rezultatų manipuliacijas.

1.2.13. Apibendrinimas

Siekiant sistemiškai įvertinti duomenų viliojimo atakų įvairovę, pateikiama 1 lentelėje, kurioje šie atakų tipai klasifikuojami pagal naudojamas perdavimo priemones ir pagrindinius požymius. Svarbu pažymėti, kad kai kurios atakos pasižymi hibridiniu pobūdžiu – jos tuo pačiu metu apima tiek socialinės inžinerijos, tiek techninės klastos elementus, pavyzdžiui, duomenų viliojimas kenkėjiškomis programomis ar socialiniuose tinkluose. Tokia klasifikacija leidžia geriau suprasti atakų veikimo principus ir galimas apsaugos priemones, kurios turi būti pritaikytos norint apsaugoti nuo skirtingų grėsmių.

1 lentelė. Duomenų viliojimo atakų tipai, naudojamos priemonės ir požymiai

Atakos tipas	**Naudojama priemonė**	**Požymis**
Duomenų viliojimas elektroniniais laiškais	El. paštas	Socialinė inžinerija
Tikslingai nukreiptos duomenų viliojimo atakos	El. paštas	Socialinė inžinerija
Tikslingai nukreiptos atakos į svarbius asmenis ar įstaigas	El. paštas	Socialinė inžinerija

Atakos tipas	Naudojama priemonė	Požymis
Duomenų viliojimas klonavimo pagalba	El. paštas	Techninė klasta
Duomenų viliojimas SMS žinutėmis	SMS žinutės	Socialinė inžinerija
Duomenų viliojimas balso pranešimais ar skambučiais	Skambučiai / balso pranešimai	Socialinė inžinerija
Duomenų viliojimas naudojantis socialiniais tinklais	Socialiniai tinklai	Socialinė inžinerija / Kenkėjiškos programos
Duomenų viliojimas kenkėjiškomis programomis	Kenkėjiškos nuorodos / el. paštas	Techninė klasta
Peradresavimo duomenų viliojimo atakos	DNS / tinklo srautas	Techninė klasta
Duomenų viliojimas įterpiant turinį	Svetainės turinys	Techninė klasta
Duomenų viliojimas tarpininko metodu	Tarpinis serveris / tinklo srautas	Techninė klasta
Duomenų viliojimas paieškų sistemose	Paieškos sistemos	Socialinė inžinerija

1.3. Duomenų viliojimo atakose naudojami apgaulės metodai

Sukčiai naudoja įvairius apgaulės metodus, manipuliuodami naudotojų emocijomis ir priimamais sprendimais. Duomenų viliojimo atakoms elektroniniai laiškai yra sukurti taip, jog atrodytų patikimi ir skubūs, dažnai pasitelkdami žinomų ir organizacijų ar asmenų vardus [KK25]. Vienos iš populiariausių naudojamų taktikų yra:

- Apsimetinėjimas institucijomis ar patikimais asmenimis — apsimetama patikimu subjektu (banku, mokesčių inspekcija ar direktoriumi), siekiant įgyti naudotojo pasitikėjimą. Užpuolikai naudoja oficialius logotipus ir prekės ženklus, kad pranešimai atrodytų autentiški [KK25].
- Baimė ir neatidėliotinumas — skubūs grėsmės ar saugumo perspėjimai (pvz., Jūsų paskyra bus ištrinta šiandien“), kuriais siekiama priversti naudotojus veikti neatidėliotinai ir be jokių patikrinimų. Žinutės keliančios praradimų ar baudų baimę skatina greitai reaguoti kritiškai neįvertinus situacijos [KK25].
- Smalsumas arba godumas — naudotojai viliojami puikiais pasiūlymais, apdovanojimais ar išskirtine informacija. Pavyzdžiui, žadama informacija ar apdovanojimas yra pasiekiamas vos vienu mygtuko paspaudimu [KK25].

Šie psichologiniai veiksniai naudojasi mąstymo šališkumais ir emocinėmis reakcijomis, mažindamos tikimybę, kad naudotojas patikrins turinį [KK25]. Tikslingo duomenų viliojimo

laiškai atrodo itin susiję su gavėju, todėl sukelia dar didesnį pasitikėjimą ir skubumo jausmą, todėl pavykusių atakų rodiklis yra dar didesnis [KK25]. Suprasti šias apgaulingas strategijas yra labai svarbu norint sukurti atsakomąsias priemones, kurios galėtų įspėti ir mokyti naudotojus apie tokio tipo manipuliacijas. Kiekvienas duomenų viliojimo atakos tipas gali pasitelkti skirtingus apgaulės metodus, todėl yra naudinga šiuos ryšius apibendrinti sistemingai. 2 lentelėje pateikiama, kokiais psichologiniais ar socialiniais manipuliacijos principais remiasi skirtingos duomenų viliojimo atakos. Tai leidžia lengviau suprasti, kokie emociniai ar mąstymo aspektai dažniausiai išnaudojami atakų metu.

2 lentelė. Duomenų viliojimo atakų tipai ir jiems būdingi apgaulės metodai

Atakos tipas	**Naudojami apgaulės metodai**
Duomenų viliojimas elektroniniais laiškais	Apsimetinėjimas institucija, baimės ir skubos kūrimas, smalsumo išnaudojimas
Tikslingai nukreiptos duomenų viliojimo atakos	Apsimetinėjimas žinomu asmeniu, laiško asmeninimas, pasitikėjimo kūrimas
Tikslingai nukreiptos atakos į svarbius asmenis ar įstaigas	Suasmenintas bendravimas, oficialių laiškų imitavimas, skubos įspūdis
Duomenų viliojimas klonavimo pagalba	Pasitikėjimo iš ankstesnio bendravimo išnaudojimas, siuntėjo imitavimas
Duomenų viliojimas SMS žinutėmis	Skubos kūrimas, baimės sukėlimas, suklastoti siuntėjai
Duomenų viliojimas balso pranešimais ar skambučiais	Spaudimas, apsimetinėjimas patikimais asmenimis, panikos sukėlimas
Duomenų viliojimas naudojantis socialiniais tinklais	Apsimetinėjimas draugais ar įmonėmis, smalsumo ir patiklumo išnaudojimas
Duomenų viliojimas kenkėjiškomis programomis	Smalsumo ir patiklumo išnaudojimas
Peradresavimo duomenų viliojimo atakos	Pasitikėjimo technologijomis išnaudojimas
Duomenų viliojimas įterpiant turinį	Patikimos svetainės naudojimas
Duomenų viliojimas tarpininko metodu	Apsimetinėjimas tarpininkaujama sistema
Duomenų viliojimas paieškų sistemose	Smalsumo išnaudojimas, pasiūlymai ar nuolaidos

1.4. Naudotojo pažeidžiamumui įtaką darantys veiksniai

Nevisi naudotojai yra vienodai pažeidžiami duomenų viliojimo atakų. Keletas žmogiškųjų veiksnių daro įtaką tikimybei tapti šios atakos auka. Svarbus vaidmuo tenka kontekstui — jeigu sukčiavimo elektroninio laiško tema atitinka naudotojo vykdomus veiksmus ar lūkesčius, yra didesnė tikimybė, kad naudotojas bus apgautas. Atliktas praktinis tyrimas parodė, kad žmonės kurie paspaudė ir tie, kurie nepaspaudė ant kenksmingos nuorodos atkreipia dėmesį į skirtingus ženklus bei tas pačias užuominas interpretuoja skirtingai priklausomai nuo jų darbo konteksto

[GST+18]. Linkę spausti tokias nuorodas naudotojai dažnai baiminasi praleisti ką nors svarbaus (pvz., nepateikti atsakymo laiku), o atsargesni naudotojai labiau rūpinasi galimomis grėsmėmis, kylančiomis paspaudus tokio tipo nuorodas [GST+18].

Demografiniai ir asmenybės veiksniai taip pat buvo tiriami. Kai kurie tyrimai parodė amžiaus skirtumus, pavyzdžiui, jauni suaugusieji (18–25 metų) yra labiau pažeidžiami duomenų viliojimo atakų nei vyresni asmenys [DFM+22]. Jaunesni naudotojai dažnai geriau išmano technologijas, tačiau yra pernelyg pasitikintys savimi, o vyresnieji dažnai linkę būti atsargesni. Visgi senyvo amžiaus asmenys, turintys menką kompiuterinį raštingumą, taip pat dažnai tampa atakų taikiniais [DFM+22]. Be to, lytis ir išsilavinimas taip pat gali turėti įtakos — tyrimai rodo, kad moterys dažniau būna atsargesnės nei vyrai, o didesnį informuotumą apie kibernetinį saugumą turintys naudotojai geriau atpažįsta grėsmes [KK25]. Asmenybės bruožai, tokie kaip impulsyvumas ir polinkis pasitikėti bei tuometinė emocinė būsena (stresas, išsiblaškymas) gali turėti poveikį naudotojo pastabumui ir kritiškam elektroninių laiškų vertinimui [KK25]. Naudotojai, turintys greitai priimti sprendimus ar patiriantys didelį kognityvinį krūvį, gali ignoruoti perspėjimus, todėl tampa labiau pažeidžiami. Žmonių pažeidžiamumas sukčiavimui elektroniniu paštu yra įvairiapusis – jį lemia psichologiniai veiksniai, situacijos kontekstas ir individualios naudotojų savybės [KK25]. Šios įžvalgos pabrėžia, kad būtina taikyti tikslines, į naudotojus orientuotas apsaugos priemones, kurios atsižvelgia į konkrečius pažeidžiamumus. Žemiau pateikta konceptualioji 3 lentelė rodo kokie pažeidžiamumai dažniausiai pasireiškia skirtingose naudotojų grupėse ir kokie veiksniai lemia jų atsparumą ar jautrumą atakoms.

3 lentelė. Naudotojų grupės ir joms būdingi pažeidžiamumai duomenų viliojimo atakoms

Veiksnys	**Rizikos grupė**	**Galimas pažeidžiamumas**
Amžius	Jauni suaugusieji (18–25 m.)	Perteklinis pasitikėjimas, impulsyvumas
	Senyvo amžiaus naudotojai	Menkas IT raštingumas, nesupratimas apie grėsmes
Technologinis raštingumas	Mažai patirties turintys naudotojai	Negebėjimas atpažinti sukčiavimo požymių
Emocinė būsena / stresas	Stresą ar spaudimą patiriantys naudotojai	Sumažėjęs kritinis vertinimas, didesnis impulsyvumas
Asmenybės bruožai	Impulsyvūs, linkę pasitikėti naudotojai	Nepatikrina informacijos, reaguoja spontaniškai
Lytis	Vyrai	Kai kurie tyrimai rodo mažesnį atsargumą
Darbo kontekstas	Naudotojai spaudžiami terminų, atsakomybės	Veikia skubotai, bijo praleisti svarbią informaciją
Informuotumas apie kibernetinį saugumą	Žemo informuotumo naudotojai	Nepastebi apgaulės požymių, nežino geriausiųjų praktikų

1.5. Naudotojų mokymas ir sąmoningumo apie sukčiavimą ugdymas

Atsižvelgiant į tai, kad duomenų viliojimo atakos yra orientuotos į žmones, pagrindinė priemonė kovai prieš jas yra naudotojų švietimas – siekiama juos išmokyti atpažinti ir išvengti sukčiavimo bandymų. Saugumo mokymai tapo pagrindine apsaugos nuo jų priemone įvairiose organizacijose.

1.5.1. Simuliaciniai mokymai

Naudotojams periodiškai siunčiami netikri duomenų viliojimo elektroniniai laiškai. Jei naudotojas paspaudžia ant nuorodos, jam yra pateikiama mokomoji žinutė. Vienas iš pavyzdžių – PhishGuru, sistema, kuri po klaidingo paspaudimo pateikia trumpą žinutę arba komiksą [DFM+22]. Tyrimai rodo, kad šis metodas veiksmingas — naudotojai, kurie mokėsi PhishGuru pagalba, vėliau žymiai rečiau spaudė nuorodas duomenų viliojimo atakų laiškuose [DFM+22]. Jaunesniems naudotojams, kurie rečiau susiduria su tokio tipo atakomis, skirtumas tarp pradinių ir galutinių rezultatų buvo itin žymus [DFM+22].

1.5.2. Sužaidybintas mokymasis

Duomenų viliojimo atakų apsaugos švietimo pavertimas žaidimu yra veiksmingas gerinant įsitraukimą ir žinių įsisavinimą. Vienas iš tokių žaidimų yra „Anti-Phishing Phil“, sukurtas Sheng ir kt. (2007), kuriame dalyviai klausimų-atsakymų formatu mokomi atpažinti įtartinas nuorodas [DFM+22]. Žaidusieji šį žaidimą rodė geresnį gebėjimą atskirti duomenų viliojimo svetaines nei nežaidusieji, o tai rodo, kad interaktyvūs žaidimai gali sustiprinti saugumo požymių įsisavinimą [DFM+22]. Taip pat yra ir vaidmenų žaidimų, tokių kaip „What.Hack“, kurie leidžia naudotojams suvaidinti scenarijus, kuriuose jie ginasi nuo duomenų viliojimo atakų. Vienas tyrimas parodė, kad šis vaidybinis žaidimas buvo veiksmingesnis saugantis nuo duomenų viliojimo atakų ir labiau įtraukiantis nei tradiciniai mokymai skaidrių ar dokumentų forma [DFM+22]. Mokymų sužaidybinimas kelia motyvaciją ir gali būti pritaikytas skirtingoms auditorijoms, pavyzdžiui, mokiniams ar studentams, įmonių darbuotojams, senjorams.

1.5.3. Mokymai tinkamu momentu *(angl. Just-in-Time)*

Įvairios priemonės mokomąjį turinį integruoja tiesiai į saugumo pranešimus ar elektroninius laiškus, kad naudotojai gautų naudingą informaciją sprendimo priėmimo metu. Viename tyrime perspėjimo puslapiuose buvo pridėti aiškinamieji pranešimai, kurie parodė, kodėl svetainė gali būti kenksminga. Šis būdas davė teigiamų trumpalaikių ir ilgalaikių rezultatų naudotojų gebėjime atpažinti duomenų viliojimo svetaines [DFM+22]. Suderinant įspėjimus ir aiškinamąjį dizainą ar klausimus, naudotojai yra apsaugomi nuorodos spaudimo metu ir taip pat yra išmokomi kur reikėtų atkreipti dėmesį ateityje. Kiti tyrimai lygino pavyzdžiais ir istorijomis bei faktais ir patarimais pagrįstos informacijos apie saugą teikimo būdus ir nustatė, kad mokymosi veiksmingumui yra itin svarbu forma, kuria informacija yra pateikiama bei duomenų šaltinio patikimumas [DFM+22]. Įtraukiantis turinys, kartu su įsakiomis žinutėmis gali pagerinti naudotojų įsisavinimą pamokoms

apie duomenų viliojimo atakas. Tinkamas naudotojų mokymas gali reikšmingai sumažinti ir jų pažeidžiamumą šioms atakoms [DFM+22]. Tačiau vis tiek lieka iššūkiai, tokie kaip mokymų veiksmingumas ilgalaikėje perspektyvoje, retais ar ne tam kontekstui pritaikytais mokymais pasitikintys naudotojai daro klaidų kitokioje aplinkoje. Todėl susidomėjimas mokymų derinimu su realaus laiko pagalba vis didėja.

1.5.4. Apibendrinimas

Naudotojų mokymas yra esminė priemonė siekiant sumažinti duomenų viliojimo atakų sėkmę, nes šios atakos remiasi žmogaus sprendimų silpnybėmis, o ne vien techniniais pažeidžiamumais. Apžvelgti metodai — simuliaciniai mokymai, sužaidybinimas ir mokymas tinkamu momentu — parodė, kad į naudotoją orientuotas švietimas gali reikšmingai pagerinti gebėjimą atpažinti kibernetines grėsmes.

Tačiau išryškėja ir apribojimai — ilgalaikis šių mokymų poveikis dažnai silpsta, o veiksmingumas priklauso nuo turinio aktualumo, pateikimo būdo ir naudotojų įsitraukimo. Todėl svarbu, kad mokymas būtų neatsiejama kasdienės naudotojo sąsajos dalis.

Kuriami įrankiai kovai su duomenų viliojimo atakomis turėtų integruoti naudotojų mokymą į naudotojo sąsajas, kad kiekvienas perspėjimas ne tik saugotų, bet ir ugdytų gebėjimą savarankiškai atpažinti grėsmes. Šiuo būdu siekiama kurti ne tik pasyvią, bet ir mokomąją apsaugą.

1.6. Įrankiai ir paslaugos, skirtos apsaugai nuo duomenų viliojimo atakų

Be naudotojų mokymų, yra ir įvairių techninių įrankių ir paslaugų, skirtų apsaugoti naudotojus nuo duomenų viliojimo atakų. Dabartinės naršyklės ir įvairūs elektroninio pašto paslaugų teikėjai integruoja apsaugas nuo šio tipo atakų. Vienas labiausiai naudojamų mechanizmų yra blokavimo sąrašas *(angl. blocklists)*, dar kitaip vadinamas juoduoju sąrašu *(angl. blacklists)*, kuriais pateikiamos žinomos duomenų viliojimo svetainės ir adresatai. „Google Safe Browsing“, kurią naudoja „Chrome“, „Firefox“ ir taip pat keletas kitų naršyklių, nuolatos palaiko kenksmingų svetainių sąrašą. Ši sistema automatiškai įspėja ar net blokuoja naudotojo veiksmus, kai šis bando atidaryti duomenų viliojimui skirtos svetainės nuorodą. Tokios blokavimo sąrašais grįstos paslaugos yra gan veiksmingos kovojant prieš jau žinomas atakas, pasižymi aukštu aptikimo procentu bei padeda sustabdyti daugelio tokių atakų plitimą dar prieš didžiajai daliai naudotojų susiduriant su jomis. [RTN25].

Kitos viešai prieinamos paslaugos, „PhishTank“ ir „OpenPhish“, renka duomenis apie patvirtintas duomenų viliojimo atakų svetaines ir teikia sistemų integracijoms tinkamą prieigą prie šių duomenų įvairiems saugumo moduliams kurti [RTN25]. Nemaža dalis šiuolaikinių elektroninių paštų tiekėjų filtrų taip pat naudojasi šiomis ar panašiomis duomenų bazėmis, kad galėtų karantinuoti duomenų viliojimo atakų laiškus jiems dar nepasiekus naudotojo pašto dėžutės.

Tačiau tokie blokavimo sąrašais pagrįsti įrankiai turi reikšmingų apribojimų. Jie skirti apsaugoti tik nuo tokių duomenų viliojimo svetainių, kurios jau yra nustatytos ir patikrintos. Naujausi ar labai tikslinei auditorijai skirti tinklapiai dažnai gali būti neįtraukti į šias duomenų bazes. Įgudę sukčiai gali greitai generuoti naujas nuorodas ar naudoti įvairias technikas kaip

adresų trumpinimas ar peradresavimas, kad kuriam laikui išvengtų įtraukimo į šias duomenų bazes [BBK+25]. Tyrimai parodė, kad duomenų viliojimo svetainės gali veikti ne tik kelias valandas, bet ir dienas prieš patenkant į blokavimo sąrašus, todėl naudotojai kurį laiką yra neapsaugoti nuo jų [RTN25].

Taip pat iškyla kita problema — įspėjimai dažnai yra pernelyg subendrinti, ribotai paaiškinti. Įprastai naršyklės įspėjimas tik nurodo, kad svetainė yra įtariama kaip duomenų viliojimo atakos dalis ir rekomenduoja jos neatidaryti [RTN25]. Tai sustabdo dali naudotojų, tačiau nemaža dalis vis tiek ignoruoja pranešimą, ypatingai jeigu perspėjimo priežastis nėra aiški arba įspėjimai rodomi nuolatos, sukeldami „įspėjimų nuovargį“ *(angl. warning fatigue)* [RTN25]. Tyrimai rodo, kad kai naudotojams trūksta konteksto ar priežasties, jie gali spausti nuorodas vedini paprasčiausio smalsumo ar per didelio pasitikėjimo svetaine [RTN25]. Taip pat naudotojai gali per daug pasitikėti ir automatizuotais įrankiais, todėl jei svetainė ar elektroninis laiškas nėra blokuojamas, jie gali pamanyti, kad tai yra saugu, neturėdami jokių įgūdžių atpažinti duomenų viliojimo atakų požymius [RTN25].

Be interneto naršyklių įspėjimų, yra ir kovai prieš duomenų viliojimo atakas skirti įskiepiai bei įrankių juostos. Yra kuriami tiek komerciniai įrankiai, tiek tyrimų prototipai. Ankstyvieji šių atakų tyrimai, pavyzdžiui, Wu ir kt. (2006) parodė, kad didelė dalis saugumo įrankių juostų nebuvo veiksmingos, nes naudotojai tiesiog nepastebėjo jų rodomų ženklų arba jais tiesiog nepasitikėjo [WMG06]. Tai tik parodo, kad piktogramų ar įvairių ženklų pridėjimas nėra pakankamas, jeigu jie nėra lengvai suprantami ir aiškūs naudotojui.

Šiuometiniai įrankiai suteikia būtiną automatinės apsaugos sluoksnį, filtruodami pavojingą turinį ir įspėdami naudotojus. Tačiau dažnai tokie įrankiai neapsaugo nuo naujų ar tikslinių atakų, taip pat neugdo naudotojo gebėjimų savarankiškai identifikuoti nuo grėsmių.

1.7. Naudotojo sąsajos dizainas apsaugai nuo duomenų viliojimo atakų

Perspektyvi kryptis, stiprinanti naudotojų apsaugą nuo duomenų viliojimo atakų yra saugumo įspėjimų ir įvairių su jais susijusių rodiklių naudotojo sąsajoje tobulinimas. Jų tikslas yra kurti naudotojo sąsajas, kurios ne tik saugo naudotoją nuo duomenų viliojimo atakų, bet taip pat ir moko jį atpažinti tokias grėsmes. Naujausiuose tyrimuose tyrimuose išryškėja dvi pagrindinės kryptys — aiškinamieji pranešimai bei atakoms atsparaus naudotojo sąsajos dizaino kūrimas.

1.7.1. Aiškinamieji pranešimai ir intuityvus dizainas

Tradiciniai saugumo pranešimai įprastai įspėja naudotoją apie kylantį pavojų, bet retai kada suteikia detalią informaciją apie grėsmę. Naujausi žmogaus ir kompiuterio sąveikos tyrimai rodo, kad saugumo pranešimai turėtų būti aiškinamieji, tokiu būdu naudotojai suprastų, kodėl šis turinys yra laikomas kenksmingu [RTN25]. Pateikiant aiškų pranešimo pagrindimą ar paryškinant įtartinus elementus, naudotojo sąsaja gali padėti suformuoti ilgalaikius prisiminimus, kurie vėliau jiems padėtų išvengti panašaus tipo grėsmių [RTN25].

Desolda ir kt. (2022) pristatė įspėjamųjų pranešimų langų prototipus, kuriais aiškiai nurodomi duomenų viliojimo požymiai svetainėje, tokius kaip klaidinančios nuorodos ir suklastoti logotipai,

o ne tiesiog pateikia bendrinį pranešimą apie grėsmę keliančią svetainę [DAA+22]. Tyrime, kuriame dalyvavo 150 asmenų, šie aiškinamieji pranešimai buvo lyginami su standartiniais „Chrome“ ir „Firefox“ pranešimais. Rezultatai parodė, kad naudotojams aiškinamieji pranešimai buvo labiau suprantami ir padėjo sumažinti paspaudimų ant duomenų viliojimo nuorodų skaičių [DAA+22]. Dalyviai, pamatę juo geriau suvokė grėsmingų svetainių požymius ir rečiau ignoravo tokio tipo pranešimus [DAA+22; RTN25]. Verta paminėti, kad naudotojai, turintys žemesnį kibernetinio saugumo žinių kiekį jautė naudą. Jie įgijo pasitikėjimo ir vėliau geriau gebėjo atpažinti duomenų viliojimo atakas, net nematydami šių pranešimų [DAA+22].

Kitas tyrimas, atliktas Aneke ir kt. (2020) siūlo pranešimus apie duomenų viliojimo atakas, kurie įterpia trumpas mokomąsias žinutes ar patarimus į pranešimą. [AAD20]. Pagrindinė mintis — aptikus sukčiavimo bandymą, pranešimas ne tik blokuoja patekimą į svetainę, tačiau ir tekstu pamoko naudotoją. Tokio įspėjimo pavyzdys galėtų būti: *„Ši svetainė yra įtartina, nes jos adresas yra login-facebook.com – tai nėra oficialus „Facebook“ domenas.“*

Tokiu būdu sistema pateikia trumpą pamoką tinkamu momentu, atitinkančią anksčiau aptartą mokymo būdą. Pirminiai tyrimai parodė, kad šis metodas gerina naudotojų ilgalaikį atsparumą atakoms. Taip pat matoma, kad naudotojai labiau pasitiki pateikiamais įspėjamaisiais pranešimais, kai supranta jų pagrindimą, priešingai nei ryškiai raudoni įspėjimai be jokio paaiškinimo [DAA+22].

Vis dėlto, iškyla iššūkių aiškinamuosiuose dizainuose — didelis informacijos kiekis gali perkrauti ar suklaidinti naudotoją, tačiau mažas jos kiekis neteikia didelės ilgalaikės naudos. Verta pabrėžti, kad svarbu kruopščiai atrinkti aiškinamus elementus ir kaip tai daryti, atsižvelgiant į naudotojų dėmesio sutelkimo trukmę bei susidomėjimą [DAA+22].

1.7.2. Atsparaus sukčiavimui naudotojo sąsajos dizaino modeliai

Kartu su įspėjamųjų pranešimų patobulinimais tyrėjai taip pat nagrinėja ir įvairius naudotojo sąsajos dizainus, kurie gali padėti naudotojams lengviau pastebėti ir taip išvengti duomenų viliojimo bandymų. Šie metodai skirti paslėptų požymių atskleidimui naudotojams.

Vienas iš būdų — tikrųjų domenų paryškinimas svetainės adreso juostoje, pavyzdžiui, pateikiant jį kitu šriftu ar spalva, kad naudotojas atkreiptų dėmesį ar galėtų įsitikinti, kad yra tikroje svetainėje [KOO22] Jeigu naudotojas žino tikrąjį svetainės domeną, pavyzdžiui, *facebook.com*, jis gali lengvai pastebėti, kad nuoroda atrodo įtartina, jeigu joje mato *facebook.com.securelogin.ru* , kai pagrindinis domenas (*facebook.com* ar *securelogin.ru*) yra išskirtas.

Praktikoje matoma, kad domeno paryškinimas dalinai padeda, tačiau nėra garantuotas problemos sprendimo būdas. Viename tyrime buvo nustatyta, kad naudotojai, kurie dažnai tikrina svetainės adreso juostą, geriau atpažino netikras svetaines, tačiau vien tik paryškintas domenas nežymiai sumažino bendrą pažeidžiamumo lygį [XPY+17]. Daugumai dalyvių vis tiek nepavyko atpažinti duomenų viliojimo svetainių, net kai domenas buvo paryškintas, kas rodo, kad jie nesuprato, kur turėtų atkreipti dėmesį ar nesupranta domeno svarbos. Tai parodo, kad domeno paryškinimas negali būti vienintelis įrankis siekiant užkirsti kelią duomenų viliojimo atakoms [XPY+17].

Kitas naudotojo sąsajos sprendimas bando įterpti sukčiavimo požymius į ekrane naudotojo

matomą turinį. Vienas tokių — skaidrios eilutės *(angl. transparent-string)* metodas, kurį pristatė Vossing ir kt. (2022) [VKL+22]. Jis įterpia pilną nuorodą tekstu šalia įprasto hipersaito elektroniniuose laiškuose. Tokiu būdu naudotojui nereikia užvesti pelės ir žiūrėti į būsenos juostą, kad pamatytų kur bus peradresuojama. Sistema modifikuoja laiško turinį, kad šalia nuorodos, pavyzdžiui „Prisijungti prie Facebook“ būtų rodomas tekstas [facebook.com/login] [BBK+25]. Taip, net jeigu nuorodoje ir rašoma, kad reikėtų ją paspausti norint prisijungti, naudotojas iš karto mato, kur iš tikrųjų bus nuvestas šiuo atveju. Pamatę įtartiną, nesusijusį su siekiamu rezultatu domeną ar nuorodą, naudotojas gali sustoti ir netapti atakos auka [BBK+25]. Dėl šios užuominos buvimo matomoje vietoje ir tame pačiame kontekste, nereikia papildomų naudotojų veiksmų pastebėti apgaulingoms nuorodoms, kuriose naudojami įvairūs subdomenai ar rašybos klaidos. Tyrime, kuriame naudotojai buvo suskirstyti į dvi grupes, skaidrios eilutės metodas naudotojo sąsajoje padėjo naudotojams teisingai identifikuoti 79 % duomenų viliojimo atakų atvejus, lyginant su 60 % kontrolinėje grupėje, kuri matė įprastus elektroninius laiškus. Tyrimas parodė, kad net ir pakankamai nedidelis naudotojo sąsajos pakeitimas gali gan ženkliai padidinti šių atakų atpažinimo tikimybę [BBK+25].

Kitas panašus modelis yra įspėjimų ar įvairių rodiklių pateikimas kiek įmanoma arčiau kenksmingo elemento. Petelka ir kt. (2019) tyrė įterptinius duomenų viliojimo atakų įspėjimus elektroninių laiškų sistemose, kur vietoje bendrinės pranešimo juostos laiško viršuje jis buvo pateiktas šalia įtartinos nuorodos arba tuo metu, kai naudotojas užveda ant jos [PZS19]. Tyrėjai taip pat bandė priverstinio dėmesio dizainus, kurie neleidžia paspausti ant nuorodos, kol naudotojas nepažymi, kad pasitiki šiuo laišku ar nuoroda, taip priversdami naudotoją atkreipti dėmesį į įspėjimą [PZS19].

Tyrime nustatyta, kad tikslingai į nuorodas nukreipti įspėjamieji pranešimai žymiai sumažino paspaudimų ant duomenų viliojimo nuorodų skaičių, lyginant su įprastais įspėjamaisiais pranešimais laiško viršuje. Veiksmingiausias dizainas vertė naudotojus atlikti papildomus veiksmus su įspėjimo pranešimu ir peržiūrėti tikrąjį nuorodos adresą prieš leidžiant vykdyti tolimesnius veiksmus. Tai rodo, kad saugumo užuominų įterpimas į naudotojų veiksmų seką bei jų pateikimas tikslingoje vietoje yra pranašesnis nei antraplaniai pranešimai [PZS19].

Kiti naudotojo sąsajos sprendimai naudoja vaizdinius rizikos požymius, tokius kaip piktogramos ar spalvos šalia nuorodų ar siuntėjų vardų, standartizuotus išdėstymus patikimiems pranešimams bei įspėjimus apie išorinius elektroninius laiškus. Kai kurios pašto sistemos skirtos verslui pažymi laiškus antrašte „Išorinis siuntėjas“, skatindamos darbuotojus būti labiau kritiškiems tokio tipo laiškams. Tyrimai šioje srityje pabrėžia, kad duomenų viliojimo atakų požymiai turi būti matomi, įtraukti į esamą kontekstą ir lengvai suprantami naudotojui.

1.7.3. Integraciniai metodai ir nauji sprendimai

Literatūra rodo, kad gera apsaugos nuo duomenų viliojimo atakų priemonė turėtų suderinti automatizuotą grėsmių aptikimą su naudotojo mokymu ir lengvai matomais pranešimais. Pastarųjų metų pažanga yra nukreipta būtent šia kryptimi. Vienas iš pavyzdžių yra PhishXplain (PXP) sistema, kuri papildo naršyklės duomenų viliojimo aptikimo metodus dirbtinio intelekto generuojamais

paaiškinimais [RTN25]. Aptikus kenksmingą svetainę, PXP ją ne tik blokuoja bet ir parodo įtartinus elementus puslapyje, pateikia trumpus paaiškinimus, kodėl šie elementai rodo galimą duomenų viliojimo ataką, sugeneruotus dirbtinio intelekto kalbos modelių pagalba [RTN25].

Atlikus mėnesio trukmės testavimą su daugiau nei 7000 svetainių, sistemai pavyko pateikti tinkamus aiškinamuosius pranešimus didžiajai daliai atakų požymių. Taip pat tyrime aiškinamuosius įspėjimus gavę per PXP sistemą naudotojai žymiai geriau identifikavo duomenų viliojimo svetaines net ir tada, kai įspėjimai buvo neberodomi [RTN25]. Tai rodo, kad aiškinamasis metodas susietas su dirbtinio intelekto generuojamais pranešimais padėjo įgyti ilgalaikių žinių ir įgūdžių, o tai ir yra į naudotoją orientuoto saugumo dizaino tikslas. Taip pat naudotojai rodė didesnį pasitikėjimą tokio įrankio sukurtais pranešimais, o tai svarbu įrankio naudojimo ilgalaikiškumui. Jeigu naudotojai pasitiki saugumo sistema, jie labiau linkę atsižvelgti į jos įspėjimus [RTN25].

1.7.4. Apibendrinimas

Apibendrinant, galima išskirti tris pagrindines kryptis, kaip naudotojo sąsajos dizainas gali padėti apsisaugoti nuo duomenų viliojimo atakų: aiškinamieji pranešimai, atakoms atsparus dizainas ir integraciniai sprendimai.

Pirmiausia, aiškinamieji pranešimai pasižymi tuo, kad jie ne tik įspėja apie grėsmę, bet ir paaiškina, kodėl konkretus turinys yra įtartinas. Tyrimai rodo, kad tokie pranešimai padeda geriau suprasti grėsmes, sustiprina naudotojų gebėjimą kritiškai vertinti informaciją ir suformuoja ilgalaikius saugumo įgūdžius. Vis dėlto, jų sėkmė priklauso nuo informacijos kiekio ir pateikimo – pernelyg išsamūs pranešimai gali apkrauti naudotoją, o per trumpi – likti neaiškūs.

Antra, įvairūs dizaino sprendimai, pavyzdžiui, domeno paryškinimas, papildomos informacijos apie nuorodas rodymas (skaidrios eilutės), ar įspėjimų pateikimas šalia įtartino elemento, padeda atskleisti paslėptus sukčiavimo požymius. Šie sprendimai veikia tuomet, kai naudotojui informacija pateikiama tiksliai ten, kur ji reikalinga, taip sumažinant pastangų kiekį informacijai įvertinti. Vis dėlto, nė viena priemonė nėra savaime pakankama – veiksmingiausi yra jų deriniai.

Trečia, integraciniai sprendimai, tokie kaip „PhishXplain“, apjungia automatizuotą grėsmių aptikimą su dirbtinio intelekto generuojamais paaiškinimais. Tai leidžia ne tik blokuoti pavojingas svetaines, bet ir išmokyti naudotojus, kaip tokias grėsmes atpažinti savarankiškai. Tyrimai rodo, kad tokie sprendimai padeda įgyti ilgalaikių įgūdžių ir stiprina pasitikėjimą saugumo sistema – o tai itin svarbu, kad naudotojai neignoruotų įspėjimų.

1.8. Elgesio paskatos ir įtikinamasis projektavimas

Mokslininkai pritaikė elgesio ekonomikos koncepcijas kibernetiniam saugumui, sukurdami paskatas *(angl. nudges)*, kurios naudotojus skatintų rinktis saugesnes alternatyvas, nesuvaržant jų laisvės. Naršyklės ar elektroninio pašto sistemų plėtiniai gali pažymėti svarbiausias laiško dalis, pavyzdžiui, siuntėjo adresą ar nuorodas, arba rodyti patarimus realiuoju laiku. Tokie būdai rodo didesnį duomenų viliojimo atpažinimo rodiklį [AA24]. Šios paskatos remiasi elgesio ekonomikos ir psichologijos principais, tokiais kaip numatytieji pasirinkimai, ryškios ir aiškios užuominos, raginimai tinkamu metu, ir nukreipia naudotojo elgseną saugia linkme. Modeliai, tokie kaip

MINDSPACE, pabrėžia pranešėjo įtaką, numatytuosius nustatymus ir išskirtinumą [CBJ+14]. Taikant šiuos principus paskata naudotojui gali iš karto parinkti saugų variantą arba išryškinti įspėjimus, taip formuodama aplinką norimiems rezultatams, tačiau paliekant naudotojui teisę rinktis [CBJ+14]. Nustatyta, kad gerai suprojektuotos paskatos, tokios kaip iškylantys langai ar užuominos sprendimo metu, verčia naudotoją sąmoningiau vertinti riziką, neverčiant jo tapti saugumo ekspertais.

Šiuolaikinės apsaugos priemonės nuo duomenų viliojimo atakų įspėjamuosius pranešimus suvokia ne kaip nekintančius tekstus, o kaip galimybę įtikinti. Wang ir kt. (2022) sukūrė 14 skirtingų įspėjimų, pagrįstų įvairiomis psichologiniais įtikinimo būdais, ir patikrino juos žvilgsnio sekimo bei A/B bandymais [WSW+22]. Rezultatai atskleidė, kad įspėjimai, pasitelkę autoritetingumą, socialinį įrodymą ar provokuojantį klausimą, reikšmingai pralenkė bendrinius pranešimus. Piktogramos su aiškiais simboliais taip pat patraukė dėmesį labiau nei standartinės „Alipay“ platformoje tokie įtikinamieji įspėjimai apgaulės sėkmę sumažino 33,2 % [WSW+22]. Įrodyta, kad traktuojant įspėjimą kaip komunikacijos formą ir pasitelkiant subtilų įtikinimą, pavyzdžiui autoritetingą toną ar kitų naudotojo elgesio pavyzdžius, jo veiksmingumas ženkliai išauga [WSW+22].

Taip pat yra matoma tendencija integruoti elgesio ekonomikos idėjas į saugų projektavimą. Skatinimas rinktis (Thalerio ir Sunsteino pasiūlytas būdas nukreipti pasirinkimus), rėminimo efektai ir numatytasis šališkumas yra taikomos projektuojant naudotojo sąsajas, skatinančias saugius pasirinkimus [CBJ+14]. Pavyzdžiui, saugaus varianto nustatymas kaip numatytojo gali ženkliai padidinti jo pasirinkimą [VCM18]. Saugumo sąsajose taip pat tyrinėjama įspėjimų pateikimas kaip galimų nuostolių patyrimas, išnaudojant naudotojų nenorą juos patirti, o suasmenintos paskatos pritaiko patarimus pagal individualų naudotojo sprendimo stilių, didindamos pasitikėjimą įrankiu ir skatina laikytis saugumo reikalavimų, palyginus su vienodais, visiems skirtais įspėjimais. Bendrai, naudojantis elgsenos ekonomikos ir įtikinamojo projektavimo priemonėmis, naujausios apsaugos priemonės nuo duomenų viliojimo atakų bando subtiliai „pastumti“ naudotojus saugaus elgesio link taip, kad pastarieji jaustųsi gerai. Toks į žmogų orientuotas požiūris pripažįsta naudotojų polinkį į šališkumus ir impulsyvius veiksmus, tačiau jais pasinaudoja saugumo tikslais, pavyzdžiui naudojantis socialinėmis normomis („90 % žmonių nespaudžia šios nuorodos“) arba teikiant grįžtamąjį ryšį po rizikingos nuorodos paspaudimo, kad sustiprintų mokymąsi [KSB+24; SZN+21].

1.9. Sukčių taktikos ir psichologija

1.9.1. Įtikinamų pranešimų kūrimas

Duomenų viliojimo atakų autoriai meistriškai išnaudoja žmogaus psichologiją. Analizuojant beveik 200 tokių atakų elektroninių laiškų temų, labiausiai paplitę įtikinimo metodai yra autoritetingumas, stiprios emocijos, tokios kaip baimė ir jaudulys, pasitikėjimo ir sąžiningumo pranešimai [FT18]. Pavyzdžiui, apsimetama banko darbuotoju ar IT pagalba, pasinaudojant paklusnumu, o grėsmės, tokios kaip galimas paskyros uždarymas ar baudos, kelia skubumo pojūtį ir trukdo kritiškai apsvarstyti situaciją. Artėjančių galutinų terminų ir vienkartinių galimybių pranešimai užblokuoja naudotojo galimybę įvertinti laiško turinį [Nai15]. Socialiniai kabliukai yra taip pat

labai paplitę. Eksperimente, kuriame dalyviai patys kūrė laiškus duomenų viliojimo atakoms, bendro pomėgio ar pažįstamo asmens paminėjimas reikšmingai padidino tokių atakų sėkmę [RG18]. Sukčiai taip pat naudoja nuolankumo ir socialinio spaudimo metodus bei mainų principą — sakoma, kad didelė dalis kolegų jau užpildė duomenis ar nurodoma, kad bus grąžinta mokesčių permoka ir prašoma skubiai įvesti duomenis, kad būtų galima ją pervesti [AA24].

1.9.2. Socialinės inžinerijos principų taikymas

Šiuolaikiniai sukčiavimai ir duomenų viliojimo atakos remiasi gerai žinomais socialinės inžinerijos principais, iš dalies atitinkančiais Cialdini šešis įtakos principus — autoritetingumas, mainai, ribotumas, panašumas, socialinis patvirtinimas ir nuoseklumas. Neseniai publikuotuose tyrimuose pateikiama konkrečių įrodymų apie šių principų taikymą [AA24]. Pavyzdžiui, 2024 m. apžvalgoje nurodoma, kad sukčiai dažnai išnaudoja autoritetingumą, panašumą, ribotumą, socialinį patvirtinimą ir mainus, kad apgaulės būdu įgytų aukų pasitikėjimą. Autoritetingumas dažnai pasireiškia apsimetimu aukšto rango pareigūnu ar patikima įmone, o socialinis patvirtinimas rodomas frazėmis, kurios nurodo, jog visi darbuotojai ar draugai jau atliko šį veiksmą. Ribotumo ir skubos elementas pasiūlymuose ar grėsmėse skatina paniką, nors viename iš tyrimų matoma, kad naudotojai nurodo ribotumą kaip mažiausiai patikimą, nes sukelta dirbtina skuba kelia įtarimų [AA24]. Mainų principas matomas laiškuose, žadančiuose atlygį ar paslaugą, siekiant suveikti instinktyviam norui duoti kažką atgal, pavyzdžiui, prisijungimo duomenis ar finansinę informaciją [FT18]. Panašumo taktikos naudojimas pastebimai auga — sukčiai išnaudoja asmens duomenis, neretai paimtus iš socialinių tinklų, kad žinutės atrodytų labiau suasmenintos ir draugiškos arba yra perimamas aukos draugo paskyros valdymas. Pastarasis metodas yra ypatingai veiksmingas, socialinių tinklų duomenų viliojimo atakų atveju kenkėjiška nuoroda, atsiųsta iš pažįstamo asmens, ženkliai padidina atakos sėkmės tikimybę. 2024 m. atliktas tyrimas tarp jaunų suaugusiųjų nustatė 100 % numatytąją tikimybę paspausti nuorodą, jei žinutė yra iš pažįstamo asmens, palyginti su labai maža tikimybe gaunant pranešimą iš nepažįstamo siuntėjo [KSB+24]. Žmonės beveik visada pasitikėjo ir spaudė nuorodas iš pažįstamų siuntėjų, o tai rodo pasitikėjimo principo galią. Sukčiai išnaudoja šį pasitikėjimo faktorių pažeisdami vieną paskyrą ir tuomet naudodamiesi ja siunčia žinutes jos kontaktams.

1.9.3. Socialinės inžinerijos principų taikymas

Šiuolaikiniai sukčiavimai ir duomenų viliojimo atakos remiasi gerai žinomais socialinės inžinerijos principais, iš dalies atitinkančiais Cialdini šešis įtakos principus – autoritetingumas, mainai, ribotumas, panašumas, socialinis patvirtinimas ir nuoseklumas. Neseniai publikuotuose tyrimuose pateikiama konkrečių įrodymų apie šių principų taikymą [AA24].

1. Autoritetingumas — Althobaiti ir kt. 2024 m. apžvalgoje nurodoma, kad sukčiai dažnai išnaudoja autoritetingumą, panašumą, ribotumą, socialinį patvirtinimą ir mainus, kad apgaulės būdu įgytų aukų pasitikėjimą. Autoritetingumas dažnai pasireiškia apsimetimu aukšto rango pareigūnu ar patikima įmone. [AA24].

2. Socialinis patvirtinimas — socialinis patvirtinimas rodomas frazėmis, kurios nurodo, jog visi darbuotojai ar draugai jau atliko šį veiksmą.
3. Ribotumas — ribotumo ir skubos elementas pasiūlymuose ar grėsmėse skatina paniką, nors viename iš tyrimų matoma, kad naudotojai nurodo ribotumą kaip mažiausiai patikimą, nes sukelta dirbtina skuba kelia įtarimų [AA24].
4. Mainai — mainų principas matomas laiškuose, žadančiuose atlygį ar paslaugą, siekiant suveikti instinktyviam norui duoti kažką atgal, pavyzdžiui, prisijungimo duomenis ar finansinę informaciją [FT18].
5. Panašumas — panašumo taktikos naudojimas pastebimai auga – sukčiai išnaudoja asmens duomenis, neretai paimtus iš socialinių tinklų, kad žinutės atrodytų labiau suasmenintos ir draugiškos arba yra perimamas aukos draugo paskyros valdymas. Pastarasis metodas yra ypatingai veiksmingas – socialinių tinklų duomenų viliojimo atakų atveju kenkėjiška nuoroda, atsiųsta iš pažįstamo asmens, ženkliai padidina atakos sėkmės tikimybę.
6. Pasitikėjimas — Althobaiti ir kt. 2024 m. atliktas tyrimas tarp jaunų suaugusiųjų nustatė 100 % numatytąją tikimybę paspausti nuorodą, jei žinutė yra iš pažįstamo asmens, palyginti su labai maža tikimybe gaunant pranešimą iš nepažįstamo siuntėjo [KSB$^+$24]. Žmonės beveik visada pasitikėjo ir spaudė nuorodas iš pažįstamų siuntėjų, o tai rodo pasitikėjimo principo galią. Sukčiai išnaudoja šį pasitikėjimo faktorių pažeisdami vieną paskyrą ir tuomet naudodamiesi ja siunčia žinutes jos kontaktams.

1.9.4. Atakų sudėtingėjimo tendencijos

Pastaraisiais metais duomenų viliojimo atakos tapo gerokai sudėtingesnės ir dar labiau išplito. Sukčiai pereina nuo primityvių „Nigerijos princo“ laiškų prie tuometinį kontekstą ir kasdienybę imituojančių temų. Akivaizdžios gramatinės, skyrybos ir stiliaus klaidos retėja [AA24]. Užpuolikai puikiai kopijuoja organizacijų prekės ženklus, naudoja tvarkingus elektroninių laiškų šablonus ir formatavimą. Taip pat padaugėjo kenksmingų pranešimų perdavimo kanalų — skambučiai, SMS žinutės, QR kodai, socialinių tinklų žinutės. Sukčiai pažįsta savo bandomą pasiekti auditoriją, tyrinėja taikinius, išnaudoja labiausiai pažeidžiamus jų gyvenimo momentus, krizes ir aktualius įvykius, tokius kaip gamtos stichijos, pandemijos ir karas, kad padidintų sėkmingos atakos tikimybę. Įvairios technologijos, tokios kaip dirbtinis intelektas, padeda jiems rašyti sklandesnius tekstus įvairiomis kalbomis, sunkiau atskiriamus nuo tikrų pranešimų [AA24; Nai15].

1.10. Pasitikėjimo ir saugumo projektavimas

1.10.1. Veiksniai lemiantys naudotojų pasitikėjimą turiniu

Vertindami svetainių ir elektroninių laiškų patikimumą, naudotojai remiasi tiek vaizdinėmis tiek kontekstinėmis užuominomis. Didžiausią pasitikėjimo efektą sukelia pažįstamas arba visuomenėje patikimas siuntėjas [KSB$^+$24]. Todėl patikimos organizacijos naudoja nuoseklius prekės ženklus, logotipus, oficialius domenus bei suasmenintus kreipinius autentiškumui parodyti.

Tyrimai kurie analizuoja duomenų viliojimo atakas skirsto pasitikėjimo požymius į tokias

kategorijas kaip atpažįstamumas, panašumas ir profesionalus įvaizdis bei pojūtis. [Nai15]. Esant šiems požymiams, naudotojai yra mažiau linkę nuodugniai tikrinti laiško turinį. Nepasitikėjimo požymiais dažniausiai įvardinami bendriniai pasisveikinimai, prasta gramatika ar skyryba, neatitinkančios nuorodos ar trūkstami logotipai, jie kelia įtarimų ir mažina pasitikėjimą. Naudotojai dažnai nurodo tokias neatitiktis kaip pagrindines nepasitikėjimo priežastis.

Vaizdinė estetika dažnai atlieka esminį vaidmenį. Gerai išplanuotas, neturintis klaidų elektroninis laiškas arba profesionaliai atrodanti svetainė atrodo patikimesnė nei neapgalvotas ar netvarkingas dizainas. Tačiau sukčiai bando nukopijuoti minėtus pasitikėjimo požymius, o tikrieji kūrėjai stengiasi išmokyti naudotojus atpažinti neatitikimus. Pastebima, kad daugelis sėkmingų duomenų viliojimo atakų turi matomų klaidų ir neatitikimų, tačiau vis tiek suveikia, nes naudotojai skuba ar yra mažai atidūs. Tai rodo, kad nors ir naudotojai geba aptikti nepasitikėjimo požymius, dažnai to tiesiog nedaro, jei nėra įpratę [Nai15].

Siekiant projektuoti sistemas, skatinančias pasitikėjimą, svarbu ne tik užtikrinti, kad tikros svetainės ar elektroniniai laiškai atrodytų patikimai, tačiau ir mokyti naudotojus ieškoti tam tikrų požymių bei atkreipti dėmesį, kai jie yra praleidžiami. Pavyzdžiui, bankai nuolatos primena savo klientams, kad į juos bus kreipiamasi ir nebus prašoma slaptažodžio elektroniniu paštu — tokiu būdu yra formuojamas pasitikėjimo modelis, dėl kurio bet kokie nukrypimai atkreipia dėmesį ar tampa akivaizdūs.

1.10.2. Saugumo estetika

Neskaitant pranešimų turinio, saugumo rodiklių ir sąsajų dizainas reikšmingai veikia naudotojų pasitikėjimą. Saugumo mechanizmai turi būti ne tik vaizdingai aiškūs, tačiau ir psichologiškai priimtini. Jeigu saugumo rodiklis, pavyzdžiui, naršyklės HTTPS spynos piktograma ar antivirusinės programos pranešimas, yra pernelyg techniškas ar subtilus, naudotojai gali jo nepastebėti. Jeigu jis yra pernelyg įkyrus ar bauginantis — naudotojas gali pasitikėti pranešimu ir ieškoti būdų jį apeiti. Todėl dizaino pusiausvyra yra labai svarbi.

Pastarųjų metų tyrimai nagrinėjo kaip spalvos, ikonografija ir dizaino maketo sprendimai gali veiksmingai perduoti saugumo informaciją. Vienas iš tyrimų parodė, kad naudojantis labiau dėmesį patraukiančias piktogramas, pavyzdžiui, ryškią įspėjimo žymę vietoje paprasto trikampio su šauktuku, į įspėjimus atkreiptas dėmesys ir nurodymų laikymasis žymiai pagerėjo [WSW+22]. Įspėjamųjų pranešimų išdėstymas ir stilius lemia, ar naudotojai juos perskaitys ar ignoruos. Saugumo estetika apima ir saugumo intuityvumo projektavimą. Gerai suprojektuotas prisijungimo langas su aiškiu prekės ženklu ir neperkrautu dizainu gali patikinti naudotoją, kad tai tikra svetainė, tuo tarpu labai paini ar pasenusi svetainė gali kelti įtarimų.

Kita vertus, sukčiai taip pat išnaudoja šiuos projektavimo sprendimus. Kuria iki kiekvieno pikselio apgalvotas duomenų viliojimo svetaines ar labai tikslias realių svetainių kopijas, tokiu būdu apgaudami net ir atidžius naudotojus. Tai paskatino kurti apsaugos nuo duomenų viliojimo atakų naršyklėse priemones, kurios vaizdiškai pabrėžia patikrintas svetaines. Pavyzdžiui, jeigu svetainės sertifikatas yra tikras, įmonės pavadinimas yra rodomas žalia spalva, siekiant pasinaudoti estetika pasitikėjimui sukurti. Vizualūs dizaino elementai stipriai veikia naudotojo saugumo pojūtį,

todėl projektuotojai stengiasi kurti patikimas naudotojų sąsajas su aiškiai pabrėžiamais saugumo ir pavojaus būsenomis, nesukeldami nereikalingos įtampos. Jų tikslas yra sukurti naudotojo sąsają, kuri perteikia saugumo jausmą ir ugdo naudotojų pasitikėjimą tuo pat metu įspėdama apie tikrąją riziką.

1.10.3. Saugumo įspėjimų ir klaidų psichologinis poveikis

Saugumo įspėjimai ir klaidų pranešimai turi gilų psichologinį poveikį naudotojams. Jie gali tiek įtikinti ir apsaugoti, tiek suerzinti ir sumažinti į juos kreipiamą dėmesį. Tyrimai atkreipia dėmesį į įspėjimų perteklių ir įpratimą. Jei naudotojai per dažnai mato saugumo pranešimus, ypač tuos, kurie neturi jokių realių pasekmių, jie pradeda įspėjimus ignoruoti [WSW+22]. Dauguma įspėjimų praktikoje tampa neveiksmingais, nes naudotojai prie jų pripranta. Pranešimai tampa tiesiog foniniu triukšmu jų aplinkoje. Tai kelia rimtą projektavimo problemą — pirmą kartą ryškus raudonas tekstas ir šauktukai gali atkreipti dėmesį, tačiau dvidešimtą kartą naudotojas gali juos praleisti, ypač jei prieš tai buvę pranešimai nepažymėjo realių grėsmių. Todėl siekiant išlaikyti įspėjimų patikimumą, svarbu sumažinti klaidingus teigiamus atvejus ir naudotojo darbą pertraukti tik esant tikrai kritinėms situacijoms. Taip pat yra taikomi baimės mechanizmai. Tokio tipo pranešimai dažnai naudoja gąsdinančią kalbą ir vaizdinius įspėjimus, siekdami priversti naudotojus vykdyti pranešime esančius veiksmus. Nors tokio tipo pranešimai gali motyvuoti naudotojus atlikti reikiamus veiksmus, tačiau pernelyg stipri baimė gali sukelti paniką, kurios metu naudotojas priims skubotus sprendimus, arba nepasitikėjimą, nes naudotojas laikys tokį pranešimą apgavyste ar perdėtu saugojimu. Įtikinamų pranešimų tyrimai rodo, kad geresni rezultatai pasiekiami suderinant refleksiją ir konkrečias rekomendacijas, o ne vien baimės skleidimą [RFC+18]. Pavyzdžiui, naršyklės perspėjimas apie galimą kenkėjišką svetainę, ne su standartine „Ši svetainė gali pakenkti jūsų įrenginiui“ žinute, tačiau su ramiu paaiškinimu ir rekomendacija „Grįžkite į saugią svetainę“ veiksmingiau skatina naudotojus išeiti iš rizikingos svetainė nesukeliant perteklinės panikos.

Kitas psichologinis veiksnys yra naudotojų pasitikėjimas saugumo sistemomis. Jeigu sistemos yra pernelyg sudėtingos ir neaiškios ar dažnai daro klaidas, naudotojai praranda pasitikėjimą jomis. Vienas tyrimas apie iššokančius saugumo pranešimus „Windows“ operacinėje sistemoje parodė, kad interaktyvūs ir naudotoją įtraukiantys įspėjamieji pranešimai, pavyzdžiui reikalaujantys įvesti veiksmo priežastį ar suteikiantys galimybę saugiai patikrinti failą prieš suteikiant pilną redagavimo galimybę, įtraukia naudotojus ir didina saugių veiksmų skaičių, lyginant su įprastais tekstiniais pranešimais [SPD+23]. Tačiau jei naudotojai nuolatos jaučiasi kovojantys su saugumo sistema dėl netikrų pavojų arba sudėtingos techninės kalbos pranešimuose, jie gali priprasti apeiti šias sistemas. Klaidingai teigiami duomenų viliojimo svetainių blokavimai mažina pasitikėjimą sistemos tikslumu ir lemia naudotojų neatidumą tikros atakos metu.

Šiuolaikiniai tyrimai pabrėžia psichologinę įspėjimų projektavimo svarbą — reikia naudoti aiškią bet ne per daug gąsdinančią kalbą, grafinius elementus, kurie perteikia riziką jos neperdėdami ir kartais net humorą naudotojo nerimui sumažinti. Svarbu, kad po įspėjimo būtų suteikiamas teigiamas grįžtamasis ryšys už pasirinktą saugų veiksmą, taip stiprinant saugią elgseną.

1.11. Apibendrinimas

Šios literatūros apžvalgos tikslas išanalizuoti mokslines publikacijas apie duomenų viliojimo atakas, naudotojų mokymą, naudotojo sąsajos dizaino principus bei aiškinamuosius ir duomenų viliojimo atakoms atsparius dizaino modelius, kad būtų atskleisti šiuo metu taikomi metodai ir jų spragos. Analizuoti pagrindiniai techniniai, psichologiniai ir dizaino aspektai, kurie daro įtaką atakų sėkmei, taip pat įvertintos egzistuojančios apsaugos priemonės, naudotojo švietimo metodai bei naudotojo sąsajos dizaino sprendimai.

Pagrindinis tikslas buvo išsiaiškinti, kaip galima integruoti aiškinamąjį dizainą, į naudotoją orientuotus įspėjimus ir sukčiavimui atsparius naudotojo sąsajos elementus taip, kad jie ne tik apsaugotų nuo atakų realiu metu, bet ir lavintų naudotojų gebėjimą savarankiškai atpažinti grėsmes. Pagrindinės išvados:

- Duomenų viliojimo atakos yra įvairios ir nuolat tobulėja, kombinuodamos technines ir socialinės inžinerijos priemones, išnaudodamos žmogaus psichologiją, emocijas ir įpročius.
- Techninės priemonės (pvz., blokavimo sąrašai) gali apsaugoti nuo žinomų grėsmių, bet jos neveiksmingos prieš naujas, tikslingas ar greitai besikeičiančias atakas.
- Naudotojų mokymas (simuliacijos, žaidimai, mokymai tinkamu metu) yra veiksmingas, bet reikalauja nuoseklaus integravimo į kasdienę naudotojo sąsają, kitaip nauda nebūna ilgalaikė.
- Aiškinamieji įspėjimai ir intuityvus sąsajų dizainas (pvz., paaiškinimai, kodėl svetainė įtartina) padeda ne tik apsisaugoti, bet ir sustiprina naudotojų žinias bei pasitikėjimą.
- Psichologinis aspektas – pasitikėjimo, dėmesio ir motyvacijos valdymas yra esminis veiksmingo dizaino komponentas. Gerai parinkti elgsenos skatinimo metodai ir įtikinamasis projektavimas ženkliai padidina naudotojų atsparumą.
- Veiksmingiausi sprendimai yra hibridiniai – jie derina automatizuotą aptikimą, naudotojo įsitraukimą, paaiškinimus, psichologinius motyvus bei aiškų dizainą.

Literatūros analizė parodė, kad norint veiksmingai apsaugoti naudotojus nuo duomenų viliojimo, reikia kurti kompleksinius sprendimus, kurie vienu metu užkerta kelią grėsmei, paaiškina naudotojui jos kilmę ir stiprina jo gebėjimą priimti saugius sprendimus ateityje. Ši apžvalga sudarė pagrindą gairėms, skirtoms naršyklės įskiepiui, kuris derins aiškinamąjį dizainą, psichologiškai pagrįstus įspėjimus ir realaus laiko aptikimą, taip padėdamas naudotojui tapti aktyvia apsaugos nuo duomenų viliojimo atakų grandimi.

2. Projektavimo metodo kūrimas

2.1. Kriterijai įrankio vertinimui apsaugai nuo duomenų viliojimo atakų

Remiantis atlikta literatūros analize, buvo suformuoti kriterijai, skirti įvertinti duomenų viliojimo atakų apsaugos įrankių veiksmingumą. Šie kriterijai kyla iš analizuotų tyrimų ir geriausių praktikų, apimančių technines galimybes, naudotojo sąsajos dizainą, edukacinius aspektus bei psichologinius principus. Kriterijų formulavimo metu buvo siekiama apimti visus esminius apsaugos aspektus, kurie buvo identifikuoti literatūros analizėje apsaugai nuo duomenų viliojimo atakų.

2.1.1. Kriterijų išvedimas iš literatūros

Kriterijai buvo išvesti analizuojant literatūros apžvalgoje aptartus pagrindinius aspektus ir atsižvelgiant į įvairių tyrimų rekomendacijas į vieningą vertinimo sistemą. Kriterijai pateikiami 4 lentelėje. Kiekviena kriterijų grupė atspindi konkretų literatūroje identifikuotą apsaugos aspektą.

Techniniai aptikimo kriterijai (1-5 kriterijai) remiasi techniniu duomenų viliojimo atakų aptikimo skyriuje aptartais metodais ir atspindi pagrindinę įrankio funkciją – gebėjimą identifikuoti grėsmes. Literatūra [KOO22] pabrėžia, kad veiksmingi įrankiai turi derinti kelis metodus: blokavimo sąrašus žinomoms atakoms aptikti, mašininio mokymosi algoritmus naujoms, anksčiau nematytoms atakoms atpažinti, ir aktyvų veiksmų stebėjimą, kad užkirstų kelią jautrios informacijos įvedimui į įtartinas svetaines. Šie kriterijai vertina ne tik aptikimo faktą, bet ir aptikimo metodų įvairovę bei gilumą. Tyrimai rodo, kad vien blokavimo sąrašais besiremiančios sistemos negali apsaugoti nuo nuolat kintančių ir naujų atakų, todėl būtina daugialypė apsauga. Taip pat svarbu, kad sistema gebėtų atpažinti ne tik pačią svetainę, bet ir pavojingus naudotojo veiksmus, pavyzdžiui, slaptažodžio įvedimą į nepatikimą formą.

Aiškinamojo projektavimo kriterijai (6-8 kriterijai) kyla iš aiškinamųjų pranešimų ir mokymo tinkamu metu tyrimų, kurie pabrėžia ne tik apsaugos, bet ir mokymo svarbą. PhishXplain [RTN25] ir panašūs sprendimai parodė, kad įrankiai turi ne tik blokuoti grėsmes, bet ir paaiškinti naudotojui, kodėl konkretus turinys yra pavojingas, paryškinant konkrečius įtartinus elementus puslapyje. Šis požiūris remiasi konstruktyvia mokymosi teorija – kai naudotojai supranta grėsmės priežastis, jie įgyja gebėjimą atpažinti panašias situacijas ateityje. Tyrimai [DAA+22] rodo, kad aiškinamieji pranešimai žymiai pagerina ilgalaikį naudotojų atsparumą atakoms, lyginant su vien blokavimo mechanizmais. Taip pat svarbu, kad paaiškinimai būtų pateikiami kontekste – šalia pavojingo elemento, o ne bendrine pranešimo juosta, kad naudotojui būtų aiškus ryšys tarp įspėjimo ir konkretaus elemento.

Atakoms atsparaus projektavimo kriterijai (9-10 kriterijai) atspindi tyrimų apie domeno paryškinimą [XPY+17] ir skaidrių eilučių metodą [BBK+25; PZS19] išvadas bei pabrėžia proaktyvaus dizaino svarbą. Šie kriterijai užtikrina, kad saugumo informacija būtų matoma ir lengvai pastebima naudotojo darbo sraute, net neįvykus atakos bandymui. Domeno vizualus išryškinimas padeda naudotojams greičiau identifikuoti netikrus domenus, o pilno URL rodymas šalia nuorodų elektroniniuose laiškuose leidžia iš karto matyti, kur veda nuoroda, nereikalaujant papildomų

naudotojo veiksmų. Šie kriterijai vertina įrankio gebėjimą integruoti saugumo informaciją į natūralų naudotojo darbo srautą, o ne kaip trukdantį elementą.

Mokomieji kriterijai (11–13 kriterijai) remiasi naudotojų mokymo skyriuje aptartais principais ir pabrėžia ilgalaikio mokymosi aspektą. Tyrimai apie mokymą tinkamu metu ir simuliacijas parodė, kad veiksmingiausi įrankiai ne tik apsaugo realiu metu, bet ir sistemingai moko naudotojus savarankiškai atpažinti grėsmes, pateikdami suprantamą, netechninį turinį [DFM+22]. Svarbu, kad mokymas būtų integruotas į darbinį procesą, o ne būtų atskiras, periodiškais įvykis. Taip pat būtina išvengti informacijos perkrovos – per daug techninės informacijos gali sumažinti naudotojų susidomėjimą ir mokymosi veiksmingumą. Šie kriterijai vertina, ar įrankis padeda formuoti ilgalaikius įgūdžius, kurie veiktų net ir be įrankio pagalbos, taip stiprinant bendrą organizacijos ar individo saugumo kultūrą.

Psichologiniai ir elgesio kriterijai (14–20 kriterijai) išplaukia iš elgesio ekonomikos ir įtikinamojo projektavimo tyrimų [CBJ+14; SPD+23; WSW+22] ir atspindi žmogaus psichologijos svarbą saugumo sprendimuose. Šie kriterijai užtikrina, kad įrankiai veiksmingai panaudotų įtikinimo elementus (autoritetą, socialinį įrodymą, mainus), išvengtų įspėjimų pertekliaus, kuris sukelia pripratimą ir ignoravimą, bei teiktų teigiamą grįžtamąjį ryšį už saugius pasirinkimus, taip formuodami saugų naudotojų elgesį be pernelyg didelio naudotojų apkrovimo. Tyrimai rodo, kad netinkamai suprojektuoti įspėjimai gali būti žalingi – dažni klaidingai teigiami rezultatai ar per agresyvūs įspėjimai sumažina pasitikėjimą sistema ir skatina naudotojus ieškoti būdų ją apeiti [RFC+18]. Todėl šie kriterijai vertina psichologinius įrankio dizaino aspektus: ar jis priverstinai atkreipia dėmesį tik esant tikrai grėsmei, ar reikalauja sąmoningo patvirtinimo prieš rizikingus veiksmus, ar naudoja paskatų (angl. nudges) principus saugiam elgesiui skatinti. Teigiamas grįžtamasis ryšys yra ypač svarbus – kai naudotojai gauna patvirtinimą už saugius pasirinkimus, jie jaučia kompetencijos ir kontrolės jausmą, kas stiprina jų motyvaciją likti budriais

Integraciniai ir pritaikymo kriterijai (21–24 kriterijai) atspindi hibridinių sprendimų, tokių kaip PhishXplain [RTN25], svarbą ir rodo, kad duomenų viliojimo atakos vyksta įvairiuose kanaluose ir nuolat tobulėja. Veiksmingi įrankiai turi derinti automatizuotą aptikimą su dirbtinio intelekto galimybėmis kontekstiniams paaiškinimams kurti, veikti keliuose komunikacijos kanaluose (el. paštas, naršyklė, SMS, socialiniai tinklai) ir gebėti prisitaikyti prie naujų grėsmių naudojant grįžtamąjį ryšį ir nuolatinį mokymąsi. Šie kriterijai atspindi šiuolaikinį požiūrį, kad apsauga negali būti statinė – sistemai reikia gebėti mokytis iš naujų atakų, pritaikyti savo aptikimo metodus ir komunikaciją individualiam naudotojui. Dirbtinio intelekto panaudojimas leidžia generuoti kontekstiškai tinkamus paaiškinimus realiuoju laiku, o veikimas keliuose kanaluose užtikrina nuoseklią apsaugą.

Kiekvienas kriterijus buvo suformuluotas kaip aiškus, išmatuojamas parametras (Taip/Ne/Iš dalies), leidžiantis objektyviai palyginti skirtingų įrankių galimybes ir nustatyti jų stipriąsias bei silpnąsias puses apsaugos nuo duomenų viliojimo atakų kontekste. Kai kurie kriterijai gali būti įvertinti kaip iš dalies įgyvendinti, nes daugelis įrankių gali turėti tam tikras galimybes, bet ne visiškai jas realizuoti. Tai ypač aktualu vertinant tokius aspektus kaip aiškinimo išsamumas, mokymo funkcijos ar psichologinių principų taikymas, kur gali būti įvairių įgyvendinimo lygių.

4 lentelė. Vertinimo kriterijai ir jų šaltiniai

Nr.	Kriterijus	Šaltiniai
K1	Ar aptinka duomenų viliojimo svetainę?	[KOO22]
K2	Ar žinoma duomenų viliojimo svetainė yra blokavimo sąraše?	[KOO22]
K3	Ar aptinka anksčiau nežinomas duomenų viliojimo svetaines?	[KOO22]
K4	Ar blokuoja/įspėja prieš įvedant jautrią informaciją?	[KOO22]
K5	Ar naudoja kelis analitinius metodus duomenų viliojimui aptikti?	[KOO22]
K6	Ar paaiškina kodėl turinys yra pavojingas?	[DAA+22; RTN25]
K7	Ar paryškina įtartinus elementus?	[RTN25]
K8	Ar įspėjimai pateikiami šalia pavojingo elemento?	[PZS19; RTN25]
K9	Ar tikrasis domenas vizualiai išryškintas?	[XPY+17]
K10	Ar rodo pilną URL šalia nuorodų?	[BBK+25; PZS19]
K11	Ar įspėjimai suprantami (ne pernelyg techniniai)?	[DFM+22]
K12	Ar teikia mokomąjį turinį sprendimo priėmimo metu?	[DFM+22]
K13	Ar padeda naudotojams mokytis atpažinti grėsmes savarankiškai?	[DFM+22]
K14	Ar išvengia perteklinio įspėjimų kiekio?	[RFC+18]
K15	Ar reikalauja naudotojo patvirtinimo prieš tęsiant?	[CBJ+14]
K16	Ar naudoja elgesio ekonomikos principus?	[CBJ+14]
K17	Ar naudoja įtikinimo elementus (autoritetas, socialinis įrodymas)?	[SPD+23; WSW+22]
K18	Ar priverstinai atkreipia dėmesį prieš rizikingus veiksmus?	[CBJ+14]
K19	Ar reikalauja papildomų naudotojo veiksmų/patvirtinimo?	[CBJ+14]
K20	Ar teikia teigiamą grįžtamąjį ryšį už saugius pasirinkimus?	[SPD+23]
K21	Ar automatiškai aptinka grėsmes?	[RTN25]
K22	Ar naudoja DI kontekstiniams paaiškinimams generuoti?	[RTN25]
K23	Ar apsaugo keliuose kanaluose (el. paštas, svetainės, SMS)?	[RTN25]
K24	Ar geba automatiškai atnaujinti savo aptikimo modelius bei naudoti grįžtamąjį ryšį?	[RTN25]

2.2. Įrankių analizė pagal kriterijus

Remiantis suformuluotais kriterijais, buvo atlikta dvylikos duomenų viliojimo atakų apsaugos įrankių analizė. Analizuoti įrankiai apima naršyklių integruotas apsaugos sistemas (Chrome Safe Browsing, Edge SmartScreen, Firefox Protection), naršyklių plėtinius (Netcraft Extension, PhishTank Sitechecker, Avast Online Security), elektroninio pašto apsaugos sprendimus (Gmail Protection, Defender for Office 365, Proofpoint TAP, Mimecast TTP), mobiliąją apsaugą (Lookout Mobile) bei telekomunikacijų operatoriaus sprendimą (Telia Safe). Kiekvienas įrankis buvo vertinamas pagal visus 24 kriterijus, remiantis viešai prieinama informacija apie jų funkcionalumą, techninius aprašymus bei naudotojo dokumentaciją. Vertinant naršyklių integruotas apsaugas, remtasi Google Safe Browsing [Goo25b], Microsoft Defender SmartScreen [Mic25b] ir Mozilla Phishing and Malware Protection [Moz25] oficialia dokumentacija. Naršyklių plėtinių anali-

zė pagrįsta Netcraft [Net25], PhishTank [Phi25] ir Avast Online Security [Ava25] oficialiomis svetainėmis ir aprašymais.

El. pašto apsaugos sprendimų vertinimui naudota informacija iš Gmail [Goo25a], Microsoft Defender for Office 365 [Mic25a], Proofpoint Targeted Attack Protection [Pro25] ir Mimecast Targeted Threat Protection [Mim25] techninių specifikacijų. Mobiliosios apsaugos srityje analizuotas Lookout Mobile Security [Loo25], o telekomunikacijų sprendimų – Telia Safe, kurio techninis pagrindas remiasi F-Secure Secure, dokumentacija [F-S25].

Vertinimas atliktas taikant trilaipsnę skalę: „Taip“ (funkcionalumas pilnai įgyvendintas), „Ne“ (funkcionalumas neįgyvendintas arba nėra viešai prieinamos informacijos apie jį), „Iš dalies“ (funkcionalumas įgyvendintas dalinai arba ribotai).

Siekiant vienodo vertinimo metodo, „Iš dalies“ įvertinimas taikomas tokiais atvejais:

- K3, K5, K24: Naudoja tik dalį metodų (pvz., tik blokavimo sąrašus ir ribotas euristikas, bet ne mašininio mokymosi algoritmus);
- K4: Įspėja apie jautrios informacijos įvedimą, bet nesiūlo aktyvaus blokavimo;
- K6: Pateikia bendrinius paaiškinimus („svetainė pranešta kaip įtartina“), bet nedetalizuoja konkrečių įtartinų elementų ar požymių;
- K7, K8: Paryškina arba rodo įspėjimus tik tam tikrose situacijose, bet ne nuosekliai visame įrankyje;
- K9, K10: Vizualiniai elementai egzistuoja, bet nėra pakankamai aiškūs ar išskirti, kad veiksmingai padėtų naudotojams;
- K11: Įspėjimai dažniausiai suprantami, bet kartais gali būti per techniški ar per bendrini;
- K12, K13: Teikia tam tikrus edukacinius patarimus ar statistiką, bet jie nėra integruoti į realaus laiko sprendimų priėmimą;
- K14: Dažniausiai išvengia perteklinio įspėjimų kiekio, bet kartais gali rodyti per daug įspėjimų mažo pavojaus situacijose;
- K15–K20: Naudoja kai kuriuos psichologinius principus (spalvos, perspėjimai), bet neturi visapusiško sisteminio įgyvendinimo;
- K22: Turi ribotą DI funkcionalumą tam tikroms užduotims, bet negeneruoja dinaminių kontekstui pritaikytų paaiškinimų;
- K23: Apsaugo vieną ar du kanalus (pvz., tik el. paštą), bet neturi visapusiškos daugiakanalės apsaugos.

2.2.1. Lyginamoji įrankių analizės lentelė

5 lentelė. Duomenų viliojimo atakų apsaugos įrankių palyginimas pagal suformuluotus kriterijus

Kriterijus	Chrome SB	Edge SS	Firefox P	Netcraft	PhishTank	Avast OS	Gmail P	Defender 365	Proofpoint	Mimecast	Lookout	Telia Safe
Techniniai aptikimo kriterijai												
1. Aptinka svetainę	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip
2. Blokavimo sąrašas	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip
3. Nežinomos atakos	Taip	Taip	Taip	Iš dalies	Ne	Taip	Taip	Taip	Taip	Taip	Taip	Iš dalies
4. Blokuoja įvestį	Taip	Taip	Taip	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Iš dalies	Ne
5. Keli metodai	Taip	Taip	Taip	Iš dalies	Ne	Taip	Taip	Taip	Taip	Taip	Taip	Iš dalies
Aiškinamojo projektavimo kriterijai												
6. Paaiškina pavojų	Ne	Iš dalies	Ne	Iš dalies	Ne	Iš dalies	Iš dalies	Iš dalies	Taip	Taip	Iš dalies	Ne
7. Paryškina elementus	Ne	Ne	Ne	Taip	Ne	Ne	Ne	Ne	Iš dalies	Iš dalies	Ne	Ne
8. Įspėjimai šalia elemento	Ne	Ne	Ne	Taip	Ne	Iš dalies	Ne	Ne	Iš dalies	Iš dalies	Ne	Ne
Atakoms atsparaus projektavimo kriterijai												
9. Domeno išryškinimas	Iš dalies	Iš dalies	Ne	Taip	Ne	Iš dalies	Ne	Ne	Ne	Ne	Ne	Ne
10. Rodo pilną URL	Ne	Ne	Ne	Taip	Taip	Iš dalies	Ne	Iš dalies	Iš dalies	Iš dalies	Ne	Ne
Mokomieji kriterijai												
11. Suprantami įspėjimai	Taip	Taip	Taip	Taip	Iš dalies	Taip	Taip	Taip	Taip	Taip	Taip	Taip
12. Mokymasis sprendimo metu	Ne	Ne	Ne	Iš dalies	Ne	Ne	Ne	Iš dalies	Taip	Taip	Ne	Ne
13. Savarankiškas atpažinimas	Ne	Ne	Ne	Iš dalies	Ne	Ne	Ne	Iš dalies	Taip	Taip	Iš dalies	Ne
Psichologiniai ir elgesio kriterijai												
14. Išvengia pertekliaus	Taip	Taip	Taip	Iš dalies	Taip	Iš dalies	Taip	Taip	Taip	Taip	Taip	Taip
15. Reikalauja patvirtinimo	Taip	Taip	Taip	Ne	Ne	Ne	Ne	Ne	Iš dalies	Iš dalies	Ne	Ne
16. Elgesio ekonomika	Ne	Iš dalies	Ne	Ne	Ne	Ne	Ne	Iš dalies	Taip	Taip	Ne	Ne
17. Įtikinimo elementai	Ne	Iš dalies	Ne	Ne	Ne	Iš dalies	Ne	Iš dalies	Taip	Taip	Ne	Ne
18. Priverstinis dėmesys	Taip	Taip	Taip	Ne	Ne	Ne	Ne	Iš dalies	Iš dalies	Iš dalies	Ne	Ne
19. Papildomi veiksmai	Taip	Taip	Taip	Ne	Ne	Ne	Ne	Ne	Iš dalies	Iš dalies	Ne	Ne
20. Teigiamas ryšys	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Iš dalies	Iš dalies	Ne	Ne
Integraciniai ir pritaikymo kriterijai												
21. Automatinis aptikimas	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip	Taip
22. DI paaiškinimai	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Iš dalies	Taip	Taip	Ne	Ne
23. Keli kanalai	Ne	Ne	Ne	Ne	Ne	Ne	Iš dalies	Taip	Taip	Taip	Taip	Iš dalies
24. Modelių atnaujinimas	Taip	Taip	Taip	Iš dalies	Iš dalies	Taip	Taip	Taip	Taip	Taip	Taip	Iš dalies
Iš viso „Taip“	11	11	10	8	4	7	7	9	14	14	8	5
Iš viso „Iš dalies“	2	6	0	9	2	7	3	9	9	9	4	3
Iš viso „Ne“	11	7	14	7	18	10	14	6	1	1	12	16
Bendras įvertinimas	12	14	10	12.5	5	10.5	8.5	13.5	18.5	18.5	10	6.5

2.2.2. Analizės rezultatai

Atlikus lyginamąją įrankių analizę, išryškėjo kelios esminės tendencijos ir įžvalgos apie esamas duomenų viliojimo atakų apsaugos priemones.

Beveik visi analizuoti įrankiai demonstruoja stiprų techninį pagrindą – visi geba aptikti duomenų viliojimo svetaines ir naudoja blokavimo sąrašus. Tačiau gebėjimas aptikti anksčiau nežinomas atakas varijuoja: naršyklių integruotos sistemos (Chrome, Edge, Firefox) ir verslui skirti sprendimai (Defender 365, Proofpoint, Mimecast) naudoja pažangius mašininio mokymosi algoritmus, tuo tarpu paprastesni įrankiai (PhishTank) remiasi tik žinomų atakų duomenų bazėmis. Įdomu tai, kad tik naršyklės (Chrome, Edge, Firefox) geba blokuoti jautrios informacijos įvedimą į įtartinas formas.

Didžiausias visų įrankių trūkumas – menkas aiškinamojo dizaino įgyvendinimas. Tik

pažangiausi verslui skirti sprendimai (Proofpoint TAP, Mimecast TTP) sistemingai paaiškina, kodėl turinys yra pavojingas, ir paryškina įtartinus elementus. Netcraft Extension išsiskiria kaip vienintelis naršyklės plėtinys, kuris paryškina įtartinus elementus ir pateikia įspėjimus šalia jų. Tai rodo, kad dauguma įrankių vis dar remiasi tradiciniais blokavimo ar bendriniais įspėjimų mechanizmais, nesuteikdami naudotojams mokymosi vertės.

Mokymo kriterijai taip pat yra viena silpniausių visų įrankių sričių. Nors dauguma įrankių pateikia suprantamus įspėjimus, labai mažai jų (tik Proofpoint ir Mimecast) teikia mokomąjį turinį sprendimo priėmimo metu ar padeda naudotojams mokytis savarankiškai atpažinti grėsmes. Tai rodo didelę galimybę tobulinti įrankius integruojant mokymo komponentus, kurie literatūros analizėje buvo identifikuoti kaip labai veiksmingi ilgalaikiam atsparumui didinti.

Psichologiniai ir elgesio kriterijai taip pat menkai įgyvendinti daugelyje įrankių. Tik Proofpoint ir Mimecast sistemingai naudoja elgesio ekonomikos principus ir įtikinimo elementus savo įspėjimuose. Naršyklių sprendimai (Chrome, Edge, Firefox) reikalauja naudotojo patvirtinimo prieš tęsiant į įtartinas svetaines, kas yra priverstinio dėmesio forma, tačiau jie neteikia jokio teigiamo grįžtamojo ryšio už saugius pasirinkimus. Tai rodo, kad dauguma įrankių orientuojasi į reaktyvią apsaugą, o ne į proaktyvų saugaus elgesio formavimą.

Verslui skirti sprendimai (Defender 365, Proofpoint, Mimecast) bei mobiliosios apsaugos įrankis (Lookout) išsiskiria gebėjimu apsaugoti keliuose kanaluose – ne tik naršyklėje, bet ir elektroniniame pašte, SMS žinutėse bei kitose platformose. Tai atitinka literatūroje pabrėžtą poreikį įvairiapusei apsaugai, nes duomenų viliojimo atakos vyksta įvairiuose kanaluose. Dirbtinio intelekto naudojimas kontekstiniams paaiškinimams generuoti vis dar yra reta praktika – tik Proofpoint ir Mimecast dalinai taiko šias technologijas.

Analizė atskleidžia aiškų atotrūkį tarp literatūroje rekomenduojamų geriausių praktikų ir realiai įgyvendinamo funkcionalumo. Pažangiausi įrankiai (Proofpoint TAP ir Mimecast TTP) pasiekė po 18.5 balų įvertinimą iš 24 galimų (77 %), kas rodo, kad net ir geriausi sprendimai turi galimybių tobulėti. Paprastesni sprendimai, tokie kaip PhishTank (5 balai) ar Telia Safe (6.5 balo), daugiausia teikia bazinę apsaugą be jokio mokomojo ar emocinio papildymo.

Ypač ryški spraga yra aiškinamojo dizaino, mokymo funkcijų ir psichologinių principų pritaikymo srityse – būtent tose srityse, kurias literatūros analizė identifikavo kaip kritinius ilgalaikiam naudotojų atsparumui formuoti. Tai rodo aiškią galimybę kurti naujus įrankius, kurie derins stiprų techninį pagrindą su į naudotoją orientuotu dizainu, mokymu ir psichologiniais principais.

2.3. Projektavimo gairės naršyklės įskiepiui

Remiantis atlikta literatūros analize ir esamų įrankių vertinimu, suformuluotos išsamios projektavimo gairės naršyklės įskiepiui, skirtam apsaugoti naudotojus nuo duomenų viliojimo atakų. Šios gairės apima techninius, dizaino, mokymo, psichologinius, integracijos ir pasitikėjimo aspektus, siekiant sukurti visapusišką ir į naudotoją orientuotą sprendimą. Šios gairės sudaro pagrindą kuriamam naršyklės įskiepiui, kuris ne tik veikia kaip pasyvi apsaugos priemonė, bet ir ugdo naudotojo sąmoningumą ir gebėjimą savarankiškai apsisaugoti nuo duomenų viliojimo atakų.

2.3.1. Techninis aptikimas ir blokavimas

G1. Daugiasluoksnis aptikimas. Įskiepis privalo derinti kelis grėsmių aptikimo metodus: viešai prieinamus ir nuolat atnaujinamus blokavimo sąrašus (pvz., OpenPhish, PhishTank) žinomoms atakoms blokuoti ir mašininio mokymosi algoritmus arba euristinius metodus naujoms *(angl. zero-day)* atakoms atpažinti automatiškai. (K1, K2, K3, K5, K21)

G2. Realaus laiko URL analizė. Įskiepis turi aktyviai analizuoti URL struktūrą, ieškodamas būdingų sukčiavimo požymių: klaidinančių subdomenų, specialiųjų simbolių naudojimo (pvz., homografų atakos) ar raktinių žodžių, nesusijusių su tikruoju domenu. (K3, K5)

G3. Svetainės turinio analizė. Sistema turi skenuoti svetainės HTML kodą ir matomą turinį, ieškodama įtartinų formų, oficialių logotipų neatitikimų, prastos gramatikos ar svetainės struktūros, kuri imituoja populiarias prisijungimo formas. (K3, K5)

G4. Išsami SSL/TLS sertifikatų patikra. Be standartinės sertifikato galiojimo patikros, įskiepis turi vertinti sertifikato amžių, išdavėją ir tipą. Naujai išduoti ar nemokami sertifikatai įtartinose svetainėse turėtų kelti didesnį pavojaus lygį.(K5)

G5. Aktyvi apsauga įvedimo metu. Įskiepis turi ne tik blokuoti svetainę, bet ir aktyviai stebėti naudotojo veiksmus. Aptikus bandymą įvesti jautrią informaciją (slaptažodį, banko kortelės duomenis) į įtartiną formą, įskiepis privalo nedelsiant blokuoti įvesties lauką ir pateikti aiškų kontekstinį įspėjimą. (K4)

G6. Kenkėjiško kodo aptikimas. Įskiepis turėtų atlikti bazinį puslapio kodo skenavimą, ieškant įterptų kenkėjiškų skriptų *(angl. malicious scripts)* ar paslėptų elementų, kurie gali rinkti duomenis be naudotojo žinios. (K5)

G7. Peradresavimų grandinės sekimas. Sistema turi analizuoti visą peradresavimų grandinę nuo pradinės nuorodos iki galutinio puslapio, nes sukčiai dažnai naudoja kelis peradresavimus, kad paslėptų galutinį kenkėjišką adresą. (K5)

2.3.2. Aiškinamasis ir atakoms atsparus dizainas

G8. Aiškinamieji įspėjimai. Užuot rodžius bendrinį įspėjimą („Pavojinga svetainė“), įskiepis turi aiškiai nurodyti, kodėl svetainė yra įtartina. Pavyzdžiui: „Šios svetainės domenas *(tikras-bankas.lt)* yra panašus į oficialų, bet toks nėra“, „Svetainė naudoja nesaugų HTTP protokolą“. (K6)

G9. Vizualus grėsmės elementų paryškinimas. Įspėjimo lange turėtų būti vizualiai paryškinti (pvz., apvesti raudona linija ar nuspalvinti geltonai) konkretūs grėsmę keliantys elementai – įtartinas domenas, neatitinkantis logotipas ar klaidinantis tekstas. (K7)

G10. Kontekstiniai įspėjimai. Įspėjimai turi būti pateikiami kuo arčiau pavojingo elemento. Pavyzdžiui, įspėjimo piktograma ar pranešimas turėtų atsirasti šalia įtartinos nuorodos ar įvesties lauko, o ne bendroje naršyklės pranešimų juostoje. (K8)

G11. Domeno išryškinimas adreso juostoje. Naršyklės adreso juostoje tikrasis svetainės domenas turi būti vizualiai paryškintas (kita spalva ar pastorintu šriftu), kad naudotojas galėtų lengvai jį atskirti nuo subdomenų. (K9)

G12. Skaidrios nuorodos. El. laiškuose ar socialinių tinklų pranešimuose, šalia paslėptų nuorodų (hipersaitų) turi būti pateikiamas pilnas URL adresas. (K10)

G13. Rizikos lygio indikatorius. Įskiepis turėtų naudoti intuityvią spalvų kodavimo sistemą (pavyzdžiui, žalia — saugu, geltona — atkreipkite dėmesį, raudona — pavojinga) arba piktogramas, kurios nuolat rodytų lankomos svetainės patikimumo lygį. (K11)

G14. Standartizuoti patikimumo ženklai. Dažnai lankomoms ir patikimoms svetainėms įskiepis galėtų rodytų specialų patikimumo ženklą, taip sukuriant kontrastą su nepatikimomis svetainėmis. (K11)

2.3.3. Mokomieji aspektai ir naudotojo ugdymas

G15. Mokymas tinkamu momentu *(angl. Just-in-Time)*. Kiekvienas įspėjimas turi būti ir pamokymas. Po paaiškinimo, kodėl svetainė yra blokuojama ar įtartina, reikėtų pateikti patarimą, pavyzdžiui „Visada patikrinkite ar domenas sutampa su oficialiu įmonės pavadinimu“.(K12, K13)

G16. Interaktyvūs mokomieji moduliai. Įskiepis galėtų turėti atskirą skiltį su trumpais mokymais ar testais, pavyzdžiui „Atpažink tikrą svetainę“, kuriuos naudotojas galėtų periodiškai atlikti. (K13)

G17. Suprantama kalba. Visa įskiepio pateikiama informacija turi būti parašyta paprasta, netechnine kalba, kad būtų suprantama įvairaus lygio naudotojams. (K11)

G18. Simuliacinės atakos. Įskiepis galėtų periodiškai (su naudotojo sutikimu) atlikti simuliacines duomenų viliojimo atakas, kad patikrintų ir lavintų naudotojo budrumą realioje aplinkoje. (K13)

G19. Asmeninė statistika ir progreso siekimas. Naudotojui turėtų būti prieinama jo asmeninės statistika: kiek atakų pavyko išvengti, su kokiomis grėsmėmis dažniausiai susiduriama, kaip pagerino atpažinimo įgūdžiai, tokiu būdu didinant įsitraukimą ir motyvaciją. (K13)

2.3.4. Psichologiniai ir elgesio principai

G20. Priverstinis dėmesys ir patvirtinimas. Prieš leidžiant naudotojui pateikti į potencialiai pavojingą svetainę, įskiepis turi reikalauti sąmoningo patvirtinimo. (K15, K18, K19)

G21. Įtikinamasis projektavimas. Įspėjimuose naudotoji psichologinius įtikinimo elementus, pavyzdžiui, socialinį įrodymą „95 % naudotojų šią svetainę pažymėjo kaip pavojingą“, arba autoriteto principą — „Kibernetinio saugumo ekspertai rekomenduoja vengti šios svetainės“. (K17)

G22. Teigiamas grįžtamasis ryšys. Įskiepis turėtų skatingi saugų elgesį. Kai naudotojas pamatęs įspėjimą nusprendžia grįžti į saugų puslapį, galima parodyti paskatinamąjį pranešimą „Puikus sprendimas! Jūs apsisaugojote nuo galimos grėsmės“. (K20)

G23. Įspėjimų nuovargio *(angl. warning fatigue)* vengimas. Įskiepis turi būti suprojektuotas taip, kad išvengtų klaidingų teigiamų įspėjimų. Per dažni įspėjimai sukelia susierzinimą, dėl kurio naudotojai pradeda ignoruoti net ir tikrus įspėjimus. (K14)

G24. Personalizuotos paskatos. Sistema turėtų prisitaikyti prie naudotojo elgesio. Pavyzdžiui, jeigu naudotojas per dažnai ignoruoja įspėjimus, įskiepis galėtų pradėti naudotis griežtesnius įspėjimus arba reikalauti patvirtinimo. (K16)

2.3.5. Integracija, pritaikomumas ir personalizavimas

G25. Daugiakanalė apsauga. Įskiepis turėtų veikti ir populiariausiose el. pašto svetainėse (Gmail, Outlook naršyklėje) ir socialiniais tinklais, kad padėtų analizuoti pavojingas nuorodas. (K23)

G26. Nuolatinis mokymasis ir atsinaujinimas. Įskiepis turi turėti mechanizmą, kuris leistų naudotojui lengvai pranešti apie įtartinas svetaines. Taip pat šie duomenys kartu su automatiškai surinkta informacija apie naujas atakas turi būti naudojami nuolatiniam modelio aptikimo tobulinimui. (K24)

G27. Personalizuoti nustatymai. Naudotojas turi galėti pasirinkti saugumo lygį ir kokio tipo pranešimus norėtų matyti.

G28. Dirbtinio intelekto panaudojimas paaiškinimams. Įskiepis galėtų naudoti dirbtinio intelekto sąsajas sukurti dinamiškesnius ir kontekstą atitinkančius paaiškinimus apie grėsmes, pritaikytus konkrečiai situacijai. (K22)

2.3.6. Pasitikėjimo ir patikimumo kūrimas

G29. Skaidrumas ir privatumas. Įskiepis turi aiškiai informuoti naudotoją, kokius duomenis ir kaip renka, kam naudoja. Turi būti pateikta aiški privatumo politika, o duomenų rinkimas minimalus, būtinas apsaugos funkcijai vykdyti.

G30. Minimalus našumo poveikis. Įskiepis turi būti optimizuotas taip, kad darytų kuo mažesnį poveikį naršymo greičiui ir sistemos resursams, kad naudotojai nebūtų linkę jo išjungti dėl greitaveikos.

G31. Lengvas valdymas. Įskiepio sąsaja turi būti intuityvi ir paprasta, svarbiausios funkcijos, tokios kaip laikinas išjungimas ar svetainės įtraukimas į leidžiamų ar blokuojamų sąrašą turi, būti lengvai pasiekiamos vienu ar dviem paspaudimais.

2.4. Prototipai ir jų testavimas

Siekiant įvertinti suformuluotų projektavimo gairių praktinį pritaikomumą ir surinkti naudotojų atsiliepimus dėl jų tobulinimo, bus sukurti du alternatyvūs maketai naudojant prototipavimo įrankį Figma. Maketai vizualizuos skirtingus gairių įgyvendinimo būdus, leidžiančius palyginti, kurios gairės ir kokie jų pateikimo metodai yra veiksmingiausi. Testavimo tikslas – ne vertinti galutinio produkto naudojamumą, o gauti grįžtamąjį ryšį apie pačias projektavimo gaires ir jų tarpusavio sąveiką.

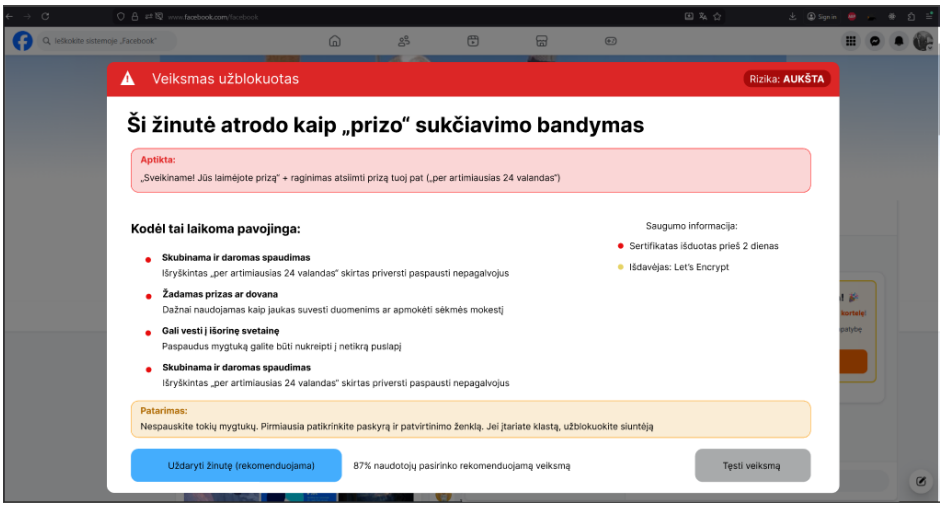
2.4.1. Alternatyvių maketų dizaino metodai

Abu maketai remiasi tomis pačiomis projektavimo gairėmis, tačiau skirtingai interpretuoja jų įgyvendinimą. Tai leis įvertinti, kuris požiūris geriau atitinka naudotojų lūkesčius ir veiksmingiau perteikia saugumo informaciją.

A metodas: Proaktyvus blokavimo modelis su išsamiais paaiškinimais

Šis metodas orientuotas į maksimalią apsaugą ir išsamų informavimą, kai įskiepis automatiškai blokuoja prieigą prie įtartinų svetainių ir pateikia visą turimą informaciją viename lange. Kai naudotojas bando patekti į įtartiną svetainę, jis susiduria su pilno ekrano blokavimo langu, kuris reikalauja priverstinio dėmesio ir patvirtinimo (G20) – naudotojas negali tęsti be sąmoningo sprendimo. Šiame lange pateikiami išsamūs aiškinamieji įspėjimai su konkrečiomis priežastimis (G8), pavyzdžiui, „Šios svetainės domenas panašus į oficialų, bet nesutampa“ arba „SSL sertifikatas išduotas tik prieš 2 dienas“. Grėsmės elementai yra vizualiai paryškinti (G9).

Maketą papildo rizikos lygio indikatorius su spalvų kodavimu ir aiškiomis piktogramomis (G13), suteikiantis greitą vizualų supratimą apie pavojaus lygį. Kiekvienas įspėjimas yra papildytas mokymu tinkamu momentu (G15) – patarimais, pavyzdžiui, „Visada patikrinkite, ar domenas tiksliai sutampa su oficialiu įmonės pavadinimu“. Taip pat integruoti įtikinimo elementai (G21), tokie kaip socialinis įrodymas „87 % naudotojų pažymėjo šią svetainę kaip pavojingą“ ir autoriteto principas. Visa informacija pateikiama suprantama kalba (G17) be techninių terminų, pritaikyta įvairaus lygio naudotojams. Norint tęsti į svetainę, reikalingas patvirtinimas (G20) – pirmiausia reikia perskaityti paaiškinimą, tada patvirtinti. Papildomai rodoma detalesnė SSL/TLS sertifikato informacija (G4), įskaitant sertifikato amžių, išdavėją. Šio metodo pavyzdys pateikiamas 1 paveikslėlyje.



1 pav. A metodo pavyzdys

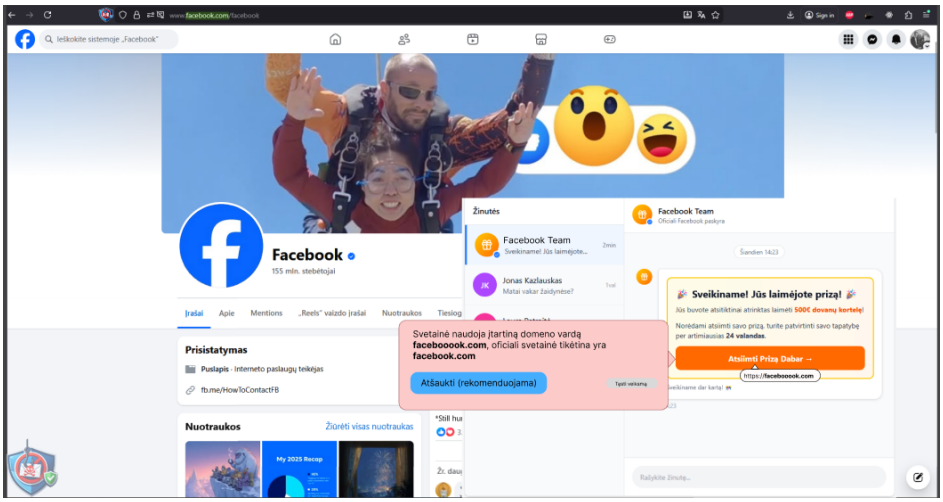
B metodas: Kontekstinis įspėjimų modelis su laipsniška informacija

Šis metodas orientuotas į mažiau įkyrią, tačiau nuolat matomą apsaugą, kai informacija pateikiama laipsniškai ir kontekstiškai, nepertraukiant naudotojo veiksmų. Pagrindinis elementas yra domeno išryškinimas adreso juostoje (G11), kur tikrasis domenas paryškintas pastorintu šriftu ir kita spalva, o subdomenai rodomi mažesniu šriftu, leidžiant naudotojui greitai identifikuoti tikrąjį domeną. Puslapio apačioje kairiame kampe nuolat matomas apsaugos programos logotipas su rizikos lygio indikatoriumi su intuityvia spalvų kodavimo sistema (G13), kuris nereikalauja papildomo dėmesio, tačiau suteikia pastovų saugumo būsenos suvokimą.

Dažnai lankomoms ir patikimoms svetainėms rodomi standartizuoti patikimumo ženklai (G14), kurie sukuria kontrastą su nepatikimomis svetainėmis. Užuot blokavęs visą svetainę,

metodas naudoja kontekstinius įspėjimus (G10), kurie pasirodo šalia konkrečių pavojingų elementų – įspėjimo piktograma ir trumpas pranešimas šalia įtartinos nuorodos ar įvesties lauko. El. laiškuose ar socialinių tinklų pranešimuose šalia nuorodų rodomas pilnas URL adresas (G12), užtikrinant skaidrias nuorodas.

Informacija atskleidžiama laipsniškai pagal aiškinamųjų įspėjimų principą (G8) – pirmiausia rodomas trumpas įspėjimas prie elemento, išsamesnis paaiškinimas pasirodo paspaudus. Aktyvi apsauga įvedimo metu (G5) reiškia, kad įspėjimas pasirodo tik kai naudotojas tikrai bando įvesti duomenis į įtartiną formą, ir tuomet įvesties laukas pažymimas. Metodas taip pat integruoja teigiamą grįžtamąjį ryšį (G22) – kai naudotojas pasirenka grįžti į saugų puslapį, rodomas paskatinamasis pranešimas, pavyzdžiui, „Puikus sprendimas!“. B metodo pavyzdys pateikiamas 2 paveikslėlyje.



2 pav. B metodo pavyzdys

2.4.2. Gairių testavimo scenarijai

Maketus lydi scenarijai, modeliuojantys realias situacijas, kuriose naudotojai gali susidurti su duomenų viliojimo atakomis. Kiekvienas scenarijus skirtas patikrinti konkrečias projektavimo gaires ir surinkti atsiliepimus apie jų aiškumą bei naudingumą.

1 scenarijus: Banko prisijungimo puslapis

Naudotojas gauna elektroninį laišką, informuojantį apie būtinybę patvirtinti banko sąskaitos duomenis. Laiškas atrodo oficialus, tačiau nuoroda veda į suklastotą banko prisijungimo puslapį su panašiu, bet ne identiškų domenu (pvz., `swedbank-lt.com` vietoj `swedbank.lt`).

Testuojamos gairės: G2 (URL analizė), G8 (aiškinamieji įspėjimai), G11 (domeno išryškinimas).

2 scenarijus: Socialinio tinklo pranešimas

Naudotojas mato pranešimą socialiniame tinkle apie laimėtą prizą. Nuoroda veda į puslapį, kuris imituoja „Facebook“ prisijungimo langą, tačiau yra kitame domene ir naudoja oficialų logotipą neteisėtai.

Testuojamos gairės: G3 (turinio analizė), G9 (grėsmės elementų paryškinimas), G12 (skaidrios nuorodos).

3 scenarijus: Siuntų tarnybos pranešimas

Naudotojas gauna SMS tipo pranešimą apie siuntą, kuriai reikia sumokėti muitą. Nuoroda veda į puslapį su mokėjimo forma, kuris naudoja HTTP (ne HTTPS) protokolą.

Testuojamos gairės: G4 (SSL sertifikatų patikra), G5 (aktyvi apsauga įvedimo metu), G10 (kontekstiniai įspėjimai).

4 scenarijus: Tikra svetainė

Kontrolinis scenarijus, kuriame naudotojas lankosi tikrame, saugiame banko ar elektroninės parduotuvės puslapyje. Šis scenarijus skirtas įvertinti, ar įskiepis nesukelia klaidingų teigiamų įspėjimų ir kaip veikia patikimumo ženklai.

Testuojamos gairės: G14 (patikimumo ženklai), G23 (įspėjimų nuovargio vengimas).

5 scenarijus: Homografinė ataka su peradresavimais

Sudėtingesnis scenarijus, kuriame domeno pavadinime naudojami panašūs simboliai iš kitų rašmenų sistemų (pvz., kirilicos „a“ vietoj lotyniškos „a“), o prieš pasiekiant galutinį puslapį vyksta keli peradresavimai.

Testuojamos gairės: G2 (URL analizė – homografai), G7 (peradresavimų grandinės sekimas), G17 (suprantama kalba paaiškinant sudėtingą grėsmę).

2.4.3. Testavimo užduotys

Kiekvienam scenarijui suformuluotos konkrečios užduotys, kurias dalyviai atliks naudodami abu maketus. Po kiekvienos užduoties dalyviai atsakys į klausimus apie gairių įgyvendinimo aiškumą ir naudingumą.

1 užduotis (1 scenarijus): „Gavote elektroninį laišką iš savo banko su prašymu patvirtinti paskyrą. Atidarykite nuorodą ir pabandykite prisijungti prie savo banko paskyros.“

2 užduotis (2 scenarijus): „Socialiniame tinkle matote pranešimą, kad laimėjote prizą. Paspauskite nuorodą ir prisijunkite, kad atsiimtumėte laimėjimą.“

3 užduotis (3 scenarijus): „Gavote pranešimą apie siuntą, už kurią reikia sumokėti 2,99 EUR muito mokestį. Atlikite mokėjimą.“

4 užduotis (4 scenarijus): „Norite patikrinti savo banko sąskaitos likutį. Eikite į oficialų banko puslapį ir prisijunkite.“

5 užduotis (5 scenarijus): „Gavote nuorodą į internetinę parduotuvę su ypatingu pasiūlymu. Peržiūrėkite pasiūlymą ir, jei norite, prisijunkite.“

2.4.4. Gairių vertinimo kriterijai

Testavimo metu bus renkami duomenys, skirti įvertinti ir tobulinti projektavimo gaires:

- Gairių aiškumas – ar naudotojas suprato kiekvienos gairės įgyvendinimo paskirtį ir pateiktą informaciją.
- Informacijos pakankamumas – ar pateikta informacija buvo pakankama sprendimui priimti, ar naudotojas norėjo daugiau/mažiau detalių.
- Gairių tarpusavio sąveika – ar skirtingos gairės papildo viena kitą, ar sukelia informacijos perteklių.

- Trūkstami elementai – ar naudotojai pastebėjo situacijas, kuriose trūko informacijos ar apsaugos.
- Pasiūlymai tobulinimui – kokybiniai naudotojų komentarai apie kiekvieną gairę.

2.4.5. Testavimo atlikimas

Testavimas atliktas su 10 tikslingai pagal demografinius kriterijus atrinktų dalyvių, siekiant užtikrinti reprezentatyvią imtį. Kriterijai dalyviams paskirstyti taip: pagal amžių – 3 asmenys iš 18–25 m. grupės, 4 asmenys iš 26–45 m. grupės ir 3 asmenys iš 46–65 m. grupės; pagal išsilavinimą – 3 atstovai su viduriniu/profesiniu ir 7 su aukštuoju išsilavinimu; pagal techninių įgūdžių lygį – 4 IT specialistai, 3 kasdieniai technologijų naudotojai (kitų sričių profesionalai) bei 3 nereguliarūs kompiuterio naudotojai. Kiekvienas dalyvis susipažino su užduotimi ir gavo trumpą įvadinį instruktažą apie tyrimo tikslą, nepateikiant per daug informacijos, kuri galėtų daryti įtaką jų natūraliam elgesiui.

Testuojama aplinka sukurta naudojant interaktyvius Figma maketus, kurie imituos realias situacijas. Kiekvienas dalyvis paeiliui atliko visas penkias užduotis su abiem maketais (A ir B metodu). Maketų pateikimo tvarka buvo keičiama tarp dalyvių, siekiant išvengti tvarkos įtakos rezultatams.

Po kiekvieno scenarijaus dalyviai užpildė trumpą klausimyną, kuriame įvertins (skalėje nuo 1 iki 5):

- Informacijos aiškumą ir suprantamumą;
- Grėsmės suvokimą ir atpažinimą;
- Sprendimo priėmimo palengvinimą;
- Vizualinių elementų veiksmingumą;
- Įspėjimų pastebimumą;
- Mokomosios vertės perteikimą;
- Sąsajos įkyrumą;
- Naudojimo patogumą;
- Bendrą įvertinimą.

Taip pat buvo užduodami klausimai apie informacijos pakankamumą (ar buvo per daug/per mažai informacijos, techninio detalumo, vizualinių elementų).

Testavimo metu buvo stebimas dalyvių elgesys ir fiksuojami komentarai, išsakyti „galvok garsiai“ metodu. Taip pat buvo registruojami visi sąveikos su maketu veiksmai – kokius elementus paspaudė, kiek laiko praleido skaitydami informaciją, ar ignoravo įspėjimus.

Po visų užduočių įvyko bendra apklausa, kurios metu dalyviai galėjo išsakyti bendrą nuomonę apie abiejų metodų stipriąsias ir silpnąsias puses, pasiūlyti patobulinimus. Buvo įvertinami šie aspektai:

- Bendras metodo pasirinkimas (kuris metodas jiems labiau patiko);
- Pasitikėjimo lygis kiekvienu metodu (1-5 skalė);
- Suvoktas saugumo lygis (1-5 skalė);

- Noras naudoti kasdien (1–5 skalė);
- Mokymo vertė (1–5 skalė).

Be to, buvo užduodami atviri klausimai apie trūkstamus elementus ir pasiūlymus tobulinimui, tokie kaip: „Ar buvo situacijų, kuriose norėjote daugiau informacijos?“, „Ką norėtumėte pakeisti ar papildyti?“.

Surinkti duomenys analizuoti kiekybiškai (įvertinimai skalėse, sėkmingai atliktos užduotys) ir kokybiškai (teminė dalyvių komentarų analizė). Šis derinys leido nustatyti, kurios gairės veikia geriausiai, kurias reikia tobulinti, ir kuris dizaino metodas yra pranašesnis.

2.4.6. Testavimo rezultatai

Dalyviai įvertino abu dizaino metodus pagal devynis aspektus naudodami 1–5 balų skalę. Kaip matyti 6 lentelėje, A metodas gavo aukštesnius įvertinimus grėsmės suvokimo ir atpažinimo (4.6) bei įspėjimų pastebimummo (4.8) aspektais, tačiau B metodas buvo geriau vertinamas informacijos aiškumo (4.5), sąsajos įkyrumo (4.6) ir naudojimo patogumo (4.5) aspektais. Bendras įvertinimas abiejų metodų buvo panašus – 4.2 (A metodas) ir 4.4 (B metodas).

6 lentelė. Dizaino metodų įvertinimas pagal aspektus (1–5 skalė, kur 5 – labai gerai, 1 – labai blogai)

Vertinimo aspektas	A metodas	B metodas
Informacijos aiškumas ir suprantamumas	4.2	4.5
Grėsmės suvokimas ir atpažinimas	4.6	4.1
Sprendimo priėmimo palengvinimas	4.3	4.4
Vizualinių elementų veiksmingumas	4.1	4.3
Įspėjimų pastebimumas	4.8	3.9
Mokomosios vertės perteikimas	4.4	4.2
Sąsajos įkyrumas	3.2	4.6
Naudojimo patogumas	3.8	4.5
Bendras įvertinimas	4.2	4.4

Scenarijų įvykdymo rezultatai (7 ir 8 lentelės) rodo, kaip dalyviai atliko užduotis naudodami kiekvieną metodą. A metodas pasižymėjo didesniu tikslumu visose kategorijose (88.6 % vidurkis), ypač homografinių atakų scenarijuje (78 %), tačiau užduočių atlikimas užtruko ilgiau (vidutiniškai 46.2 s). B metodas buvo greitesnis (33.0 s vidurkis), bet pasiekė mažesnį tikslumą (83.4 % vidurkis), ypač sudėtingesnėse situacijose kaip homografinė ataka (71 %).

7 lentelė. Scenarijų įvykdymo rezultatai (A metodas)

Scenarijus	Tikslumas (%)	Vid. laikas (s)	Supratimas (%)
1: Banko prisijungimas	92	45	95
2: Socialinis tinklas	88	38	91
3: Siuntų tarnyba	85	52	87
4: Tikra svetainė	100	28	98
5: Homografinė ataka	78	68	82
Vidurkis	88.6	46.2	90.6

8 lentelė. Scenarijų įvykdymo rezultatai (B metodas)

Scenarijus	Tikslumas (%)	Vid. laikas (s)	Supratimas (%)
1: Banko prisijungimas	87	32	92
2: Socialinis tinklas	82	28	88
3: Siuntų tarnyba	79	35	85
4: Tikra svetainė	98	22	96
5: Homografinė ataka	71	28	78
Vidurkis	83.4	33.0	87.8

Informacijos pakankamumo vertinimas (9 lentelė) atskleidė reikšmingus skirtumus tarp metodų. A metodas dažniau buvo kritikuojamas dėl per didelio informacijos kiekio (45 % dalyvių) ir techninio detalumo (40 %), o B metodas – dėl nepakankamo techninio detalumo (20 %) ir vizualinių elementų (25 %). Tai patvirtina poreikį lanksčiam, kelių lygių informacijos pateikimo būdui.

9 lentelė. Informacijos pakankamumas

Kriterijus	A metodas		B metodas	
	Per daug	Per mažai	Per daug	Per mažai
Informacijos kiekis (dalyvių %)	45	10	15	5
Techninis detalumas (dalyvių %)	40	5	10	20
Vizualinių elementų (dalyvių %)	15	10	15	25

Bendrame metodo pasirinkimo palyginime (10 lentelė) 60 % dalyvių teikė pirmenybę B metodui, nors A metodas buvo vertinamas kaip sukeliantis didesnį pasitikėjimą (4.5 ir 4.3). Svarbu pažymėti, kad noras naudoti įrankį kasdien buvo žymiai didesnis B metodui (4.4 ir 3.6), o mokymo vertė abiejuose metoduose buvo panaši (4.4 ir 4.3).

10 lentelė. Metodo pasirinkimo palyginimas

Aspektas	A metodas	B metodas
Bendras naudotojų pasirinkimas (%)	40	60
Pasitikėjimo lygis (1–5)	4.5	4.3
Suvoktas saugumo lygis (1–5)	4.1	4.2
Noras naudoti kasdien (1–5)	3.6	4.4
Mokymo vertė (1–5)	4.4	4.3

Dalyvių atsiliepimai apie trūkstamus elementus ir pasiūlymus tobulinimui (11 lentelė) parodė konkrečias tobulinimo kryptis. Dažniausiai minimi trūkumai buvo susiję su nepakankama informacija apie saugias svetaines (4 paminėjimai), per dažnais blokavimo langais (3 paminėjimai), ir norais turėti išsamesnę analizę bei mokomuosius patarimus (po 3 paminėjimus).

11 lentelė. Trūkstami elementai ir pasiūlymai tobulinimui

Trūkstamas elementas / problema	Paminėjimų skaičius	Pasiūlymas tobulinimui
Per dažni blokavimo langai trukdo naršymą	3	Naudoti adaptyvų blokavimą – blokuoti tik labai pavojingas svetaines, kitiems rodant įspėjimus
Sunku pastebėti įspėjimus sudėtingesnėse situacijose	1	Padidinti įspėjimų matomumą homografinėms ir sudėtingoms atakoms, naudoti animacijas
Trūksta istorijos funkcijos peržiūrėti ankstesnius įspėjimus	2	Pridėti įspėjimų istoriją su galimybe peržiūrėti, kokios grėsmės buvo aptiktos
Nepakanka informacijos, kodėl svetainė saugi (tikroje svetainėje)	4	Aiškiau parodyti patikimumo indikatorius ir paaiškinti, kodėl svetainė laikoma saugia
Norėtųsi turėti galimybę matyti išsamesnę analizę iš karto	3	Pridėti papildomą mygtuką „Rodyti detalią analizę“ šalia trumpų įspėjimų
Neaišku, kaip pranešti apie klaidingą įspėjimą	1	Integruoti grįžtamojo ryšio mechanizmą su mygtuku „Pranešti apie klaidą“
Trūksta mokomųjų patarimų, kaip bendrai atpažinti sukčiavimą	3	Pridėti trumpą interaktyvų vadovą pirmojo naudojimo metu ir periodinius patarimus

Remiantis testavimo rezultatais, projektavimo gairės bus patikslintos ir papildytos. Bus nustatyta, kurias gaires reikia sujungti, kurias išplėsti, o kurios galbūt yra perteklinės. Taip pat bus pasirinktas tinkamesnis dizaino metodas arba sukurtas hibridinis sprendimas galutiniam įskiepiui kurti.

2.4.7. Gairių tobulinimas

Remiantis testavimo rezultatais ir naudotojų grįžtamuoju ryšiu, pradinės projektavimo gairės (G1–G31) buvo peržiūrėtos ir patikslintos. Testavimas parodė, kad nei vienas iš metodų (A ar B) atskirai nėra optimalus – geriausias sprendimas derina abiejų metodų stipriąsias puses, išvengiant jų trūkumų.

Patobulinta G14 (Standartizuoti patikimumo ženklai) – Aktyvūs patikimumo indikatoriai

Ne tik įtartinoms svetainėms rodyti įspėjimus, bet ir patikimoms svetainėms aiškiai pažymėti patikimumo indikatorius su trumpu paaiškinimu, kodėl svetainė laikoma saugia (pvz., „Oficialus banko domenas, galiojantis SSL sertifikatas, reguliariai lankoma“).

Pagrindimas: dalyviai nurodė, kad tikroje svetainėje norėjo aiškesnio patvirtinimo apie saugumą, ne tik grėsmių nebuvimą.

Patobulinta G2 ir G11 (URL analizė ir domeno išryškinimas) – Tobulinta vizualizacija homografinėms atakoms

Homografinių atakų atveju ne tik paryškinti domeną, bet ir vizualiai parodyti skirtumą naudojant spalvų kodavimą skirtingiems simbolių rinkiniams (pvz., kirilica viena spalva, lotyniška abėcėlė kita) arba rodyti įspėjimą „Domene naudojami simboliai iš skirtingų rašmenų“.

Pagrindimas: Homografinių atakų scenarijai turėjo žemiausius tikslumų rezultatus, rodant poreikį aiškesnei vizualizacijai.

Patobulinta G16 (Interaktyvūs mokomieji moduliai) – Pirmojo naudojimo vadovas ir periodinis mokymas

Pirmą kartą naudojant įskiepį, pateikti trumpą (2–3 min.) interaktyvų vadovą, kuris paaiškina pagrindinius duomenų viliojimo požymius ir parodo, kaip veikia įskiepis. Vėliau periodiškai (pvz., kartą per savaitę) rodyti trumpus edukacinius patarimus apie naujas sukčiavimo schemas.

Pagrindimas: dalyviai norėjo bendresnių mokomųjų patarimų, ne tik kontekstinių. Tai padėtų didinti ilgalaikę naudotojų kompetenciją. Taip pat trūksta pamokymų kaip naudotis įrankiu.

Patobulinta G24 (Personalizuotos paskatos) – Adaptyvus elgesio pritaikymas

Sistema turi stebėti naudotojo elgesį ir prisitaikyti: jei naudotojas dažnai ignoruoja įspėjimus, padidinti jų intensyvumą arba pasiūlyti papildomą mokymą. Jei naudotojas sėkmingai atpažįsta grėsmes, sumažinti informacijos kiekį ir pasiūlyti „pažangų režimą“ su kompaktiškomis nuorodomis.

Pagrindimas: Skirtingi naudotojai turi skirtingą patirties lygį, todėl vienodas požiūris nėra optimalus visiems.

Nauja gairė G32 – Išsami įspėjimų istorija

Įskiepis saugo visų aptiktų grėsmių istoriją su galimybe peržiūrėti ankstesnius įspėjimus, naudotojo veiksmus (ar ignoravo įspėjimą, ar grįžo), ir grėsmės tipus. Istorija prieinama per vieną paspaudimą ir vizualizuota grafiškai (laiko juosta ar kategorizuotas sąrašas).

Pagrindimas: dalyviai paminėjo trūkstamą istorijos funkciją. Tai padėtų naudotojams matyti savo pažangą ir mokytis iš ankstesnių situacijų.

Nauja gairė G33 – Grįžtamojo ryšio mechanizmas

Prie kiekvieno įspėjimo lengvai pasiekiamas mechanizmas pranešti apie klaidingą įspėjimą arba patvirtinti grėsmę. Tai apima mygtukus „Pranešti apie klaidą“ ir „Patvirtinti grėsmę“, kurie padeda tobulinti sistemą ir didina naudotojų pasitikėjimą.

Pagrindimas: dalyviai nurodė, kad neaišku, kaip pranešti apie klaidingą įspėjimą (susijusi su G26 nuolatiniu mokymusi). Tai svarbu norint išvengti įspėjimų nuovargio (G23).

Galutinis įskiepis kuriamas remiantis hibridiniu požiūriu, kuris derina abiejų metodų pranašumus. Aukšto pavojaus grėsmėms, tokioms kaip homografinės atakos, žinomi sukčiavimai ar nauji domenai su jautrios informacijos formomis, taikomas A metodo pilno ekrano blokavimas su išsamiu paaiškinimu ir priverstinio patvirtinimo mechanizmu (G20). Vidutinio pavojaus grėsmėms, pavyzdžiui, įtartiniems domenams ar SSL sertifikatų problemoms, naudojami B metodo kontekstiniai įspėjimai (G10) su kelių lygių informacija. Patikimos svetainės aiškiai pažymimos patikimumo indikatoriais (patobulinta G14). Visais atvejais integruojamas grįžtamojo ryšio mechanizmas (G33), saugoma įspėjimų istorija (G32) ir naudojamas adaptyvumas (patobulinta G24).

Šis hibridinis metodas užtikrina, kad naudotojai jaustųsi saugūs, bet nepatirtų per didelio įkyrumo, kartu gaudami mokomąją vertę ir norą naudoti įrankį kasdien. Testavimo rezultatai patvirtino, kad abiejų metodų derinys leidžia maksimaliai išnaudoti jų privalumus: aukštą grėsmių atpažinimo tikslumą (A metodo stiprybė) ir mažesnį kliudymą kasdieniame darbe (B metodo stiprybė). Patobulintų gairių rinkinys, apimantis 33 gaires (G1-G33), sudaro tvirtą pagrindą kurti veiksmingus ir į naudotoją orientuotus duomenų viliojimo atakų apsaugos įrankius.

2.5. Projektavimo metodas

Remiantis atlikta mokslinės literatūros analize, 12 esamų apsaugos įrankių vertinimu pagal 24 kriterijus ir dviejų alternatyvių naudotojo sąsajos maketų kokybinio vertinimo rezultatais, sukurtas projektavimo metodas, skirtas kurti naudotojo sąsajas, apsaugančias nuo duomenų viliojimo atakų. Metodą sudaro 33 projektavimo gairės (G1–G33) ir jų taikymo rekomendacijos, leidžiančios projektuotojams pritaikyti gaires konkrečiam kontekstui, tikslinei auditorijai ir kuriamo įrankio pobūdžiui. Gairės organizuotos į penkias funkcines grupes pagal jų paskirtį apsaugos sistemoje.

2.5.1. Gairių grupės ir tikslai

Projektavimo gairės suskirstytos į penkias grupes, kurių kiekviena atspindi skirtingą apsaugos aspektą:

1. **I grupė. Techninis aptikimas (G1–G7).** Šios gairės apibrėžia automatinius grėsmių aptikimo mechanizmus — daugiasluoksnį aptikimą, URL ir turinio analizę, SSL sertifikatų patikrą, aktyvią apsaugą įvedimo metu, kenkėjiško kodo aptikimą ir peradresavimų sekimą. Grupės tikslas — užtikrinti, kad sistema gebėtų techniškai identifikuoti grėsmes be naudotojo įsikišimo.
2. **II grupė. Aiškinamasis ir atsparaus dizaino projektavimas (G8–G14).** Šios gairės apibrėžia, kaip aptiktos grėsmės komunikuojamos naudotojui — aiškinamieji įspėjimai, vizualinis paryškinimas, kontekstiniai pranešimai, domeno išryškinimas, skaidrios nuo-

rodos, rizikos indikatoriai ir patikimumo ženklai. Grupės tikslas — padaryti grėsmes matomas ir suprantamas, o ne tiesiog jas blokuoti.

3. **III grupė. Mokomieji elementai (G15–G19).** Šios gairės apibrėžia edukacinę sistemos funkciją — mokymą tinkamu momentu, interaktyvius mokomuosius modulius, suprantamą kalbą, simuliacines atakas ir asmeninę statistiką. Grupės tikslas — formuoti ilgalaikius naudotojų gebėjimus savarankiškai atpažinti grėsmes.
4. **IV grupė. Psichologiniai ir elgsenos principai (G20–G24).** Šios gairės apibrėžia elgesio ekonomikos ir psichologijos principų taikymą — priverstinį patvirtinimą, įtikinamąjį projektavimą, teigiamą grįžtamąjį ryšį, įspėjimų nuovargio vengimą ir personalizuotas paskatas. Grupės tikslas — maksimizuoti apsaugos veiksmingumą pasitelkiant žinias apie žmogaus sprendimų priėmimą.
5. **V grupė. Integracija, adaptyvumas ir pasitikėjimo kūrimas (G25–G33).** Šios gairės apibrėžia sistemos integracijos, pritaikymo ir ilgalaikio patikimumo aspektus — daugiakanalę apsaugą, nuolatinį mokymąsi, personalizuotus nustatymus, dirbtinio intelekto panaudojimą, skaidrumą, našumą, lengvą valdymą, įspėjimų istoriją ir grįžtamojo ryšio mechanizmą. Grupės tikslas — užtikrinti sistemos ilgaamžiškumą, naudotojų pasitikėjimą ir gebėjimą prisitaikyti prie naujų grėsmių.

2.5.2. Galutinis gairių sąrašas

Žemiau pateikiamos visos 33 projektavimo gairės su galutinėmis formuluotėmis, sugrupuotos pagal funkcinę paskirtį.

I grupė. Techninis aptikimas

G1. Daugiasluoksnis aptikimas. Įskiepis privalo derinti kelis grėsmių aptikimo metodus: viešai prieinamus ir nuolat atnaujinamus blokavimo sąrašus (pvz., OpenPhish, PhishTank) žinomoms atakoms blokuoti ir mašininio mokymosi algoritmus arba euristinius metodus naujoms *(angl. zero-day)* atakoms atpažinti automatiškai.

G2. Realaus laiko URL analizė. Įskiepis turi aktyviai analizuoti URL struktūrą, ieškodamas būdingų sukčiavimo požymių: klaidinančių subdomenų, specialiųjų simbolių naudojimo (pvz., homografų atakos) ar raktinių žodžių, nesusijusių su tikruoju domenu. Homografinių atakų atveju turi būti taikomas spalvų kodavimas skirtingiems simbolių rinkiniams (pvz., kirilica viena spalva, lotyniška abėcėlė kita) arba rodomas įspėjimas apie skirtingus rašmenis domene.

G3. Svetainės turinio analizė. Sistema turi skenuoti svetainės HTML kodą ir matomą turinį, ieškodama įtartinų formų, oficialių logotipų neatitikimų, prastos gramatikos ar svetainės struktūros, kuri imituoja populiarias prisijungimo formas.

G4. Išsami SSL/TLS sertifikatų patikra. Be standartinės sertifikato galiojimo patikros, įskiepis turi vertinti sertifikato amžių, išdavėją ir tipą. Naujai išduoti ar nemokami sertifikatai įtartinose svetainėse turėtų kelti didesnį pavojaus lygį.

G5. Aktyvi apsauga įvedimo metu. Įskiepis turi ne tik blokuoti svetainę, bet ir aktyviai stebėti naudotojo veiksmus. Aptikus bandymą įvesti jautrią informaciją (slaptažodį, banko

kortelės duomenis) į įtartiną formą, įskiepis privalo nedelsiant blokuoti įvesties lauką ir pateikti aiškų kontekstinį įspėjimą.

G6. Kenkėjiško kodo aptikimas. Įskiepis turėtų atlikti bazinį puslapio kodo skenavimą, ieškant įterptų kenkėjiškų skriptų ar paslėptų elementų, kurie gali rinkti duomenis be naudotojo žinios.

G7. Peradresavimų grandinės sekimas. Sistema turi analizuoti visą peradresavimų grandinę nuo pradinės nuorodos iki galutinio puslapio, nes sukčiai dažnai naudoja kelis peradresavimus, kad paslėptų galutinį kenkėjišką adresą.

II grupė. Aiškinamasis ir atsparaus dizaino projektavimas

G8. Aiškinamieji įspėjimai. Užuot rodžius bendrinį įspėjimą (pvz., „Pavojinga svetainė“), įskiepis turi aiškiai nurodyti, kodėl svetainė yra įtartina. Pavyzdžiui: „Šios svetainės domenas yra panašus į oficialų, bet toks nėra“ arba „Svetainė naudoja nesaugų HTTP protokolą“.

G9. Vizualus grėsmės elementų paryškinimas. Įspėjimo lange turėtų būti vizualiai paryškinti (pvz., apvesti raudona linija ar nuspalvinti geltonai) konkretūs grėsmę keliantys elementai — įtartinas domenas, neatitinkantis logotipas ar klaidinantis tekstas.

G10. Kontekstiniai įspėjimai. Įspėjimai turi būti pateikiami kuo arčiau pavojingo elemento. Pavyzdžiui, įspėjimo piktograma ar pranešimas turėtų atsirasti šalia įtartinos nuorodos ar įvesties lauko, o ne bendroje naršyklės pranešimų juostoje.

G11. Domeno išryškinimas adreso juostoje. Naršyklės adreso juostoje tikrasis svetainės domenas turi būti vizualiai paryškintas (kita spalva ar pastorintu šriftu), kad naudotojas galėtų lengvai jį atskirti nuo subdomenų.

G12. Skaidrios nuorodos. El. laiškuose ar socialinių tinklų pranešimuose šalia paslėptų nuorodų (hipersaitų) turi būti pateikiamas pilnas URL adresas.

G13. Rizikos lygio indikatorius. Įskiepis turėtų naudoti intuityvią spalvų kodavimo sistemą (žalia — saugu, geltona — atkreipkite dėmesį, raudona — pavojinga) arba piktogramas, kurios nuolat rodytų lankomos svetainės patikimumo lygį.

G14. Aktyvūs patikimumo indikatoriai. Ne tik įtartinoms svetainėms rodyti įspėjimus, bet ir patikimoms svetainėms aiškiai pažymėti patikimumo indikatorius su trumpu paaiškinimu, kodėl svetainė laikoma saugia (pvz., „Oficialus banko domenas, galiojantis SSL sertifikatas, reguliariai lankoma“). Tai sukuria kontrastą su nepatikimomis svetainėmis ir suteikia naudotojui aiškų patvirtinimą apie saugumą.

III grupė. Mokomieji elementai

G15. Mokymas tinkamu momentu *(angl. Just-in-Time)*. Kiekvienas įspėjimas turi būti ir pamokymas. Po paaiškinimo, kodėl svetainė yra blokuojama ar įtartina, reikėtų pateikti patarimą, pavyzdžiui „Visada patikrinkite, ar domenas sutampa su oficialiu įmonės pavadinimu“.

G16. Pirmojo naudojimo vadovas ir periodinis mokymas. Pirmą kartą naudojant įskiepį, pateikti trumpą (2–3 min.) interaktyvų vadovą, kuris paaiškina pagrindinius

duomenų viliojimo požymius ir parodo, kaip veikia įskiepis. Vėliau periodiškai (pvz., kartą per savaitę) rodyti trumpus edukacinius patarimus apie naujas sukčiavimo schemas.

G17. Suprantama kalba. Visa įskiepio pateikiama informacija turi būti parašyta paprasta, netechnine kalba, kad būtų suprantama įvairaus lygio naudotojams.

G18. Simuliacinės atakos. Įskiepis galėtų periodiškai (su naudotojo sutikimu) atlikti simuliacines duomenų viliojimo atakas, kad patikrintų ir lavintų naudotojo budrumą realioje aplinkoje.

G19. Asmeninė statistika ir progreso siekimas. Naudotojui turėtų būti prieinama jo asmeninė statistika: kiek atakų pavyko išvengti, su kokiomis grėsmėmis dažniausiai susiduriama, kaip pagerino atpažinimo įgūdžius, tokiu būdu didinant įsitraukimą ir motyvaciją.

IV grupė. Psichologiniai ir elgsenos principai

G20. Priverstinis dėmesys ir patvirtinimas. Prieš leidžiant naudotojui patekti į potencialiai pavojingą svetainę, įskiepis turi reikalauti sąmoningo patvirtinimo — naudotojas negali tęsti be aktyvaus sprendimo.

G21. Įtikinamasis projektavimas. Įspėjimuose naudoti psichologinius įtikinimo elementus, pavyzdžiui, socialinį įrodymą („Dauguma naudotojų šią svetainę pažymėjo kaip pavojingą“) arba autoriteto principą („Kibernetinio saugumo ekspertai rekomenduoja vengti šios svetainės“).

G22. Teigiamas grįžtamasis ryšys. Įskiepis turėtų skatinti saugų elgesį. Kai naudotojas, pamatęs įspėjimą, nusprendžia grįžti į saugų puslapį, galima parodyti paskatinamąjį pranešimą, pavyzdžiui, „Puikus sprendimas! Jūs apsisaugojote nuo galimos grėsmės“.

G23. Įspėjimų nuovargio vengimas. Įskiepis turi būti suprojektuotas taip, kad išvengtų klaidingų teigiamų įspėjimų. Per dažni įspėjimai sukelia susierzinimą, dėl kurio naudotojai pradeda ignoruoti net ir tikrus įspėjimus.

G24. Adaptyvus elgesio pritaikymas. Sistema turi stebėti naudotojo elgesį ir prisitaikyti: jei naudotojas dažnai ignoruoja įspėjimus, padidinti jų intensyvumą arba pasiūlyti papildomą mokymą. Jei naudotojas sėkmingai atpažįsta grėsmes, sumažinti informacijos kiekį ir pasiūlyti „pažangų režimą“ su kompaktiškomis nuorodomis.

V grupė. Integracija, adaptyvumas ir pasitikėjimo kūrimas

G25. Daugiakanalė apsauga. Įskiepis turėtų veikti ir populiariausiose el. pašto svetainėse (Gmail, Outlook naršyklėje) ir socialiniuose tinkluose, kad padėtų analizuoti pavojingas nuorodas.

G26. Nuolatinis mokymasis ir atsinaujinimas. Įskiepis turi turėti mechanizmą, kuris leistų naudotojui lengvai pranešti apie įtartinas svetaines. Šie duomenys kartu su automatiškai surinkta informacija apie naujas atakas turi būti naudojami nuolatiniam aptikimo modelio tobulinimui.

G27. Personalizuoti nustatymai. Naudotojas turi galėti pasirinkti saugumo lygį ir kokio tipo pranešimus norėtų matyti.

G28. Dirbtinio intelekto panaudojimas paaiškinimams. Įskiepis galėtų naudoti dirb-

tinio intelekto sąsajas sukurti dinamiškesnius ir kontekstą atitinkančius paaiškinimus apie grėsmes, pritaikytus konkrečiai situacijai.

G29. Skaidrumas ir privatumas. Įskiepis turi aiškiai informuoti naudotoją, kokius duomenis ir kaip renka, kam naudoja. Turi būti pateikta aiški privatumo politika, o duomenų rinkimas minimalus, būtinas apsaugos funkcijai vykdyti.

G30. Minimalus našumo poveikis. Įskiepis turi būti optimizuotas taip, kad darytų kuo mažesnį poveikį naršymo greičiui ir sistemos resursams, kad naudotojai nebūtų linkę jo išjungti dėl greitaveikos.

G31. Lengvas valdymas. Įskiepio sąsaja turi būti intuityvi ir paprasta, svarbiausios funkcijos (laikinas išjungimas, svetainės įtraukimas į leidžiamų ar blokuojamų sąrašą) turi būti lengvai pasiekiamos vienu ar dviem paspaudimais.

G32. Išsami įspėjimų istorija. Įskiepis saugo visų aptiktų grėsmių istoriją su galimybe peržiūrėti ankstesnius įspėjimus, naudotojo veiksmus (ar ignoravo įspėjimą, ar grįžo) ir grėsmės tipus. Istorija prieinama per vieną paspaudimą ir vizualizuota grafiškai (laiko juosta ar kategorizuotas sąrašas).

G33. Grįžtamojo ryšio mechanizmas. Prie kiekvieno įspėjimo lengvai pasiekiamas mechanizmas pranešti apie klaidingą įspėjimą arba patvirtinti grėsmę. Tai apima mygtukus „Pranešti apie klaidą“ ir „Patvirtinti grėsmę“, kurie padeda tobulinti sistemą ir didina naudotojų pasitikėjimą.

2.5.3. Taikymo rekomendacijos

Nors visų 33 gairių įgyvendinimas užtikrina geriausią apsaugos rezultatą, praktikoje projektuotojai dažnai dirba su ribotu laiku, biudžetu ar specifine kuriamo įrankio paskirtimi. Todėl metodas pateikia taikymo rekomendacijas, padedančias pasirinkti tinkamiausią gairių rinkinį pagal kontekstą.

Minimalus gairių rinkinys. Nepriklausomai nuo kuriamo įrankio tipo ar tikslinės auditorijos, rekomenduojama visada įgyvendinti šias bazines gaires, kurios sudaro apsaugos pagrindą:

- G1 (daugiasluoksnis aptikimas) — užtikrina, kad sistema apskritai aptiktų grėsmes;
- G2 (URL analizė) — pagrindinis techninis aptikimo mechanizmas;
- G4 (SSL/TLS patikra) — bazinis svetainės patikimumo vertinimas;
- G8 (aiškinamieji įspėjimai) — užtikrina, kad naudotojas suprastų grėsmės priežastį;
- G13 (rizikos indikatorius) — greitas vizualinis saugumo būsenos suvokimas;
- G15 (mokymas tinkamu momentu) — kiekvienas įspėjimas tampa mokymosi galimybe;
- G17 (suprantama kalba) — informacija prieinama visiems naudotojams;
- G20 (priverstinis patvirtinimas) — apsauga nuo skubotų sprendimų;
- G23 (įspėjimų nuovargio vengimas) — užtikrina ilgalaikį naudotojų pasitikėjimą;
- G29 (skaidrumas ir privatumas) — etinis ir teisinis pagrindas;
- G30 (minimalus našumo poveikis) — praktinė naudojimo prielaida;
- G31 (lengvas valdymas) — sumažina barjerą naudoti įrankį.

Gairių parinkimas pagal kuriamo įrankio kontekstą:

- **El. pašto apsaugos įrankis.** Prioritetinės gairės: G1–G4, G8, G10, G12, G13, G15, G17, G20, G21, G23. Mažiau aktualios: G5, G6, G7, G11 (orientuotos į svetainių naršymą, ne el. paštą).
- **Naršyklės plėtinys.** Tinkamos visos 33 gairės. Tai išsamiausias taikymo kontekstas, kuriam metodas ir buvo sukurtas.
- **Edukacinis ar sąmoningumo didinimo įrankis.** Prioritetinės gairės: G8, G13, G15, G16, G17, G18, G19, G21, G22, G32. Mažiau aktualios: G6, G7 (techninis aptikimas ne pagrindinis tikslas).
- **Sprendimas organizacijoms.** Papildomai rekomenduojamos: G18 (simuliacinės atakos darbuotojų testavimui), G19 (statistika vadovams), G25 (daugiakanalė apsauga), G28 (DI paaiškinimams).

Gairių parinkimas pagal tikslinę auditoriją:

- **Mažos IT patirties naudotojai.** Ypač svarbios: G8 (aiškinamieji pranešimai), G15 (mokymas tinkamu momentu), G16 (pirmojo naudojimo vadovas), G17 (suprantama kalba), G21 (įtikinamasis projektavimas), G22 (teigiamas grįžtamasis ryšys). Šiai grupei G24 (adaptyvus elgesio pritaikymas) tampa kritine, nes sistema turi aktyviai prisitaikyti prie naudotojo mokymosi tempo.
- **Techniniai naudotojai.** Galima sumažinti G17 (suprantama kalba) svarbą, nes šie naudotojai supranta techninius terminus. Rekomenduojama stiprinti G28 (DI paaiškinimams) ir G27 (personalizuoti nustatymai), suteikiant daugiau konfigūravimo laisvės.
- **Mišri auditorija.** G24 (adaptyvus elgesio pritaikymas) tampa esmine gaire, nes sistema turi gebėti automatiškai pritaikyti informacijos kiekį ir detalumo lygį pagal kiekvieno naudotojo elgesį ir pažangą.

Geriausi rezultatai pasiekiami įgyvendinus visas 33 gaires. Tačiau dalinis įgyvendinimas yra priimtinas, kai kuriamo įrankio apimtis yra ribota. Rekomenduojama visada pradėti nuo minimalaus gairių rinkinio (12 bazinių gairių), tada papildyti grupėmis, aktualiomis konkrečiam kontekstui. Gairės yra tarpusavyje suderintos ir papildančios viena kitą — pavyzdžiui, G8 (aiškinamieji įspėjimai) veikia efektyviau kartu su G15 (mokymas tinkamu momentu), o G20 (priverstinis patvirtinimas) geriau pasitarnauja, kai derinamas su G23 (įspėjimų nuovargio vengimas).

3. Įskiepio kūrimas ir testavimas

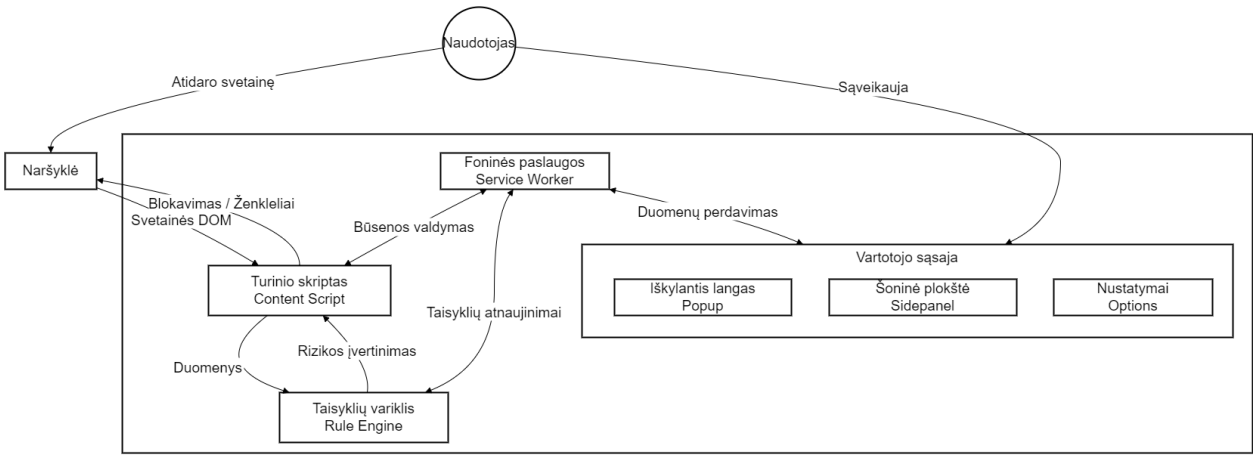
3.1. Architektūra ir techninis įgyvendinimas

Remiantis testavimo metu nustatytu hibridiniu projektavimo metodu, kuris derina A metodo (proaktyvus blokavimas) ir B metodo (kontekstiniai įspėjimai) pranašumus, buvo sukurtas *Chrome* naršyklės plėtinys. Plėtinys remiasi *Manifest V3* (MV3) standartais ir yra orientuotas į lokalų, tikslų atakų aptikimą, nereikalaujantį išorinių paslaugų.

Plėtinio architektūra susideda iš keturių pagrindinių komponentų:

1. Foninės paslaugos (*Service Worker*) — veikia nuolat ir tvarko bendrosios sistemos logiką, taisyklių valdymą ir duomenų saugojimą. Šis komponentas tiesiogiai neblokuoja grėsmių, bet suteikia informaciją turinio ir naudotojo sąsajos komponentams.
2. Turinio skriptai (*Content Script*) — injektuojami į kiekvieną puslapį ir atlieka realaus laiko domenų, URL, formų ir atvaizduoto teksto analizę.
3. Naudotojo sąsaja (*UI*) — sudaro trys dalys:
   - Iškylantis langas (*Popup*) — greitas atsakas apie esamos svetainės statusą.
   - Šoninė sekcija (*Side Panel*) — išsami ataskaita su visomis aptiktomis grėsmėmis ir paaiškinimais.
   - Nustatymai (*Options*) — naudotojai gali pasirinkti kalbą, valdyti taisykles ir reguliuoti jautrumą.
4. Taisyklių variklis (*Rule Engine*) — deterministinis vertintojas, kuris grindžiamas sudarytomis projektavimo gairėmis (G1-G33) ir atlieka šiuos patikrinimus:
   - G2, G3, G4, G11 — URL analizė, domeno panašumas, SSL sertifikatai, homografinės atakos.
   - G5 — formų jautrių duomenų aptikimas.
   - G6 — kenkėjiško kodo fragmentų aptikimas.
   - G8, G9 — teksto ir vizualinių indikatorių analizė.

Plėtinio duomenų srautas aprašytas schemoje (3 pav.): turinio skriptas nuskaito svetainės elementus, perduoda juos varikliui, kuris grąžina rizikos įvertinimą ir paaiškinimą. Šie duomenys rodomi kontekstiniais ženkleliais (B metodas žemesnei rizikai) arba pilno ekrano blokavimo langais (A metodas aukštai rizikai, remiantis patobulintomis G20 ir G24 gairėmis). Visais atvejais integruojamas grįžtamojo ryšio mechanizmas (G33) ir saugoma įspėjimų istorija (G32).



3 pav. Naršyklės plėtinio architektūra

3.2. Plėtinio pagrindinės funkcijos

Plėtinys naudotojui teikia keturis pagrindinius funkcionalumus:

3.2.1. Realaus laiko aptikimas

Kiekvienoje svetainėje, kurią naudotojas aplanko, plėtinys automatiškai atlieka šiuos patikrinimus remdamasis projektavimo gairėmis:

- URL analizė (G2, G11) — tikrina domeno panašumą su žinomais draudžiamais domenais, homografinius simbolius, neįprastus subdomenus.
- SSL sertifikatų patikra (G4) — nustato nesaugias HTTP jungtis, pasibaigusio galiojimo ar suklastotus sertifikatus.
- Formos analizė (G5) — aptinka jautrius laukus (slaptažodis, banko kortelė) ir, jei jie rasti įtartinose formose, parodo įspėjimą.
- Teksto analizė (G8, G15) — ieško skubos žodžių, baimės, grasinimo, autoriteto ar atlygio elementų.
- Domeno patikimumo patikra (G14) — rodo žalia spalva žinomai saugias svetaines su standartizuotais patikimumo ženklais.

3.2.2. Sluoksniniai aiškinamieji įspėjimai

Plėtinys naudoja kelių lygių informacijos strategiją (G8):

1. Pirmasis lygis — įterptas ženklelis šalia žymeklio, rodantis grėsmės lygį spalva (vizualizacija pateikiama 4 pav.):
   - Žalia (*Safe*) — oficiali, patikima svetainė.
   - Geltona (*Caution*) — įtartina, reikalingas dėmesys.
   - Raudona (*Dangerous*) — didelė rizika, rekomenduojama neatsidaryti.

2. Antrasis lygis — trumpas pranešimas šalia ženklelio, pvz., „Šios svetainės domenas panašus į banko, bet su juo nesutampa“.
3. Trečiasis lygis — išsami analizė šoninėje sekcijoje (4 pav.) su konkrečiomis nustatytomis priežastimis, rekomendacijomis ir aiškinimais (G8, G15).

3.2.3. Šoninė ataskaita

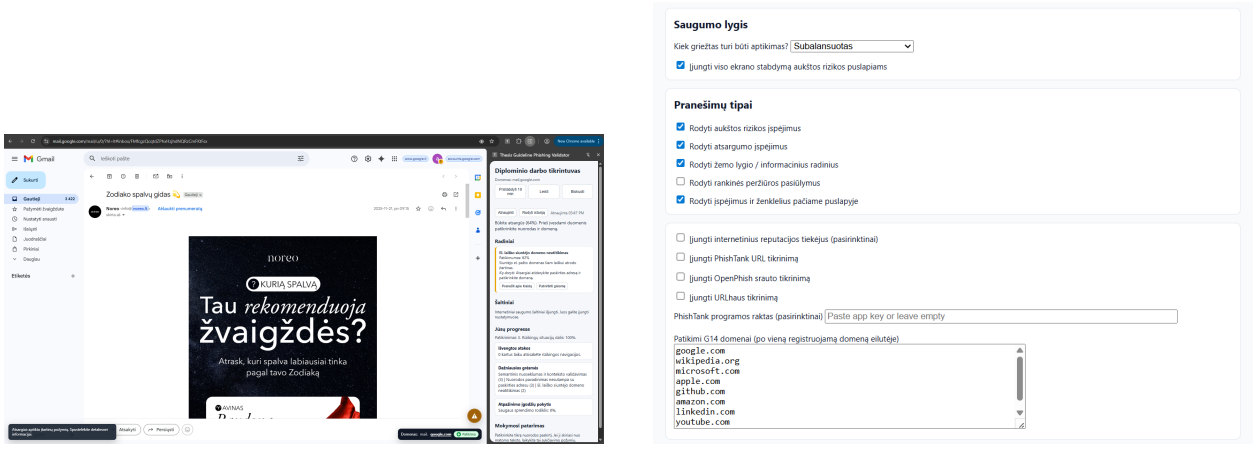
Naudotojui paspaudus ženklelį arba atidarant plėtinio meniu, atidaroma visa ataskaita su:

- Visomis aptiktomis grėsmėmis ir jų detalėmis;
- Spalvų žymėjimo sistema: žalia (saugu), geltona (įtartina), raudona (pavojinga) (G13);
- Mokomaisiais patarimais ir rekomendacijomis, kaip apsisaugoti nuo panašių atakų (G15);
- Mygtukais „Pranešti apie klaidą“ ir „Patvirtinti grėsmę“ grįžtamajam ryšiui (G33);
- Ankstesnių įspėjimų istorijos peržiūra su laiko žymėmis (G32).

3.2.4. Personalizuoti nustatymai

Naudotojai gali:

- Pasirinkti kalbą (lietuvių, anglų);
- Reguliuoti plėtinio jautrumą (laisvas, normalus, griežtas režimai), adaptuojant sistemą prie naudotojo patirties lygio (G24);
- Leisti arba riboti svetaines pagal taisykles;
- Konfigūruoti, kaip dažnai rodyti mokomuosius patarimus.



4 pav. Plėtinio naudotojo sąsajos: įspėjimai ir šoninė sekcija (kairėje) bei nustatymai (dešinėje)

3.3. Testavimo metodologija

3.3.1. Testavimo tikslas ir fazės

Testavimo tikslas buvo nustatyti, ar sukurtas plėtinys padeda naudotojams atpažinti duomenų viliojimo atakas ir ar jis lavina ilgalaikius identifikavimo įgūdžius. Testavimas buvo atliekamas trimis etapais, iš kurių galima apskaičiuoti plėtinio naudą:

1. Kontrolinis etapas (*Baseline*) — naudotojai bandė atpažinti duomenų viliojimo atakas savarankiškai, be plėtinio pagalbos. Šis etapas nustato jų pradinį lygį, gebėjimą atpažinti grėsmes.
2. Mokymasis su plėtiniu (*Learning*) — tie patys naudotojai susidūrė su tais pačiais scenarijais naudodami plėtinį. Jų tikslas buvo pamatyti, kiek plėtinys padeda iš karto.
3. Išmoktų pamokų taikymas (*Retention*) — dalyviai vėl bandė atpažinti atakas savarankiškai (be plėtinio), tačiau šį kartą jie naudojosi žiniomis, kurias įgijo ankstesniame etape. Šis etapas parodo, ar įgytos žinios išlieka atsisakius pagalbinio įrankio.

Šis trifazis modelis leidžia išmatuoti tris skirtingus rodiklius:

- Tiesioginė plėtinio nauda — skirtumas tarp kontrolinės stadijos ir mokymosi stadijos.
- Mokymosi greitis — ar naudotojai greitai suprato rodytą informaciją.
- Žinių išlaikymas — ar naudotojai sugeba veiksmingai pritaikyti išmoktą informaciją net neturėdami plėtinio apsaugos.

3.3.2. Dalyviai

Testavime dalyvavo 10 naudotojų, sugrupuotų pagal IT kompetenciją:

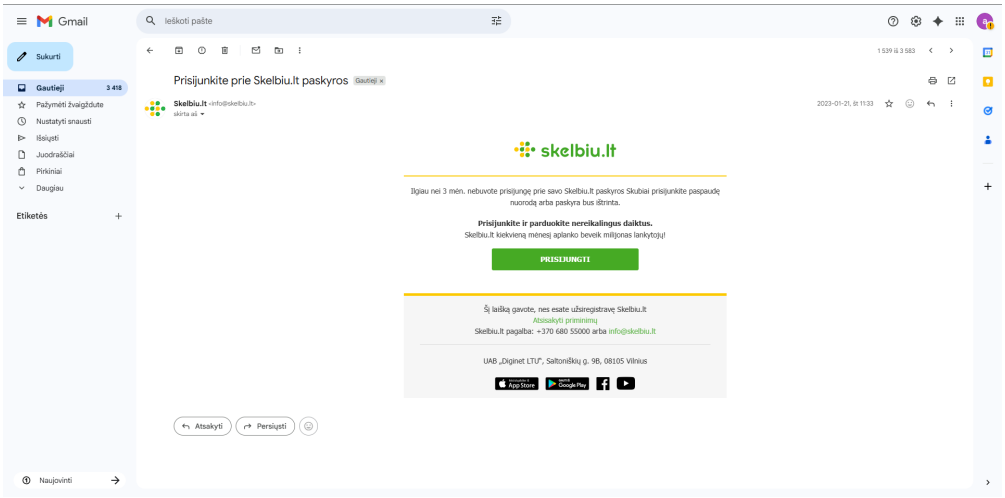
- 4 IT specialistai — profesionalai, kurie dirba informacinių technologijų arba kibernetinio saugumo srityje;
- 3 kasdieniai naudotojai — asmenys, kurie kompiuteriu naudojasi kasdien darbo tikslais, tačiau tai nėra jų pagrindinė specializacija;
- 3 nereguliarūs naudotojai — asmenys, kurie kompiuteriu ir elektroniniu paštu naudojasi retai, dažniausiai asmeniniams tikslams.

Ši demografija buvo pasirinkta siekiant suprasti plėtinio naudą skirtingoms naudotojų grupėms.

3.3.3. Testavimo scenarijai

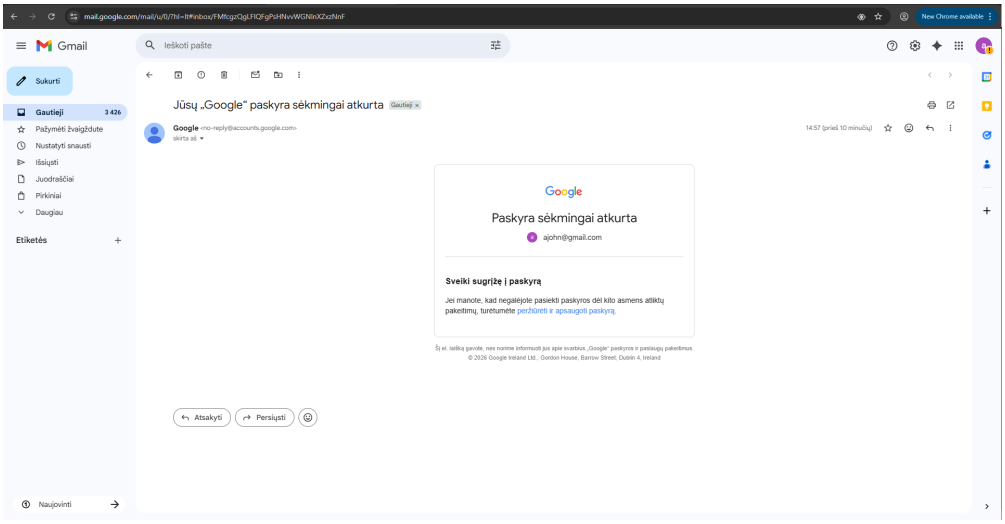
Testavime buvo naudoti 5 el. paštu grįsti scenarijai, kurie reprezentuoja realias duomenų viliojimo atakas. Kadangi analizuojamos 3 fazės pagal kiekvieno asmens mokymąsi susidūrus su tuo pačiu tipu atakos, kiekvienam asmeniui kiekviena fazė pateikė po 2 praktines kiekvieno scenarijaus užduotis (iškart generuodama 10 atsakymų iš kiekvieno žmogaus per fazę). Tokiu būdu vienos fazės skaičiavimams buvo remiamasi 100 gautų atsakymų bendrai.

1 scenarijus: Prisijungimo patvirtinimas. Naudotojui buvo parodytas el. laiškas, kuriame reikalaujama „patvirtinti paskyrą“, kartu su pateikta nuoroda (5 pav.). Nuoroda vedė į suklastotą puslapį, kuris imitavo oficialią sistemą, tačiau domenas buvo tik panašus, bet ne identiškas (pvz., `paslauga-lt.com` vietoj `paslauga.lt`). Testuojamos taisyklės: G2 (URL analizė), G8 (aiškinamieji pranešimai), G11 (domeno išryškinimas).



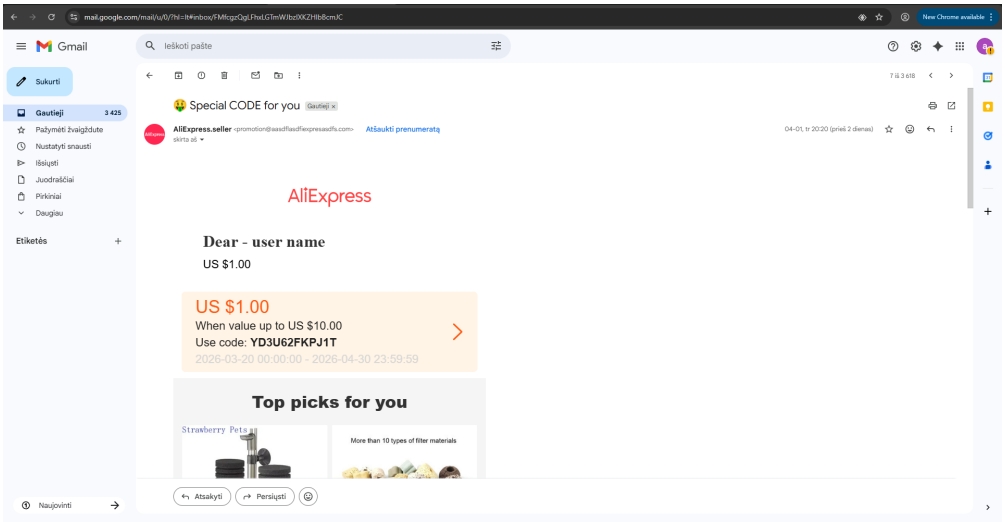
5 pav. 1 scenarijus: Prisijungimo patvirtinimas

2 scenarijus: Paskyros atkūrimo patvirtinimas. Naudotojas gauna laišką apie pakeistą paskyros slaptažodį su prašymu jį patvirtinti (6 pav.). Laiškas atrodo oficialus, nuoroda veda į puslapį su prisijungimo forma, kuris naudoja tokią pat URL struktūrą kaip tikras puslapis, tačiau yra iš kito serverio. Testuojamos taisyklės: G3 (turinio analizė), G5 (aktyvi formų apsauga), G10 (kontekstiniai pranešimai).



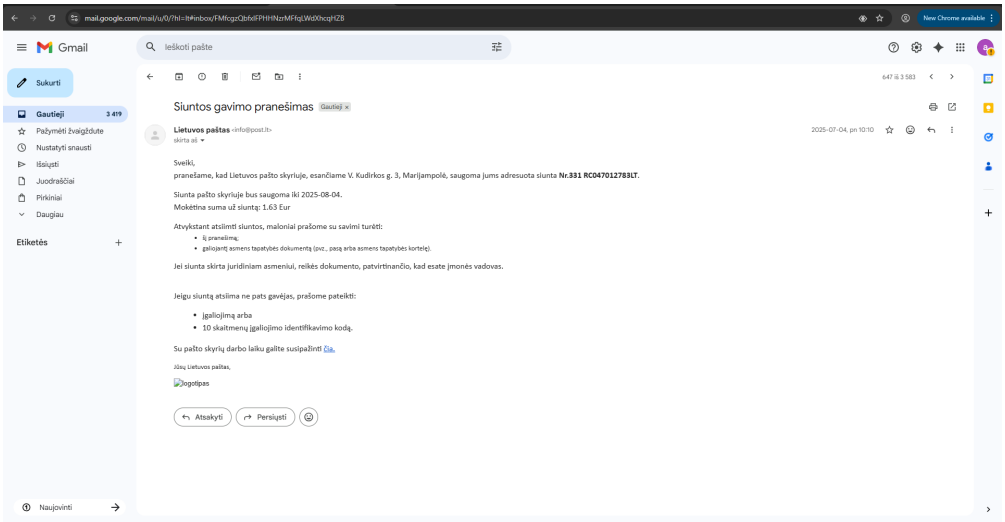
6 pav. 2 scenarijus: Paskyros atkūrimo patvirtinimas

3 scenarijus: Specialūs pasiūlymai ir kodai. Naudotojas gauna intriguojantį laišką su „išskirtinėmis nuolaidomis" ir mygtuku „Patvirtinti dabar" (7 pav.). Laiške gausu skubos, autoriteto ir bauginimo elementų. Nuoroda veda į puslapį su forma, renkančia SMS kodus arba banko duomenis. Testuojamos taisyklės: G1 (psichologiniai signalai), G5 (formų apsauga), G8 (aiškinamieji pranešimai).



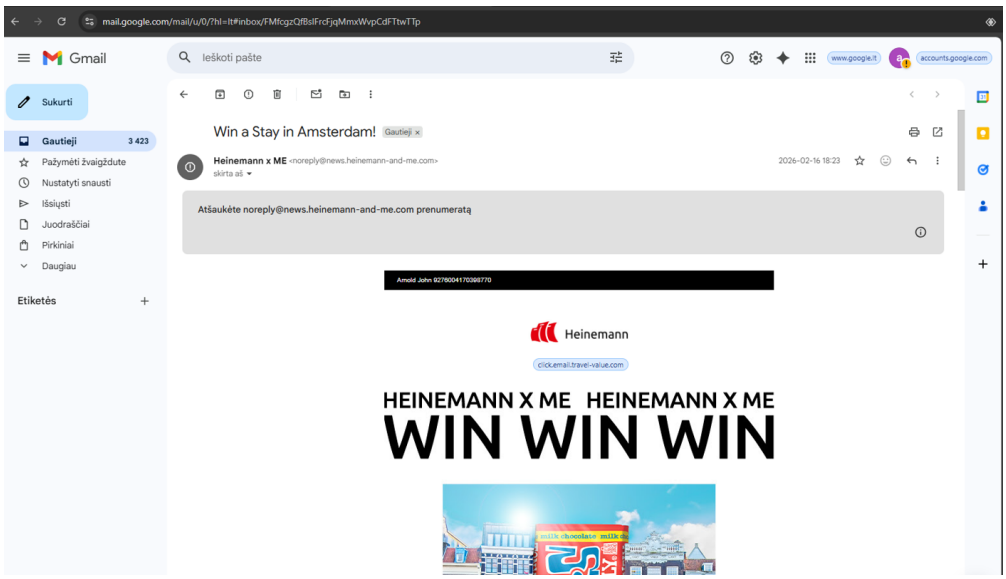
7 pav. 3 scenarijus: Specialūs pasiūlymai ir kodai

4 scenarijus: Pristatymo pranešimai. Naudotojas gauna el. laišką apie siuntą, kuri „nebuvo pristatyta nurodytu adresu“ (8 pav.). Pranešime nurodoma, kad „jei veiksmas nebus atliktas, siunta bus grąžinta siuntėjui“ ir pateikiamas mygtukas su prašymu patvirtinti pristatymo adresą. Testuojamos taisyklės: G4 (SSL sertifikatų patikra), G5 (aktyvi apsauga), G7 (nuorodų grandinės sekimas).



8 pav. 4 scenarijus: Pristatymo pranešimai

5 scenarijus: Loterijų laimėjimai. Naudotojas gauna laišką su žinia „Jūs laimėjote! Spauskite čia, kad atsiimtumėte prizą“ (9 pav.). Laiškas naudoja spalvingus elementus ir paveikslėlius, imituojančius oficialias loterijas. Nuoroda veda į puslapį, kuriame reikalaujama daug asmeninių duomenų (vardas, pavardė, el. paštas, telefono numeris, adresas). Testuojamos taisyklės: G1 (psichologinis manipuliavimas), G5 (jautrūs duomenys), G13 (rizikos indikatorius).



9 pav. 5 scenarijus: Loterijos dalyvavimas

3.4. Testavimo rezultatai

3.4.1. Bendri rezultatai pagal fazę

Žemiau esančioje lentelėje (12 lentelė) parodomi bendri 10 dalyvių rezultatai, sugrupuoti pagal testavimo fazę:

12 lentelė. Testavimo rezultatai pagal fazę (vidurkiai iš 10 dalyvių, apimančių po 10 užduočių per fazę; tikslumas matuojamas kaip teisingai identifikuotų atakų procentas iš 100 užduočių)

Etapas	Tikslumo vidurkis (%)	Vidutinis sprendimo laikas (s)	Pasitikėjimo lygis (1-5)
Kontrolinis (be plėtinio)	58.0	58.4	3.1
Mokymasis su plėtiniu	86.0	34.2	4.4
Išmoktų pamokų taikymas	76.0	39.7	4.1

Įžvalgos iš šios lentelės:

- Tiesioginė plėtinio nauda: mokymosi fazėje atpažinimo tikslumas padidėjo nuo 58.0 % iki 86.0 %, t. y. 28.0 %. Tai rodo, kad plėtinys suteikia reikšmingą tiesioginę asmeninę apsaugą.
- Mokymasis ir žinių išlaikymas: dalyvių rezultatai išmoktų pamokų fazėje (76.0 %) taip pat viršijo kontrolinę fazę 18.0 %, nors ir buvo žemesni nei mokymosi fazėje. Tai rodo, kad naudotojai sėkmingai įsisavino dalį informacijos, pateiktos plėtinio naudojimo metu.
- Sprendimo laikas: mokymosi fazėje sprendimo laikas sutrumpėjo nuo 58.4 s iki 34.2 s, kas rodo, jog naudotojai greitai išmoko pasinaudoti plėtinio teikiama informacija. Po mokymosi etapo, net ir nenaudojant plėtinio, dalyviai sprendimus priimdavo greičiau (39.7 s) nei pradiniame vertinime.
- Pasitikėjimo augimas: naudotojai žymiai labiau pasitikėjo savo sprendimais turėdami plėtinį (4.4 iš 5), tačiau išmoktomis žiniomis grįsti sprendimai taip pat išlaikė aukštą pasitikėjimo lygį (4.1 iš 5).

3.4.2. Rezultatai pagal dalyvių kategorijas

Žemiau pateikiami fazių rezultatai (13 lentelė), suskirstyti pagal naudotojų IT kompetenciją:

13 lentelė. Tikslumo rezultatai pagal dalyvio tipą ir fazę

Naudotojo tipas	Kontrolinis (%)	Su plėtiniu (%)	Išmoktų pamokų taikymas (%)
IT specialistai (n=4)	75.0	95.0	85.0
Kasdieniai naudotojai (n=3)	60.0	86.7	80.0
Nereguliarūs naudotojai (n=3)	33.3	73.3	60.0

Įžvalgos pagal dalyvio tipą:

- IT specialistai: pradiniame kontroliniame etape jie pasiekė 75.0 % tikslumą dėl savo praktinės patirties saugumo srityje. Naudodami plėtinį jie pasiekė 95.0 % atpažinimo rodiklį, ir net pašalinus plėtinį jie išlaikė 85.0 % atsparumą atakoms.
- Kasdieniai naudotojai: šioje grupėje kontrolinis tikslumas buvo vidutinis (60.0 %), tačiau plėtinio pagalba jis padidėjo iki 86.7 % — užregistruojamas reikšmingas 26.7 % šuolis. Po mokymosi ši grupė išlaikė 80.0 % tikslumą.
- Nereguliarūs naudotojai: šios grupės pradinis tikslumas buvo žemiausias (33.3 %), bet naudojant plėtinį jis pakilo iki 73.3 %, kas sudaro 40.0 % augimą. Po mokymosi etapo jie išlaikė 60.0 % atpažinimo rodiklį, ir tai vis tiek reiškia svarbų pažangos žingsnį nuo bazinės stadijos.
- Išvada: plėtinys buvo naudingas visoms naudotojų grupėms, tačiau didžiausią poveikį padarė asmenims, turintiems mažiausiai žinių. Šis atradimas patvirtina naršyklės plėtinio, kaip mokomosios priemonės, vertę.

3.4.3. Rezultatai pagal scenarijaus tipą

Kiekvienas iš 5 scenarijų buvo vertintas pagal tai, kaip dažnai naudotojai sugebėdavo teisingai identifikuoti grėsmę (14 lentelė):

14 lentelė. Tikslumo rezultatai pagal scenarijaus tipą

Scenarijus	Kontrolinis (%)	Su plėtiniu (%)	Mokymosi taikymas (%)
1. Prisijungimo patvirtinimas	60	90	80
2. Paskyros atkūrimo patvirtinimas	50	90	70
3. Specialūs pasiūlymai ir kodai	50	80	70
4. Pristatymo pranešimai	60	90	80
5. Loterijų laimėjimai	70	80	80
Vidurkis	**58**	**86**	**76**

Įžvalgos pagal scenarijus:

- Lengviausios atakos: pirmas scenarijus „Prisijungimo patvirtinimas“ buvo lengviausiai atpažįstamas — net kontroliniame etape jį nustatė 62 % naudotojų. Tai greičiausiai lemia

tai, kad tokio tipo ataka naudoja standartines paskyros prisijungimą imituojančias formas, su kuriomis naudotojai jau yra susidūrę viešojoje erdvėje.

- Sudėtingiausios atakos: antrasis scenarijus „Paskyros atkūrimas“ ir trečiasis „Specialūs pasiūlymai“ pasirodė esantys sudėtingesni — pradiniame etape juos identifikavo tik 50 % dalyvių. Net ir panaudojus plėtinį, atpažinimo rodikliai padidėjo (iki 90 % ir 80 %), bet išliko iššūkių taikant išmoktas žinias be plėtinio (70 %).
- Plėtinio poveikis: visuose scenarijuose plėtinys teikė reikšmingą naudą, padidindamas atpažinimo rodiklį vidutiniškai 28 %. Vis dėlto, sudėtingesnės manipuliacijos be ryškių vizualių trūkumų naudotojams be plėtinio vis tiek kėlė sumaištį.
- Žinių išlaikymas: po mokymosi naudotojai neturėdami plėtinio rodė prastesnius rezultatus nei atlikdami užduotis su plėtiniu. Tai rodo ribotą ilgalaikę žinių išlaikymo galią ir patvirtina nuolatinio mokymosi ir įrankių naudojimo būtinybę.

3.4.4. Projektavimo gairių veiksmingumas

Žemiau pateikiami duomenys apie tai (15 lentelė), kurias projektavimo gaires (G1-G33) naudotojai geriausiai suprato ir išskyrė kaip naudingiausias sprendžiant užduotis (vertinimas balo sistemoje nuo 1 iki 5):

15 lentelė. Projektavimo gairių supratimas ir vertinimas naudotojų grupėje

Gairė (kodas)	Supratimas (%)	Naudingumas naudotojams (1-5)
G2 — URL analizė	90	4.7
G3 — Turinio analizė	80	4.2
G4 — SSL sertifikatai	90	4.5
G5 — Aktyvi apsauga	80	4.1
G8 — Aiškinamieji pranešimai	100	4.8
G11 — Domeno išryškinimas	90	4.6
G13 — Rizikos indikatorius	90	4.4
G14 — Patikimumo ženklai	80	4.3
G15 — Mokymas tinkamu momentu	90	4.6
G20 — Priverstinis patvirtinimas	90	4.5
G32 — Įspėjimų istorija	80	3.9
G33 — Grįžtamasis ryšys	70	3.7

Įžvalgos iš gairių veiksmingumo:

- Labiausiai naudingos gairės: G8 (Aiškinamieji pranešimai) ir G2 (URL analizė) buvo pripažintos naudingiausiomis, įvertintos atitinkamai 4.8 ir 4.7 balais. Tai rodo, jog tiesioginiai ir aiškūs paaiškinimai apie tai, kodėl ataka yra pavojinga, yra kritiškai svarbūs apsaugos komponentai.
- Mažiau suprastos gairės: G33 (Grįžtamasis ryšys) ir G32 (Įspėjimų istorija) pasižymėjo žemesniais supratimo bei naudingumo rodikliais (atitinkamai 70 % ir 80 %), nors iš esmės

buvo vertinamos teigiamai. Tai sufleruoja, kad naudotojai rečiau atkreipia dėmesį į šias funkcijas arba mažiau supranta jų paskirtį.

- Mokymo reikšmė: G15 (Mokymas tinkamu momentu) gavo 4.6 naudingumo balo įvertinimą ir smarkiai prisidėjo prie rezultatų antrame etape. Tai patvirtina kontekstinio mokymo integravimo svarbą nuo pat pradžių.
- Kombinuotas veiksmingumas: plėtinys aukštą veiksmingumą pasiekė būtent dėl keturių pagrindinių elementų kombinacijos: URL analizės, aiškinamųjų pranešimų, vizualinių indikatorių ir integruoto mokymo.

3.4.5. Naudotojų grįžtamasis ryšys

Be kiekybinių duomenų gautas ir kokybinis dalyvių vertinimas:

- Teigiami komentarai:
  - 9 iš 10 dalyvių nurodė, kad spalvų sistema (žalia, geltona, raudona) buvo „labai aiški“ ir intuityvi;
  - 8 iš 10 teigiamai įvertino mokomuosius pranešimus — pastebėjo, jog išmoko „ne tik tai, ko nedaryti, bet ir kodėl“;
  - 7 iš 10 tvirtino, kad plėtinys buvo „lengvai naudojamas“ ir „netrukdė įprastam naršymui“.
- Neigiami komentarai:
  - 4 dalyviai atkreipė dėmesį į per didelį informacijos kiekį šoninėje sekcijoje — jie pageidautų detalių sumažinimo arba galimybės naudoti „kompaktišką“ režimą;
  - 3 naudotojai išreiškė nuomonę, kad mokomieji pranešimai kartais atrodė pernelyg sudėtingi ir pripildyti techninių terminų, ypač nereguliariems naudotojams;
  - 2 dalyviai pasigedo greitesnio būdo įtraukti saugius domenus į patikimų domenų sąrašą (angl. *trusted domains*).
- Tolimesni pasiūlymai:
  - Skatinti didesnį įsitraukimą mokymosi metu, pavyzdžiui, naudojant trumpus interaktyvius testus;
  - Suteikti galimybę naudotojams stebėti savo identifikavimo pasiekimus ar suvestinę statistiką;
  - Diegti tiesioginę integraciją su populiariomis elektroninio pašto paslaugomis.

Rezultatai ir išvados

Atlikus darbe numatytus uždavinius, gauti šie rezultatai:

1. Suformuluoti 24 apsaugos įrankių kokybės vertinimo kriterijai, apimantys technines, aiškinamąsias, mokymo, psichologines ir integracines savybes, remiantis atlikta mokslinės literatūros analize apie duomenų viliojimo atakas, naudotojų mokymą ir sąsajų dizaino principus.
2. Nustatyta, kurie apsaugos būdai yra sėkmingai įgyvendinami paslaugose ir kur yra esminės spragos, atlikus 12 esamų apsaugos įrankių lyginamąją analizę pagal suformuluotus kriterijus.
3. Ištyrus galimus projektinius sprendimus, buvo sukurti ir testuoti maketai, vienas grindžiamas proaktyviu blokavimu (A), kitas — kontekstiniais įspėjimais (B). Gautos naudotojų įžvalgos, jų pagrindu pasirinktas hibridinis būdas, adaptuojantis apsaugos intensyvumą pagal grėsmės lygį.
4. Sukurtas naudotojo sąsajų, apsaugančių nuo duomenų viliojimo, projektavimo metodas, apimantis 33 projektavimo gaires ir jų taikymo rekomendacijas.
5. Metodo naudojimui demonstruoti ir veiksmingumui vertinti sukurtas *Chrome* naršyklės įskiepis.

Tyrimo metu suformuluotos šios išvados:

1. Išanalizavus mokslinę literatūrą apie duomenų viliojimo atakas, naudotojų mokymą ir sąsajų projektavimo principus, nustatyta, kad apsaugos nuo duomenų viliojimo atakų problema negali būti išspręsta vien techninėmis priemonėmis — būtinas holistinis metodas, integruojantis aiškinamąjį dizainą, atakoms atsparius sąsajos šablonus ir psichologiniais principais grįstą mokymą.
2. Išanalizavus esamas apsaugos paslaugas nustatyta, kad priemonės įgyvendina techninius apsaugos sprendimus, tačiau tik 2 iš 12 iš dalies integruoja aiškinamojo dizaino ar mokymo elementus. Tai atskleidžia esminę spragą — nei viena iš analizuotų paslaugų nederina stipraus techninio aptikimo su aiškinamuoju dizainu ir mokomuoju komponentu, nors būtent šių savybių derinys literatūroje identifikuojamas kaip veiksmingiausias ilgalaikiam naudotojų atsparumui formuoti.
3. Proaktyvaus blokavimo ir kontekstinio įspėjimo principais grindžiamų maketų vertinimas parodė jų privalumus, kurie atsiskleidžia skirtinguose kontekstuose. Aukštos rizikos grėsmėms buvo teikiama pirmenybė intensyviam proaktyviam blokavimui su išsamiais paaiškinimais, mažesnės rizikos atvejams — subtilesnis kontekstinis informavimas. Šios įžvalgos sudaro pagrindą integruoti abu principus vienoje sąsajoje, siekiant užtikrinti apsaugos veiksmingumą ir kasdienio naudojimo patogumą.
4. Nustatyta, kad projektavimo gaires verta skirstyti pagal tikslinę auditoriją ir kuriamo įrankio tipą, nes skirtingoms naudotojų grupėms (skirtingų IT įgūdžių) ir įrankių tipams (naršyklės plėtinys, el. pašto apsauga, organizacinis sprendimas) aktualūs gairių poaibiai

skiriasi. Kartu nustatyta, kad bazinis 12 gairių rinkinys yra būtinas nepriklausomai nuo konteksto.

5. Įvertinus demonstracinį pavyzdį paaiškėjo, kad aiškinamųjų naudotojo sąsajų integravimas su duomenų viliojimui atspariais projektavimo metodais pagerina naudotojų gebėjimą atpažinti atakas ir formuoja savarankiškus atpažinimo įgūdžius, kurie išlieka išjungus pagalbinį įrankį. Nustatyta, jog didžiausią naudą įskiepis teikė mažiausios IT įgūdžių naudotojams, o adaptyvus įspėjimų intensyvumas leido išvengti įspėjimų nuovargio aukštesnių IT įgūdžių naudotojams.

Šaltiniai

[AA24] K. Althobaiti, N. Alsufyani. A review of organization-oriented phishing research. *PeerJ Computer Science*. 2024, tomas 10, e2487. Prieiga per internetą: https://doi.org/10.7717/peerj-cs.2487.

[AAD20] J. Aneke, C. Ardito, G. Desolda. Designing an Intelligent User Interface for Preventing Phishing Attacks. Iš: J. Abdelnour Nocera, A. Parmaxi, M. Winckler, F. Loizides, C. Ardito, G. Bhutkar, P. Dannenmann (sudarytojai). *Beyond Interactions*. Cham: Springer International Publishing, 2020, p. 97–106.

[AHN+21] Z. Alkhalil, C. Hewage, L. Nawaf, I. Khan. Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*. 2021, tomas 3. ISSN 2624-9898. Prieiga per internetą: https://doi.org/10.3389/fcomp.2021.563060.

[Ava25] Avast. *Avast One solution*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://www.avast.com/avast-one.

[BBK+25] C. Beckmann, B. Berens, N. Kühl, P. Mayer, M. Mossano, M. Volkamer. Design and Evaluation of an Anti-phishing Artifact Based on Useful Transparency. Iš: M. Mehrnezhad, S. Parkin (sudarytojai). *Socio-Technical Aspects in Security*. Cham: Springer Nature Switzerland, 2025, p. 113–133. ISBN 978-3-031-83072-3.

[CBJ+14] L. Coventry, P. Briggs, D. Jeske, A. van Moorsel. SCENE: A Structured Means for Creating and Evaluating Behavioral Nudges in a Cyber Security Environment. Iš: 2014. ISBN 978-3-319-07667-6. Prieiga per internetą: https://doi.org/10.1007/978-3-319-07668-3_23.

[DAA+22] G. Desolda, J. Aneke, C. Ardito, R. Lanzilotti, M. Costabile. Explanations in Warning Dialogs to Help Users Defend Against Phishing Attacks. *International Journal of Human-Computer Studies*. 2022, tomas 176.

[DFM+22] G. Desolda, L. Ferro, A. Marrella, M. Costabile, T. Catarci. Human Factors in Phishing Attacks: A Systematic Literature Review. *ACM Computing Surveys*. 2022, tomas 54, p. 35. Prieiga per internetą: https://doi.org/10.1145/3469886.

[F-S25] F-Secure. *F-Secure Total*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://www.f-secure.com/en/total.

[FT18] A. Ferreira, S. Teles. Persuasion: How phishing emails can influence users and bypass security measures. *International Journal of Human-Computer Studies*. 2018, tomas 125. Prieiga per internetą: https://doi.org/10.1016/j.ijhcs.2018.12.004.

[Gar10] A. García Frey. Self-explanatory user interfaces by model-driven engineering. 2010, p. 341–344. ISBN 9781450300834. Prieiga per internetą: https://doi.org/10.1145/1822018.1822076.

[Goo25a] Google. *Advanced Gmail security*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://support.google.com/a/topic/2683828?hl=en.

[Goo25b] Google. *Google Safe Browsing*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://safebrowsing.google.com/.

[GST+18] K. Greene, M. Steves, M. Theofanos, J. Kostick. User Context: An Explanatory Variable in Phishing Susceptibility. Iš: 2018. Prieiga per internetą: https://doi.org/10.14722/usec.2018.23016.

[Hon12] J. Hong. The state of phishing attacks. *Commun. ACM*. 2012, tomas 55, numeris 1, p. 74–81. ISSN 0001-0782. Prieiga per internetą: https://doi.org/10.1145/2063176.2063197.

[KK25] A. Kavvadias, T. Kotsilieris. Understanding the Role of Demographic and Psychological Factors in Users' Susceptibility to Phishing Emails: A Review. *Applied Sciences*. 2025, tomas 15, numeris 4. ISSN 2076-3417. Prieiga per internetą: https://doi.org/10.3390/app15042236.

[KOO22] I. Kara, M. Ok, A. Ozaday. Characteristics of Understanding URLs and Domain Names Features: The Detection of Phishing Websites With Machine Learning Methods. *IEEE Access*. 2022, tomas 10, p. 124420–124428. Prieiga per internetą: https://doi.org/10.1109/ACCESS.2022.3223111.

[KSB+24] J. Klütsch, J. Schwab, C. Böffel, V. Zimmermann, S. Schlittmeier. Friend or phisher: how known senders and fear of missing out affect young adults' phishing susceptibility on social media. *Humanities and Social Sciences Communications*. 2024, tomas 11. Prieiga per internetą: https://doi.org/10.1057/s41599-024-03412-8.

[Loo25] Lookout. *Lookout mobile phishing platform*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://www.lookout.com/.

[Mic25a] Microsoft. *Microsoft Defender for Office 365*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://learn.microsoft.com/en-us/defender-office-365/.

[Mic25b] Microsoft. *Microsoft Defender SmartScreen overview*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://learn.microsoft.com/en-us/windows/security/operating-system-security/virus-and-threat-protection/microsoft-defender-smartscreen/.

[Mim25] Mimecast. *Mimecast security solutions*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://www.mimecast.com.

[MKK08] E. Medvet, E. Kirda, C. Kruegel. Visual-similarity-based phishing detection. 2008. ISBN 9781605582412. Prieiga per internetą: https://doi.org/10.1145/1460877.1460905.

[Moz25] Mozilla. *Firefox phishing and malware protection*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://support.mozilla.org/en-US/kb/how-does-phishing-and-malware-protection-work.

[Mun19] A. Muntode. An Overview on Phishing- its types and Countermeasures. *International Journal of Engineering Research and Technology*. 2019, tomas V8. Prieiga per internetą: https://doi.org/10.17577/IJERTV8IS120260.

[Nai15] R. Naidoo. Analysing urgency and trust cues exploited in phishing scam designs. Iš: 2015.

[Net25] Netcraft. *Netcraft Browser Extension*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://www.netcraft.com/apps/.

[Ope25] Openphish. *Openphish Phishing Database*. 2025. Prieiga per internetą: https://www.openphish.com/phishing_database.html.

[Phi25] PhishTank. *PhishTank Phishing Database*. 2025. Prieiga per internetą: https://phishtank.org/.

[Pro25] Proofpoint. *Proofpoint security solutions*. 2025. [žiūrėta 2025-11-13]. Prieiga per internetą: https://www.proofpoint.com.

[PZS19] J. Petelka, Y. Zou, F. Schaub. Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. Iš: 2019, p. 1–15. ISBN 978-1-4503-5970-2. Prieiga per internetą: https://doi.org/10.1145/3290605.3300748.

[RFC+18] R. W. Reeder, A. P. Felt, S. Consolvo, N. Malkin, C. Thompson, S. Egelman. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. Iš: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC, Canada: Association for Computing Machinery, 2018, p. 1–13. CHI '18. ISBN 9781450356206. Prieiga per internetą: https://doi.org/10.1145/3173574.3174086.

[RG18] P. Rajivan, C. Gonzalez. Creative Persuasion: A Study on Adversarial Behaviors and Strategies in Phishing Attacks. *Frontiers in Psychology*. 2018, tomas Volume 9 - 2018. ISSN 1664-1078. Prieiga per internetą: https://doi.org/10.3389/fpsyg.2018.00135.

[RRJ+20] J. Rastenis, S. Ramanauskaitė, J. Janulevičius, A. Čenys, A. Slotkienė, K. Pakrijauskas. E-mail-Based Phishing Attack Taxonomy. *Applied Sciences*. 2020, tomas 10, numeris 7. ISSN 2076-3417. Prieiga per internetą: https://doi.org/10.3390/app10072363.

[RTN25] S. S. Roy, C. Torres, S. Nilizadeh. *"Explain, Don't Just Warn!" – A Real-Time Framework for Generating Phishing Warnings with Contextual Cues*. 2025. Prieiga per internetą: https://arxiv.org/abs/2505.06836.

[SPD+23] A. Shrestha, R. Paudel, P. Dumaru, M. N. Al-Ameen. Towards Improving the Efficacy of Windows Security Notifier for Apps from Unknown Publishers: The Role of Rhetoric. Iš: 2023, p. 101–121. ISBN 978-3-031-35821-0. Prieiga per internetą: https://doi.org/10.1007/978-3-031-35822-7_8.

[SZN+21] K. Sharma, X. Zhan, F. Nah, K. Siau, M. Cheng. Impact of digital nudging on information security behavior: an experimental study on framing and priming in cybersecurity. *Organizational Cybersecurity Journal: Practice, Process and People*. 2021, tomas ahead-of-print. Prieiga per internetą: https://doi.org/10.1108/OCJ-03-2021-0009.

[TCR24] D. Timko, D. H. Castillo, M. L. Rahman. *A Quantitative Study of SMS Phishing Detection*. 2024. Prieiga per internetą: https://arxiv.org/abs/2311.06911.

[VCM18] J. P. Vargheese, M. Collinson, J. Mastho. How can persuasion reduce user cyber security vulnerabilities? Iš: 2018, p. 35–38. Prieiga per internetą: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056904067&partnerID=40&md5=e78950665122f265ca1c0c97f75826f6.

[VKL+22] M. Vössing, N. Kühl, M. Lind, G. Satzger. Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*. 2022, tomas 24, numeris 3, p. 877–895. ISSN 1387-3326. Prieiga per internetą: https://doi.org/10.1007/s10796-022-10284-3.

[WMG06] M. Wu, R. C. Miller, S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? Iš: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Montréal, Québec, Canada: Association for Computing Machinery, 2006, p. 601–610. CHI '06. ISBN 1595933727. Prieiga per internetą: https://doi.org/10.1145/1124772.1124863.

[WSW+22] J. Wang, J. Shi, X. Wen, L. Xu, K. Zhao, F. Tao, W. Zhao, X. Qian. The Effect of Signal Icon and Persuasion Strategy on Warning Design in Online Fraud. *Computers & Security*. 2022, tomas 121, p. 102839. Prieiga per internetą: https://doi.org/10.1016/j.cose.2022.102839.

[XPY+17] A. Xiong, R. Proctor, W. Yang, N. Li. Is Domain Highlighting Actually Helpful in Identifying Phishing Web Pages? *Human Factors*. 2017, tomas 0, p. 001872081668406. Prieiga per internetą: https://doi.org/10.1177/0018720816684064.