Süni İntellekt Əsaslı Kiberhücumların Aşkarlanması və Rəqəmsal Kriminalistika Problemləri

Bəhruz Əliyev
Azərbaycan Kibertəhlükəsizlik Təşkilatları Assosiasiyası, Bakı Dövlət Universiteti, Bakı, Azərbaycan
bahruz.work@gmail.com

***Xülasə*— Məqalədə süni intellekt əsaslı kiberhücumların mahiyyəti, növləri və onların aşkarlanması ilə bağlı müasir yanaşmalar araşdırılır. Xüsusilə süni intellektdən istifadə etməklə həyata keçirilən avtomatlaşdırılmış hücumlar, deepfake texnologiyaları və sosial mühəndislik hücumlarının yeni formaları təhlil edilir. Bu cür hücumların ənənəvi təhlükəsizlik sistemlərini keçməsi və onların aşkarlanmasında süni intellekt əsaslı müdafiə mexanizmlərinin rolu izah olunur. Eyni zamanda, rəqəmsal kriminalistika çərçivəsində bu hücumların izlərinin aşkarlanması, sübutların toplanması və analizi proseslərində qarşıya çıxan çətinliklər araşdırılır.**

Açar sözlər— süni intellekt; zərərli proqram; fişinq; anomaliya; davranış; təhlükəsizlik; kriminalistika; sübutlar; insident.

I. GİRİŞ

Rəqəmsal texnologiyaların cəmiyyətin bütün sahələrinə sürətlə nüfuz etdiyi bir dövrdə kibertəhlükəsizlik məsələsi dövlət, biznes və şəxsi həyatın strateji prioritetinə çevrilmişdir. Bir zamanlar yalnız ixtisaslı hakerlərə məxsus olan hücum texnikaları bu gün süni intellekt (Sİ) vasitəsilə kütləviləşmiş, avtomatlaşdırılmış və son dərəcə adaptiv xarakter almışdır [1]. Hücumçular artıq böyük texniki bilik olmadan mürəkkəb kiberhücumlar həyata keçirə bilir. Bunu mümkün edən isə məhz Sİ texnologiyalarıdır.

Vəziyyəti daha da mürəkkəbləşdirən amil budur ki, süni intellekt həm hücum, həm də müdafiə alətinə çevrilmişdir. Bu "ikili təyinatlı istifadə" (dual-use) xarakteri kibertəhlükəsizlik idarəetməsini prinsipial yeni problemlər qarşısında qoyur [2]. Belə bir mühitdə ənənəvi müdafiə mexanizmləri: imza əsaslı antiviruslar, statik qaydalar toplusu, əl ilə aparılan monitorinq, artıq effektiv cavab vermək iqtidarında deyil [3].

Bu tədqiqatın məqsədi Sİ-dən istifadə edilməklə həyata keçirilən kiberhücumların mahiyyətini, növlərini və onların ənənəvi sistemləri necə aşdığını araşdırmaq, bu hücumların aşkarlanmasında Sİ əsaslı müdafiə mexanizmlərinin rolunu təhlil etmək, rəqəmsal kriminalistika çərçivəsində sübutların toplanması prosesini, bu sahədə mövcud çətinlikləri, habelə etik və hüquqi məsələləri sistemli şəkildə nəzərdən keçirməkdir.

II. SÜNİ İNTELLEKT ƏSASLI KİBERHÜCUMLARIN NÖVLƏRİ VƏ MEXANİZMLƏRİ

2.1 Avtomatlaşdırılmış hücumlar

Ənənəvi kiberhücumlar böyük ölçüdə əl ilə idarə olunurdu. Hücumçu hər addımı şəxsən planlaşdırır, icra edir və nəzarət edirdi. Sİ-nin tətbiqi bu prosesi kökündən dəyişdirmişdir. Müasir avtomatlaşdırılmış hücum sistemləri hədəfi avtomatik müəyyən edir, zəif nöqtələri müəyyən edir, istismar üsulunu seçir və hücumu həyata keçirir [1].

Ağıllı port skanı və kəşfiyyat: Sİ dəstəkli kəşfiyyat alətləri şəbəkə topologiyasını öyrənir, müdafiə sistemlərinin davranış nümunələrini analiz edir və aşkarlanmadan yayınmaq üçün hücum vaxtını, intensivliyini optimallaşdırır. Shodan və Censys kimi açıq kəşfiyyat platformalarından əldə edilən məlumatlar Sİ modellərindən keçirilərək hücum üçün ən perspektivli hədəflər müəyyənləşdirilir [2].

Kütləvi fərdiləşdirilmiş hücumlar: Ənənəvi kütləvi hücumlarda bütün hədəflərə eyni mesaj göndərilirdi. Sİ-nin gücü hər hədəf üçün fərdiləşdirilmiş hücum ssenariləri hazırlamağa imkan verir. Böyük Dil Modelləri (LLM) hər bir qurbanın sosial media profilini, peşə keçmişini, maraq sahələrini analiz edərək yüksək inandırıcılıqlı, fərdiləşdirilmiş fişinq məzmunu yaradır [4].

2.2 Adaptiv zərərli proqramlar (Adaptive Malware)

Müasir zərərli proqramların ən təhlükəli xüsusiyyəti mühitə uyğunlaşma qabiliyyətidir. Ənənəvi imza əsaslı antiviruslar zərərli proqramı onun xarakterik bayt ardıcıllığı, imzası vasitəsilə tanıyırdı. Adaptiv zərərli proqramlar bu imzasını dinamik olaraq dəyişdirərək aşkarlanmadan yayınır [5].

Polimorfik kod: Hər icra edilmə zamanı şifrələmə açarı dəyişir, lakin əsas funksionallıq qorunur. Metamorfik kodda isə daha irəli gedilir. Proqramın strukturu, assembly instruksiyaları, dəyişən adları tam dəyişdirilir. Sİ bu prosesi avtomatlaşdırır və sürətləndirir [5].

Mühit-şüurlu zərərli proqramlar: Sİ əsaslı zərərli proqramlar izolyasiya olunmuş sınaq mühitini (sandbox) müəyyənləşdirərək orada zərərsiz davranış nümayiş etdirir, real sistemdə isə tam funksionallığını ortaya qoyur. Bu qabiliyyət ənənəvi analizi əhəmiyyətli dərəcədə çətinləşdirir [3].

LLM əsaslı zərərli kod generasiyası: Böyük Dil Modelləri istifadəçinin insan dilli təsviri əsasında funksional proqram kodu yaza bilir. Hücumçular bu qabiliyyəti xüsusi, aşkarlanması çətin zərərli proqramların sürətlə hazırlanması üçün istifadə edirlər. Tədqiqat göstərdi ki, GitHub, Copilot kimi alətlər istifadəçinin niyyətindən asılı olmayaraq potensial zərərli kod fraqmentləri yarada bilir [6].

2.3 Deepfake texnologiyaları ilə hücumlar

Deepfake texnologiyası dərin öyrənmə alqoritmlərindən istifadə etməklə real görünüşlü süni media kontentinin (video, audio, şəkil) yaradılması texnologiyasıdır. Bu texnologiya kiberhücumların yeni kateqoriyasını formalaşdırmışdır [7].

DOI: 10.25045/forensic.gov.2026.12

Səs klonlaması (Voice cloning): Yalnız bir neçə dəqiqəlik audio nümunəsindən istifadə etməklə hər hansı bir şəxsin səsini klonlamaq hazırda texniki baxımdan mümkün hesab olunur. Bu texnologiya "vishing" (voice phishing) hücumlarında istifadə edilir. Hücumçu qurbanın yaxınlarının səsini simulyasiya edərək ciddi maliyyə əməliyyatları həyata keçirmək üçün razılıq alır [7].

Video deepfake: Yüksək keyfiyyətli saxta video indi istehlakçı səviyyəsindəki avadanlıqla yaradıla bilər. 2024-cü ildə Honq Konq şirkətinin maliyyəçisi deepfake videosu əsasında keçirilmiş videokonfrans zamanı 25 milyon dollar köçürmüş və bu, sənədləşdirilmiş ən böyük deepfake fırıldaqlarından biridir [8].

Sənəd saxtalaşdırması: Sİ üz şəkillərini, imzaları, sənəd möhürlərini ultra-realist şəkildə reproduksiya edə bilir. Biometrik doğrulama sistemlərinin deepfake-ə qarşı müqaviməti hazırda aktiv tədqiqat sahəsidir.

Multimod manipulyasiya: Ən müasir hücumlar eyni zamanda video, audio və mətn saxtasını birləşdirir. Belə kombinasiya inandırıcılığı əhəmiyyətli dərəcədə artırır [9].

2.4 Sİ dəstəkli sosial mühəndislik hücumları

Sosial mühəndislik insan psixologiyasını manipulyasiya edərək texniki boşluqlar olmadan da sistemlərə giriş əldə etmə texnikasıdır. Sİ ilə birləşdikdə tamamilə yeni keyfiyyət kəsb edir [1].

Hiperfərdiləşdirilmiş fişinq (Spear phishing): LLM-lər hədəf şəxs haqqında toplanmış açıq məlumatları (LinkedIn profili, Twitter paylaşımları, şirkət veb-saytı, patent sənədləri) analiz edərək son dərəcə inandırıcı, kontekstual cəhətdən dəqiq fişinq e-poçtları hazırlayır. Bu e-poçtlar dilçilik xüsusiyyətləri, ton və üslub baxımından qurbanın tanıdığı şəxsin yazı tərzini imitasiya edir [4].

Dinamik sosial mühəndislik: Sİ agentləri real-vaxt rejimində qurbanla dialoq aparır, onun suallarına inandırıcı cavablar verir, etirazları aradan qaldırır.

OSINT avtomatlaşdırması: Açıq mənbə kəşfiyyatı (Open Source Intelligence) prosesi Sİ vasitəsilə sürətləndirilir. Hücumçular saniyələr ərzində hədəf şəxs haqqında böyük həcmdə kontekstual məlumat toplayır ki, bu da daha inandırıcı sosial mühəndislik ssenarilərinin hazırlanmasını mümkün edir [2].

İnkar edilə bilən (Plausible deniability): Sİ tərəfindən generasiya edilmiş məzmunun müəllifi insan müəllifi ilə fərqlənmir, bu isə hücum faktını sübut etməyi son dərəcə çətinləşdirir [10].

III. Sİ ƏSASLI MÜDAFIƏ MEXANIZMLƏRI

3.1 Anomaliya aşkarlama

Anomaliya aşkarlama sistemin "normal" davranışını müəyyənləşdirərək, ondan kənarlaşmaları aşkarlama yanaşması, Sİ əsaslı müdafiənin özəyini təşkil edir [3]. Bu yanaşma imza əsaslı sistemlərdən fərqli olaraq bilinməyən hücum növlərini də aşkarlaya bilir.

Şəbəkə davranışı analizi: Şəbəkə trafikinin statistik xüsusiyyətləri paket ölçüsü paylanması, protokol, bağlantı müddəti, port istifadə nümunələri daim monitorinq edilir.

İstifadəçi davranışı analitikası: Hər istifadəçi üçün normal davranış profili yaradılır. Hansı vaxtda sistemə daxil olur, hansı fayllara müraciət edir, hansı əmrləri icra edir. Bu profildən kənarlaşmalar məsələn, gecə yarısında server məlumatlarına toplu müraciət dərhal xəbərdarlıq yaradır. Daxili təhlükə (insider threat) ssenarilərinin aşkarlanmasında bu yanaşma xüsusilə effektivdir [11].

Son istifadəçi qurğularının telemetriyası: Hər bir iş stansiyasından toplanan telemetriya məlumatları proses iyerarxiyası, fayl sistemi dəyişiklikləri, şəbəkə bağlantıları, sistem reyestri dəyişiklikləri Sİ modeli tərəfindən real-vaxt rejimində analiz edilir [3].

3.2 Deepfake aşkarlama texnikaları

Deepfake aşkarlaması aktiv tədqiqat sahəsinə çevrilmiş, bir sıra effektiv metodlar işlənib hazırlanmışdır [9].

Biometrik ardıcıllıq analizi: Real insan videosunda göz qırpma tezliyi, üz əzələlərinin hərəkəti, baş tərpənişi fizioloji qanunlara tabedir. Deepfake generasiya modelləri bu ardıcıllığı tam imitasiya edə bilmir.

Spektral analiz: Audio deepfake-lər real insan nitqindən fərqlənən frekans spektri xüsusiyyətlərinə malikdir. MFCC (Mel-Frequency Cepstral Coefficients) əsaslı modellər bu fərqi aşkarlayır [9].

Kriptoqrafik yoxlama: Media autentikasiyası üçün C2PA (Content Credentials) standartı işlənib, kamera tərəfindən yaradılan hər bir media faylına rəqəmsal imza əlavə edilir. Sonradan media manipulyasiya edildikdə bu imza pozulur [7].

Kontekstual tutarsızlıq: Deepfake video arxafonda bir sıra fiziki tutarsızlıqlar buraxır. İşıqlandırma istiqaməti, kölgə bucağı, səthin əksi fiziki cəhətdən mümkün olmayan konfiqurasiyalar yaradır. Fizika əsaslı sinir şəbəkələri bu tutarsızlıqları aşkarlayır [9].

3.3 Fişinq Aşkarlama

Sİ əsaslı fişinq aşkarlama sistemləri bir neçə paralel analiz həyata keçirir. Təbii dil emalı modelləri (NLP modelləri) e-poçt mətnini semantik cəhətdən analiz edərək manipulyasiya niyyətinin işarələrini təcililik hissi yaratma, avtoritet imitasiyası, qeyri-adi müraciətləri aşkarlayır [4]. URL nüfuzunun analizi domenin yaş tarixi, WHOIS məlumatları, SSL sertifikatı xüsusiyyətləri, homoglyph (vizual oxşar simvol) hücumları əsasında fişinq URL-lərini müəyyənləşdirir. Vizual oxşarlıq analizi isə brend imitasiyasını aşkarlamaq üçün şübhəli veb-saytların loqo, rəng sxemi və səhifə quruluşunu legitim saytlarla müqayisə edir [3].

3.4 Adaptiv müdafiə (Moving Target Defense)

Moving Target Defense (MTD) konsepsiyası hücumçuya daim dəyişən hücum səthi (attack surface) təqdim edir. Sİ bu dəyişiklikləri dinamik olaraq idarə edir: şəbəkə konfiqurasiyasını, IP ünvanlarını, port nömrələrini, protokol tənzimləmələrini müntəzəm intervallarla yeniləyir. Hücumçunun əldə etdiyi kəşfiyyat məlumatı tez bir zamanda köhnəlir, bu da hücum planlamasını çətinləşdirir [2].

IV. RƏQƏMSAL KRİMİNALİSTİKA: Sİ HÜCUMLARININ İZLƏRİNİN AŞKARLANMASI

4.1 Rəqəmsal kriminalistikanın əsas prinsipləri

Rəqəmsal kriminalistika rəqəmsal sübutların elmi əsaslarla toplanması, qorunması, analizi və məhkəmədə təqdim edilməsi prosesidir [12]. Bu sahənin dörd əsas prinsipi mövcuddur:

- Bütövlük (Integrity): Toplanmış sübutun dəyişdirilmədiyini kriptoqrafik hash funksiyaları (SHA-256) vasitəsilə sübut etmək.
- Autentiklik (Authenticity): Sübutun həqiqətən iddia edilən mənbədən gəldiyini müəyyənləşdirmək.
- Tam olmaq (Completeness): Sübutun bütün əlaqədar hissələrini əhatə etmək.
- Qəbul olunma (Admissibility): Sübutun məhkəmə qaydalarına uyğun toplanmasını təmin etmək.
- NIST SP 800-86 [12] standartı bu prinsiplər əsasında dörd mərhələli kriminalistika prosesini müəyyənləşdirir: toplayıcılıq, araşdırma, analiz, hesabat.

4.2 Sİ Hücumlarının izlərinin aşkarlanmasında xüsusi çətinliklər

Şifrələnmiş trafik: Müasir hücumların böyük əksəriyyəti HTTPS, TLS şifrələməsindən istifadə edir. Trafik məzmununu açmadan analiz aparmaq üçün metadata (paket ölçüsü, zaman möhürü, müddəti) əsaslı maşın öyrənməsi tətbiq edilir. Bu "şifrəli trafik analizi" (Encrypted Traffic Analysis) kimi tanınır [3].

Müvəqqəti sübutlar: RAM yaddaşı, proses cədvəli, şəbəkə bağlantıları sistem söndürüldükdə itir. Aktiv insidentlərdə müvəqqəti operativ yaddaşın canlı əldə olunması (live memory acquisition) prioritet əməliyyat kimi həyata keçirilməlidir [12].

Loq silinməsi: Mürəkkəb hücumçular sistem loqlarını silərək izlərini gizlədərlər. Sİ əsaslı anomaliya aşkarlama sistemləri loqlardakı boşluqların qeyri-adi loq intervallarının özünün bir hücum əlaməti olduğunu müəyyənləşdirə bilir [11].

Çoxyönlü yurisdiksiya: Bulud xidmətlərinin coğrafi paylanmış xarakteri kriminalistika üçün hüquqi çətinliklər yaradır. Müxtəlif ölkələrin qanunvericiliyi altında olan serverlərdən məlumat əldə etmək üçün beynəlxalq hüquq çərçivə tələb edir [13].

4.3 Sübut zənciri (Chain of Custody) bütövlüyünün qorunması

Rəqəmsal sübutun hüquqi dəyəri onun toplanma metoduna birbaşa bağlıdır [12]. Sübut zəncirinin qorunması aşağıdakı tədbirləri tələb edir: kriptoqrafik hash vasitəsilə sübutun bütövlüyünün ani təsdiq edilməsi; bloçkeyn (blockchain) texnologiyasından istifadə etməklə dəyişdirilməsi mümkün olmayan audit jurnalının aparılması; sübutla bağlı hər əməliyyatın (kim aldı, nə vaxt, harada saxladı) tam sənədləşdirilməsi; orijinal sübutun izolyasiya olunması.

V. ETİK VƏ HÜQUQİ PROBLEMLƏR

5.1 Süni intellektin ikili təyinatlı istifadəsi (Dual-Use Dilemma)

Süni intellektin həm hücum, həm müdafiə alətinə çevrilə bilməsi fundamental etik problem yaradır [1]. Deepfake aşkarlama alətlərini hazırlamaq üçün, əvvəlcə daha inandırıcı deepfake yaratmaq bacarığına sahib olmaq tələb olunur [9]. Zərərli proqram davranışını modelləşdirən müdafiə sistemi eyni zamanda yeni növ zərərli proqramların hazırlanmasına zəmin yaradır. Bu ikili xarakter tədqiqatçılar, hüquq-mühafizə orqanları və siyasət qurucuları üçün çətin tarazlıq problemi yaradır [2].

5.2 Sübutların hüquqi qəbul edilə bilməsi

Sİ əsaslı kriminalistika sübutlarının məhkəmədə qəbul olunması hüquqi boşluqlar yaradır. Elmi metodun yoxlanıla bilməsi, xəta nisbəti tələb oluna bilər. Dərin öyrənmə modelləri "qara qutu" xarakteri daşıdığından bu kriteriyaları tam ödəmək çətin olur [14]. İzah oluna bilən Sİ metodlarının tətbiqi bu problemi qismən həll edir.

5.3 Məxfilik hüququ ilə təhlükəsizliyin tarazlığı

Kibertəhlükəsizlik monitorinqi tez-tez məxfilik hüququ ilə toqquşur. Şəbəkə trafikinin tam analizi istifadəçilərin şəxsi kommunikasiyalarına müdaxiləni ehtiva edə bilər. Avropa İttifaqının GDPR tənzimlənməsi belə məlumat toplamanın qanuni əsaslarını, məqsəd məhdudiyyətini və saxlama müddətini ciddi çərçivəyə alır. Sİ əsaslı kriminalistika alətlərinin tətbiqindən əvvəl məxfilik qiymətləndirilməsi (Privacy Impact Assessment) zəruri addım kimi həyata keçirilməlidir [7].

5.4 Xəta riski

Sİ əsaslı sistemlərin yanlış nəticə verməsi, əslində günahsız olan şəxsi şübhəli kimi göstərməsi, ciddi ədalət problemləri doğura bilər. "False flag" hücumları hücumçunun başqasının texnikasını imitasiya etməsi, yanlış atribusiyaya (mənbənin səhv müəyyən edilməsinə) məqsədyönlü şəkildə rəvac verə bilər [10].

5.5 Sİ silahlanması problemi

Sİ-nin kibersilah kimi inkişaf etdirilməsi beynəlxalq münasibətlər kontekstindəki yeni geosiyasi riskləri ortaya qoyur. Kibersavaş ssenarilərinin tənzimlənməsi üzrə beynəlxalq hüquq normalarının olmaması vəziyyəti daha da mürəkkəbləşdirir [13].

VI. GƏLƏCƏK İNKİŞAF İSTİQAMƏTLƏRİ

Kvant kriptoqrafiyası: Post-kvant kriptoqrafiya standartlarının (NIST PQC standartlaşdırma layihəsi) tətbiqi kvant kompüterləri ilə həyata keçirilə biləcək kriptoqrafik hücumlara qarşı müdafiəni təmin edəcək [12].

Sİ müdafiə sistemləri: Müstəqil qərar qəbul edə bilən Sİ agentlərinin müdafiə infrastrukturuna inteqrasiyası, insident cavabını tamamilə avtomatlaşdırmaq, söndürülmüş sistemləri bərpa etmək, hücuma cavab tədbirləri görmək, gələcəyin kibertəhlükəsizlik sistemlərinin əsas komponentlərindən birini təşkil edəcəkdir [1].

Neyromorfik hesablama: Bioloji beyin arxitekturasını imitasiya edən neyromorfoloq çiplər (Intel Loihi, IBM

TrueNorth) enerji effektivliyini qoruyaraq real-vaxt anomaliya aşkarlaması üçün yeni imkanlar açır [14].

Bundan əlavə, dərin möhkəmləndirici öyrənmə (Deep Reinforcement Learning) metodları kibertəhlükəsizlik sistemlərində adaptiv qərarvermə, resursların dinamik bölüşdürülməsi və müdafiə mexanizmlərinin optimallaşdırılması üçün perspektivli yanaşma kimi qiymətləndirilir [15].

Süni intellektin özünü müdafiəsi: LLM-ləri manipulyasiya (prompt injection, jailbreaking) hücumlarına qarşı möhkəmləndirmək, eyni zamanda Sİ əsaslı kiberhücum alətlərinin inkişafını məhdudlaşdırmaq baxımından kritik istiqamətdir [5].

NƏTİCƏ

Süni intellekt texnologiyaları kiberhücum mənzərəsini köklü şəkildə dəyişdirmişdir. Avtomatlaşdırılmış, adaptiv, fərdiləşdirilmiş və yüksək miqyaslı hücumlar artıq ənənəvi müdafiə mexanizmlərinin funksional həddini aşmışdır. Bu çağırışa layiqli cavab yalnız Sİ-nin özü ilə verilə bilər. Anomaliya aşkarlama, davranış analitikası, deepfake müəyyənləşdirməsi və adaptiv müdafiə bu konsepsiyanın əməli ifadələridir.

Rəqəmsal kriminalistika isə yeni reallıqla üz-üzədir. Şifrələnmiş trafik, müvəqqətilik təşkil edən sübutlar, çoxyönlü yurisdiksiya problemləri, deepfake atribusiyası kimi məsələlər ənənəvi kriminalistika metodologiyasını məhdudlaşdırır. Sİ dəstəkli analiz alətlərinin tətbiqi bu çətinlikləri idarə edilə bilən səviyyəyə gətirir.

Lakin texnologiya məsələnin yalnız bir tərəfidir. Hüquqi çərçivənin müasirləşdirilməsi, etik normların formalaşdırılması, ixtisaslı kadr hazırlığı və beynəlxalq əməkdaşlığın gücləndirilməsi texniki tədbirlərlə paralel şəkildə irəliləməlidir. Sİ əsaslı kibertəhlükəsizliyin gələcəyi yalnız daha ağıllı alqoritmlərdən deyil, həm də daha güclü idarəetmə strukturlarından asılıdır.

ƏDƏBIYYAT

[1] M. Brundage et al., The Malicious Use of Artificial Intelligence:Forecasting, Prevention, and Mitigation. Oxford, U.K.: Future of Humanity Institute, 2018.

[2] R. Chesney and D. K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," California Law Review, vol. 107, no. 6, pp. 1753–1820, 2019.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[4] R. Karanjai, "Targeted phishing campaigns using large language models," arXiv preprint arXiv:2301.09558, 2023.

[5] A. J. Lohn, Hacking AI: A Primer for Policymakers on Machine Learning Cybersecurity. Washington, DC, USA: Center for Security and Emerging Technology (CSET), Georgetown University, 2020.

[6] MITRE ATT&CK Framework, "MITRE ATT&CK Framework v14," MITRE Corporation, 2024. https://attack.mitre.org..

[7] National Institute of Standards and Technology, Guide to Integrating Forensic Techniques into Incident Response, NIST Special Publication 800-86. Gaithersburg, MD, USA: NIST, 2022.

[8] H. Ning, Y. Li, F. Shi, and L. T. Yang, "Heterogeneous edge computing offloading optimization with end-side deep reinforcement learning," IEEE Transactions on Network Science and Engineering, vol. 8, no. 3, pp. 2059–2072, 2021, doi: 10.1109/TNSE.2020.3036644.

[9] H. Pearce, B. Ahmad, B. Tan, B. Dolan-Gavitt, and R. Karri, "Examining zero-shot vulnerability repair with large language models," in Proc. IEEE Symp. Security and Privacy (SP), San Francisco, CA, USA, 2023, pp. 2339–2356, doi: 10.1109/SP46215.2023.10179373.

[10] Reuters, "Hong Kong firm loses $25 million in deepfake video call scam," Reuters Technology, Feb. 4, 2024.

[11] T. Rid and B. Buchanan, "Attributing cyber attacks," Journal of Strategic Studies, vol. 38, no. 1–2, pp. 4–37, 2015, doi: 10.1080/01402390.2014.977382.

[12] M. N. Schmitt, Ed., Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. Cambridge, U.K.: Cambridge University Press, 2017.

[13] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131–148, 2020, doi: 10.1016/j.inffus.2020.07.006.

[14] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[15] A. Zaharia, M. Chow, C. Yuan, A. Tudor, A. Krishnamurthy, I. Stoica, and R. Griffith, "Adversarial machine learning in cybersecurity: A systematic survey," ACM Computing Surveys, vol. 54, no. 6, pp. 1–38, 2022.

Artificial Intelligence-Based Cyberattack Detection and Digital Forensics Challenges

Bahruz Aliyev
Azerbaijan Cybersecurity Organizations Association, Baku, Azerbaijan
Baku State University, Baku, Azerbaijan
bahruz.work@gmail.com

Abstract- This article examines the nature, types, and modern detection approaches of artificial intelligence-based cyberattacks. In particular, automated attacks carried out using artificial intelligence, deepfake technologies, and new forms of social engineering attacks are analyzed. The study explains how such attacks bypass traditional security systems and highlights the role of AI-based defense mechanisms in their detection. At the same time, within the framework of digital forensics, the challenges related to identifying traces of these attacks, collecting evidence, and conducting analysis processes are investigated.

Keywords– artificial intelligence; malware; phishing; anomaly; behavior; security; forensics; evidence; incident.