Universitatea Națională de Știință și Tehnologie
POLITEHNICA București
Facultatea de Automatică și Calculatoare, Departamentul de Calculatoare

Teză de Doctorat

în domeniul Științe inginerești, Calculatoare și Tehnologia Informației

User-centric Cybersecurity Attacks Management in Future Generation Computer Systems

Gestionarea Atacurilor de Securitate Cibernetică centrată pe Utilizator în Sistemele Informatice de Generație Viitoare

REZUMAT

Autor

Drd.ing. Răzvan-Constantin STOLERIU

Coordonator

Prof.dr.ing. Florin POP

2024
București, România

Contents

List of Figures

List of Tables

1 | Introducere

1.1 Definirea Problemei de Cercetare și Obiectivele Tezei

În zilele noastre, pe măsură ce tehnologia a evoluat prin progrese în domeniul cloud computing, ale operațiunilor bancare online, ale rețelelor de socializare și ale proceselor automate, informațiile care sunt stocate, gestionate de către diferite dispozitive IT sau care circulă prin rețele sunt în pericol. Același ritm accelerat în dezvoltarea de noi tactici și tehnici se observă și de partea adversarilor. Ei lansează atacuri cibernetice ce vizează atât utilizatorii obișnuiți și companiile, cât și guvernele pentru a fura date și a cauza daune. Amenințările cibernetice cuprind atacuri împotriva infrastructurilor critice care includ sisteme din domeniul transporturilor, comunicațiilor, apei și energiei electrice. Amenințările de securitate constau, de asemenea, și în operațiunile efectuate de atacatori pentru a frauda și a fura date de la victime [6]. Unele dintre efectele acestor atacuri ar fi întreruperea continuității activității și scăderea încrederii din partea clienților. [2]. Adversarii încearcă să nu lase urme în cadrul stațiilor compromise prin dezactivarea sau ocolirea instrumentelor de securitate existente, eliminând log-urile de securitate și ștergând conturile create și fișierele malware utilizate în timpul incidentului.

Tabelul 1.1 prezintă efectele și consecințele amenințărilor cibernetice moderne împortiva mai multor domenii.

Luând în considerare faptul că în zilele noastre adversarii dezvoltă specii malware sofisticate pentru a evita detecțiile și a împiedica analiza malware, ar trebui implementate strategii de detecție și contramăsuri solide. Dacă acestea nu pot opri complet planul adversarilor, cel puțin îl pot face mai greu de îndeplinit, iar în cele din urmă, e posibil ca aceștia să lase urme în rețea care să conducă la identificarea lor. [4].

Toate strategiile de detecție și contramăsurile pe care le-am propus împreună cu soluția de securitate pe care am proiectat-o și dezvoltat-o, au determinat diferite abordări de securitate cu rezultate specifice, fiecare din ele fiind prezentată în capitolele ce urmează. Rezultatele la care am ajuns se bazează pe întrebările și obiectivele de cercetare stabilite în tabelul 1.2.

Soluțiile noastre de detecție nu și-au propus să construiască un context centrat pe utilizator, ci consideră cazurile general valabile pentru toți utilizatorii.

Principalul scop al tezei reprezintă identificarea tipurilor de atacuri cibernetice moderne și dezvoltarea unor strategii de detecție și soluții de securitate care pot oferi rezultate precise și timp de răspuns rapid la incidente. În acest sens, am început cu un studiu al soluțiilor existente în literatură, iar apoi am testat și propus noi abordări care ar putea aduce inovație și rezultate mai bune. Am comparat sistemele pe care le-am proiectat cu cele care au fost propuse de alți cercetători din domeniu și am subliniat avantajele și îmbunătățirile pe care cele propuse de noi le au.

Am definit obiectivele tezei răspunzând la următoarele întrebări de cercetare:

1. **Care sunt tipurile de atacuri cibernetice moderne și cum ar putea organizațiile**

să se apere împotriva lor? (RQ1)

Am studiat diferite lucrări de cercetare din literatură și bloguri de la furnizori populari de securitate pentru a identifica cele mai răspândite atacuri cibernetice, pentru a determina modul în care sunt operate, care sunt cele mai inovatoare tactici și tehnici. În plus, am studiat principalii vectori de infecție utilizați de adversari pentru a obține acces la sisteme și am clasificat diferite familii de malware care targhetează diverse industrii și sisteme de operare. Am propus strategii moderne de detecție a atacurilor și contramăsuri care pot ajuta atât utilizatorii obișnuiți, cât și cercetătorii să se protejeze, să detecteze sau cel puțin să minimizeze riscul unui atac de succes care i-ar putea viza.

2. **Cum ar putea fi detectate URL-urile scurte? (RQ2)**

   Adresele URL scurte au fost concepute pentru a oferi posibilitatea utilizării unei adrese URL care are mai puține caractere și care pointează către același site precum URL-ul original. Recent, mulți furnizori de securitate au raportat utilizarea adreselor URL scurte în campanii malware. Atacatorii preferă această metodă, deoarece pot ascunde site-ul web rău intenționat, iar victimele nu știu ce se află în spatele său. Noi am propus o soluție combinată care utilizează atât algoritmi de învățare automată (ML), cât și diferite platforme open-source de Threat Intelligence pentru a detecta adrese URL scurte care sunt malware.

3. **Cum ar putea fi detectate URL-urile scurte care sunt malițioase în atacurile de tip smishing? (RQ3)**

   Potrivit unui raport de threat intelligence (TI) [15], phishing-ul a fost cel mai frecvent vector de infecție utilizat pe parcursul anului 2021. Phishing-ul are multe forme, în funcție de modul în care este efectuat. De exemplu, dacă se realizează prin mesaje SMS, atunci se numește smishing (adică sms phishing). Smishing a fost folosit în mai multe campanii pentru livrarea de malware, cum ar fi TeaBot, TangleBot și FluBot. Noi am propus un nou model de tip Cloud-Edge pentru a detecta adresele URL scurte care sunt malițioase în atacurile de tip smishing. Sistemul nostru folosește platforme de TI și algoritmi de Machine Learning.

4. **Cum să creștem eficiența sistemelor de securitate în detecția atacurilor cibernetice? (RQ4)**

   Una dintre soluțiile de securitate utilizate în detecția atacurilor cibernetice sunt sistemele de tip Endpoint Detection and Response (EDR). Acestea sunt sisteme sofisticate de detecție a intruziunilor care încearcă să găsească similarități între evenimentele din sistem și moduri de operare cunoscute ale atacatorilor. Cu toate că acestea promit rezultate satisfăcătoare în eliminarea amenințărilor de securitate cibernetică, totuși ele se confruntă cu următoarele provocări: generează un număr mare de alerte de tip false-positive, veridicitatea alertelor trebuie verificată manual de către un analist de securitate etc. Astfel, am propus o soluție distribuită bazată pe un model de tip Cloud-Edge pentru detecția atacurilor de smishing. Sistemul noastru diminuează prin soluții moderne, cum ar fi platformele de threat intelligence și algoritmii de machine learning, deficiențele sistemelor EDR clasice. În plus, sistemul pe care l-am proiectat integrează o soluție ChatBot care facilitează utilizarea și crește gradul de conștientizare a utilizatorilor. Acest proiect pe care l-am propus a făcut parte dintr-un program de tip *Proof-of-Concept* și am lucrat la elaborarea unui plan de afaceri (programul *Invest National University of Science and Technology POLITEHNICA Bucharest Proof of Concept*).

5. **Cum am putea să detectăm cyberbullying-ul în cadrul fișierelor multimedia? (RQ5)**

O formă de atac cibernetic observată recent este cyberbullying-ul. Acesta apare atunci când o persoană amenință, jignește sau îngrozește pe cineva folosind mijloace electronice. Acest comportament rău intenționat are loc de obicei în școli, universități sau acasă. El se realizează de obicei prin intermediul rețelelor de socializare (de exemplu, TikTok) unde oamenii trimit sau partajează videoclipuri sau prin intermediul fișierelor GIF. Noi am propus două soluții diferite pentru detecția cyberbullying-ului. Unul vizează fișierele GIF, în timp ce celălalt se ocupă de videoclipurile de pe TikTok.

6. **Care sunt principalele riscuri de securitate și potențialele contramăsuri pentru atacurile ce vizează infrastructurilor de scară largă? (RQ6)**

   Transportul, educația, sănătatea, guvernarea, infrastructura și multe alte domenii dintr-o infrastructură pe scară largă au trecut prin procesul de tehnologizare. În consecință, există o mulțime de programe, aplicații și procese automate care sunt critice pentru continuitatea activității și furnizarea de servicii. Importanța și capacitatea lor de utilizare pentru viața de zi cu zi a oamenilor i-au determinat pe atacatori să considere aceste domenii drept potențiale ținte pentru a face mai mult profit. Astfel, noi am propus un model de evaluare a riscului de securitate cibernetică în infrastructurile de scară largă, precum și tehnici și contramăsuri de protecție. Metodologia pe care am folosit-o în studiul nostru este calitativă și se bazează pe proiecte din baza de date CORDIS a Comisiei Europene care se referă la guvernanța securității cibernetice în cadrul orașelor inteligente.

1.2 Prezentarea Generală a Tezei de Doctorat

Obiectivele principale ale acestei teze sunt identificarea celor mai răspândite tipuri de atacuri cibernetice din epoca modernă și dezvoltarea strategiilor de detecție, a contramăsurilor și a soluțiilor de securitate robuste care oferă eficiență, rezultate precise și timp de răspuns rapid.

Primul pas în realizarea acestui obiectiv este acela de a determina cele mai importante tipuri de atacuri cibernetice, vectori de infecție, tendințe de dezvoltare malware și amenințări de securitate care vizează diverse industrii și sisteme de operare. Pe lângă aceasta, am căutat și analizat diferite strategii de detecție și contramăsuri de protecție.

Sub-obiectivele specifice sunt următoarele: detecția adreselor URL scurte care sunt malițioase, identificarea atacurilor de tip smishing, dezvoltarea unui sistem EDR de ultimă generație bazat pe agenți pentru amenințările de securitate cibernetică, detecția cyberbullying-ului în cadrul fișierelor multimedia și guvernarea securității cibernetice în infrastructurile la scară largă.

În capitolul 2, prezentăm cele mai recente tipuri de atacuri cibernetice, vectori de infecție, grupări de atacatori, tendințe de dezvoltare malware și multe alte lucruri de interes care descriu plaja de amenințări cibernetice pentru perioada 2021-2022.

În capitolul 3, evidențiem atât strategiile tradiționale, cât și cele moderne de detecție a atacurilor cibernetice, împreună cu cele mai recente contramăsuri. Ele oferă îndrumări importante pentru apărarea organizațiilor împotriva speciilor malware sofisticate.

În capitolul 4, vă prezentăm un sistem exhaustiv de detecție a URL-urilor scurte ce sunt malware. Noi integrăm sistemul propus cu platforme open-source de threat intelligence și algoritmi de Machine Learning (ML).

În capitolul 5, discutăm o soluție scalabilă bazată pe o arhitectură de tip Cloud-Edge pentru detecția adreselor URL care sunt malware în atacurile de tip smishing. Soluția propusă integrează platforme de threat intelligence și algoritmi de Machine Learning ce clasifică adresele URL în funcție de caracteristicile lor. Pentru a demonstra eficiența soluției noastre, am implementat o

aplicație Android care detectează adrese URL scurte care sunt malware în mesajele SMS și notifică utilizatorul cu privire la legitimitatea acestora. Acest capitol se bazează pe un articol de cercetare ce reprezintă o versiune revizuită și extinsă a unei lucrări intitulată *Malicious Short URLs Detection Technique* prezentată la *2023 22nd RoEduNet Conference: Networking in Education and Research (RoEduNet), Craiova, Romania, 2023*.

În capitolul 6, prezentăm două soluții pentru detecția cyberbullying-ului în fișiere multimedia folosind modele de deep learning. Prima soluție efectuează detecția agresiunii în GIF-uri, în timp ce cealaltă realizează aceeași sarcină pentru videoclipurile de pe TikTok.

În capitolul 7, discutăm despre *NextEDR* - sistem EDR de ultimă generație bazat pe agenți pentru detecția amenințărilor de securitate cibernetică. Acesta reprezintă o platformă inovatoare și interactivă de tip Cloud-Edge-Continuum Endpoint Detection and Response (EDR) pentru protejarea organizațiilor moderne de atacurile de securitatea cibernetică. Noi proiectăm un Proof-of-Concept bazat pe un agent de comunicare interactiv (ChatBot) pentru detecția phishing-ului în adrese URL scurte. Acest proiect pe care l-am propus a făcut parte dintr-un program *Proof of Concept* și am lucrat la elaborarea unui plan de afaceri. Programul se intitulează *Invest National University of Science and Technology POLITEHNICA Bucharest Proof of Concept*.

În capitolul 8, prezentăm un model pentru evaluarea riscului de securitate cibernetică în infrastructurile de scară largă, precum și diferite tehnici de protejare și contramăsuri. Metodologia pe care o folosim în studiul nostru este calitativă. Se bazează pe 66 de proiecte din baza de date CORDIS a Comisiei Europene care au legătură cu guvernanța securității cibernetice în orașele inteligente.

În capitolul 9, prezentăm principalele concluzii de cercetare ale acestei teze, dimpreună cu principalele avantaje și neajunsuri ale fiecăreia dintre abordările de securitate propuse. Obiectivul principal al tezei îl reprezintă identificarea celor mai răspândite amenințări cibernetice în epoca modernă și dezvoltarea unor strategii solide de detecție, contramăsuri și soluții robuste de securitate care oferă precizie ridicată și timp de răspuns rapid.

Table 1.1. Efectele și consecințele amenințărilor cibernetice asupra diferitor domenii.

	Universitar	Guvernamental	Medical	Rețelelor de socializare	Grupurilor mai mari de chat
Short URL în atacuri de tip smishing	1. Dacă studenții sunt păcăliți să furnizeze credențialele, atacatorii ar putea avea acces la sistemele din universitate. 2. Paginile de logare false pot fura informațiile lor de conectare.	1. Dacă angajații furnizează numele lor de utilizator și parola, adversarii ar putea avea acces în sistemele instituțiilor publice. 2. Angajații pot dezvălui informații confidențiale de la locul de muncă.	1. Medicii sunt redirecționați către site-uri medicale false. 2. Atacatorii ar putea fura date medicale.	1. Utilizatorii primesc link-uri suspecte care pot fura date sensibile. 2.Adversarii ar putea trimite link-uri pentru a fura informații de pe cardul de credit.	1. Participanții sunt predispuși la escrocherii. 2. Utilizatorii se pot teme să interacționeze cu ceilalți.
Cyberbullying prin intermediul fișierelor de tip GIF	1. Studenții își pot pierde stima de sine și devin rușinați. 2. Capacitatea de concentrare poate fi afectată.	1. Poate provoca o scădere a eficienței în muncă. 2. Relațiile cu cei din jur sunt afectate.	1. Pacienții pot citi mesaje periculoase care le-ar putea afecta starea de sănătate. 2. Starea emoțională poate fi afectată.	1. Utilizatorii primesc insulte în postările lor. 2. Poate provoca izolare socială.	1. Utilizatorii devin îngrijorați cu privire la valoarea lor. 2. Restul membrilor grupului ar putea râde de persoana agresată.
Cyberbullying prin intermediul videoclipurilor de pe TikTok	1. Provoacă o scădere a rezultatelor academice. 2. Este posibil ca elevii să abandoneze școala.	1. Conduce la deteriorarea relației cu cei din jur. 2. Este posibil să genereze neînțelegeri și scandaluri între angajați.	1. Poate provoca stări de șoc, și dureri de cap pacienților. 2. Pacienții pot refuza consulturile medicilor.	1. Experiențele utilizatorilor devin neplăcute. 2. Poate provoca anxietate pe termen lung.	1. Utilizatorii se pot simți izolați. 2. Persoanele afectate pot părăsi grupul.
Cyberbullying prin intermediul mesajelor de tip text	1. Crește nivelul de stres și depresie. 2. Elevii se pot simți inferiori față de ceilalți colegi.	1. Angajații nu mai sunt implicați atât de mult în activitățile de la muncă. 2. Angajații pot deveni mai agresivi și suspicioși față de cei de lângă.	1. Poate provoca insomnie si oboseala pacienților. 2. Pacienții pot deveni mai stresați, ceea ce îi poate determina să mănânce în exces.	1. Utilizatorii primesc comentarii negative. 2. Utilizatorii pot deveni nemulțumiți în ceea ce privește aspectul lor fizic.	1. Utilizatorii primesc insulte. 2. Ceilalți membri ai grupului ar putea amplifica hărțuirea victimei și în final să le elimine din grup.

Table 1.2. Obiectivele și metodologia tezei.

Întrebare de cercetare	Obiectiv de cercetare	Metodologie de cercetare	Caz de utilizare
RQ1	Pentru a identifica tipurile de atacuri cibernetice moderne și cum ar putea organizațiile să se apere împotriva lor.	Dovedit de literatură și rapoartele de securitate	Pentru a răspunde la această întrebare, am analizat cele mai răspândite atacuri cibernetice care definesc plaja de amenințări pentru perioada 2021-2022. Am studiat principalele tipuri de atacuri care vizează diverse organizații și industrii, vectorii de infecție utilizați pentru a obține acces în cadrul sistemelor, tendințele actuale de dezvoltare malware, cele mai active grupări de atacatori și așa mai departe. Mai mult, am propus atât strategii de detecție tradiționale, cât și moderne, alături de cele mai recente contramăsuri. Detalii despre rezultatele obținute pot fi citite în capitolele 2 și 3 și în lucrarea [20].
RQ2	Pentru a detecta URL-uri scurte care sunt malware.	Dovedit prin simulări	Pentru a răspunde acestei întrebări, am propus un sistem exhaustiv de detecție a URL-urilor scurte rău intenționate prin utilizarea unor platforme populare de Threat Intelligence (TI) cum ar fi VirusTotal și PhishTank și prin folosirea unor algoritmi de Machine Learning (ML). Sistemul nostru funcționează pentru orice URL, indiferent de serviciul de scurtare utilizat, fie public, fie privat. Pentru a demonstra eficiența soluției noastre, am dezvoltat un sistem generic de detecție care identifică adrese URL scurte rău intenționate și notifică utilizatorul cu privire la legitimitatea acestora. Detalii despre rezultatele experimentale pot fi citite în capitolul 4 și în lucrarea [19].
RQ3	Pentru a detecta URL-uri scurte în atacurile de tip smishing.	Dovedit prin simulări	Pentru a oferi un răspuns clar la această întrebare, noi am propus o tehnică scalabilă ce se bazează pe o arhitectură de tip Cloud-Edge pentru detecția adreselor URL malware în atacurile de smishing. Soluția noastră folosește platforme de Threat Intelligence (de exemplu, VirusTotal, PhishTank) și algoritmi de Machine Learning (ML) care clasifică adresele URL în funcție de caracteristicile lor. Acuratețea modelului de ML este de până la 97%. Pentru a demonstra eficiența soluției noastre, am implementat o aplicație Android care detectează adrese URL scurte din mesajele SMS care sunt malware și notifică utilizatorul cu privire la legitimitatea acestora. Aceasta reprezintă o versiune revizuită și extinsă a soluției publicate în lucrarea [19]. Mai multe detalii despre rezultatele experimentale pot fi citite în capitolul 5.
RQ4	Pentru a crește eficienta sistemelor de securitate în detecția atacurilor cibernetice.	Dovedit prin simulări	Pentru a răspunde la această întrebare, am propus NextEDR, un sistem de tip Endpoint Detection and Response (EDR) de ultimă generație, bazat pe agenți, ce are rol în detecția amenințărilor de securitate cibernetică. Este o platformă inovatoare și interactivă de tip Cloud-Edge-Continuum EDR pentru protejarea organizațiilor moderne de atacurile de securitate cibernetică. Noi am proiectat un Proof-of-Concept (PoC) bazat pe un agent de comunicare interactivă (ChatBot) pentru detecția atacurilor de tip phishing în cadrul adreselor URL scurte. Solutia noastră este o platformă multistrat centrată pe dispozitivile mobile ce se bazează pe modelul Cloud-Edge-Continuum. Acest proiect pe care l-am propus a făcut parte dintr-un program de tip Proof-of-Concept și am lucrat la elaborarea unui plan de afaceri. Dezvoltarea produsului minim viabil (MVP) pentru platforma NextEDR a implicat o rafinare iterativă bazată pe feedback-ul utilizatorilor, prioritizarea funcțiilor esențiale pentru a oferi o soluție funcțională care să răspundă eficient nevoilor principale ale utilizatorilor. Detalii despre rezultatele experimentale pot fi citite în capitolul 7 și în lucrarea [10].
RQ5	Pentru a detecta cyberbullying în cadrul fișierelor multimedia.	Dovedit prin simulări	Pentru a oferi un răspuns perspicace la această întrebare, propunem un sistem pentru detecția cyberbullying-ului în fișierele multimedia folosind modele de tip Deep Learning (DL). Acesta cuprinde două soluții: una care realizează detectarea agresiunii în GIF-uri și alta care realizează aceeași sarcină pentru videoclipurile de pe TikTok. Prima soluție folosește o arhitectură hibridă care cuprinde o rețea neuronală convoluțională (CNN) și trei rețele neuronale recurente (RNN). Al doilea sistem folosește un model bazat pe Transformer care funcționează pe caracteristici generate de CNN. Detalii despre rezultatele experimentale pot fi citite în capitolul 6 sau în lucrarea [18].
RQ6	Pentru a identifica principalele riscuri de securitate și eventualele contramăsuri pentru infrastructurile de scară largă.	Dovedit de literatură și rapoartele de securitate	Pentru a răspunde la această întrebare, am propus un model pentru evaluarea riscului de securitate în cadrul infrastructurilor de scară largă, precum și tehnici și contramăsuri de protecție. Metodologia pe care am folosit-o în studiul nostru este calitativă și se bazează pe proiecte din baza de date CORDIS a Comisiei Europene care se referă la guvernanța securității cibernetice în orașele inteligente. Detalii despre rezultatele obținute pot fi citite în capitolul 8.

2 | Analiză Critică a Atacurilor Cibernetice Moderne

În acest capitol, discutăm despre cele mai recente tipuri de atacuri cibernetice, vectori de infecție, grupări de atacatori, tendințele de dezvoltare a programelor malware și multe alte lucruri de interes care descriu plaja amenințărilor de securitate pentru perioada 2021-2022. Conținutul acestui capitol se bazează pe publicarea lucrării *Modern Cyber Security Attacks, Detection Strategies, and Countermeasures Procedures* în *2023 24th International Conference on Control Systems and Computer Science (CSCS)*.

2.1 Tendințe Actuale de Dezvoltare Malware

Autorii de malware își folosesc abilitățile pentru a construi specii mai sofisticate care pot ocoli cu ușurință sistemele de securitate implementate. În această secțiune, prezentăm tendințele moderne de dezvoltare malware, acoperind tehnicile care sunt cele mai folosite de adversari pentru a-și atinge scopul final.

Capabilități de Nivel Avansat de Evaziune a Detecțiilor

Autorii de ransomware și-au schimbat tactica de atac folosind criptarea intermitentă care este mult mai rapidă (doar unele blocuri de date sunt criptate, nu întregul sistem de fișiere). Unele schimbări au fost observate și în comunicațiile C2, adversarii făcând uz de tehnologia cloud pentru a nu fi descoperiți. Aceștia folosesc de asemenea protocolul DNS pentru a-și ascunde comunicarea (de exemplu DNS Tunneling). Pentru a împiedica analiza programelor malware, adversarii folosesc mecanisme avansate de împachetare a codului (de exemplu, UPX, Themida, MPRESS) și tehnici de ofuscare. Utilizarea diferitelor limbaje de programare, cum ar fi PureBasic sau Nim, îngreunează procesul de inginerie inversă [15].

Fișierele Malware Targhetează Programele de Virtualizare

Platformele de virtualizare preferate de adversari pentru a fi atacate sunt containerele Docker și Windows sau Kubernetes. Au fost observate mai multe familii de ransomware care vizează serverele VMWare ESXi bazate pe Linux. În loc să cripteze sistemul de operare care rulează în interiorul acestora, adversarii au ales să folosească acest tip de atac împotriva fișierelor ce sunt specifice mașinii virtuale (VM).

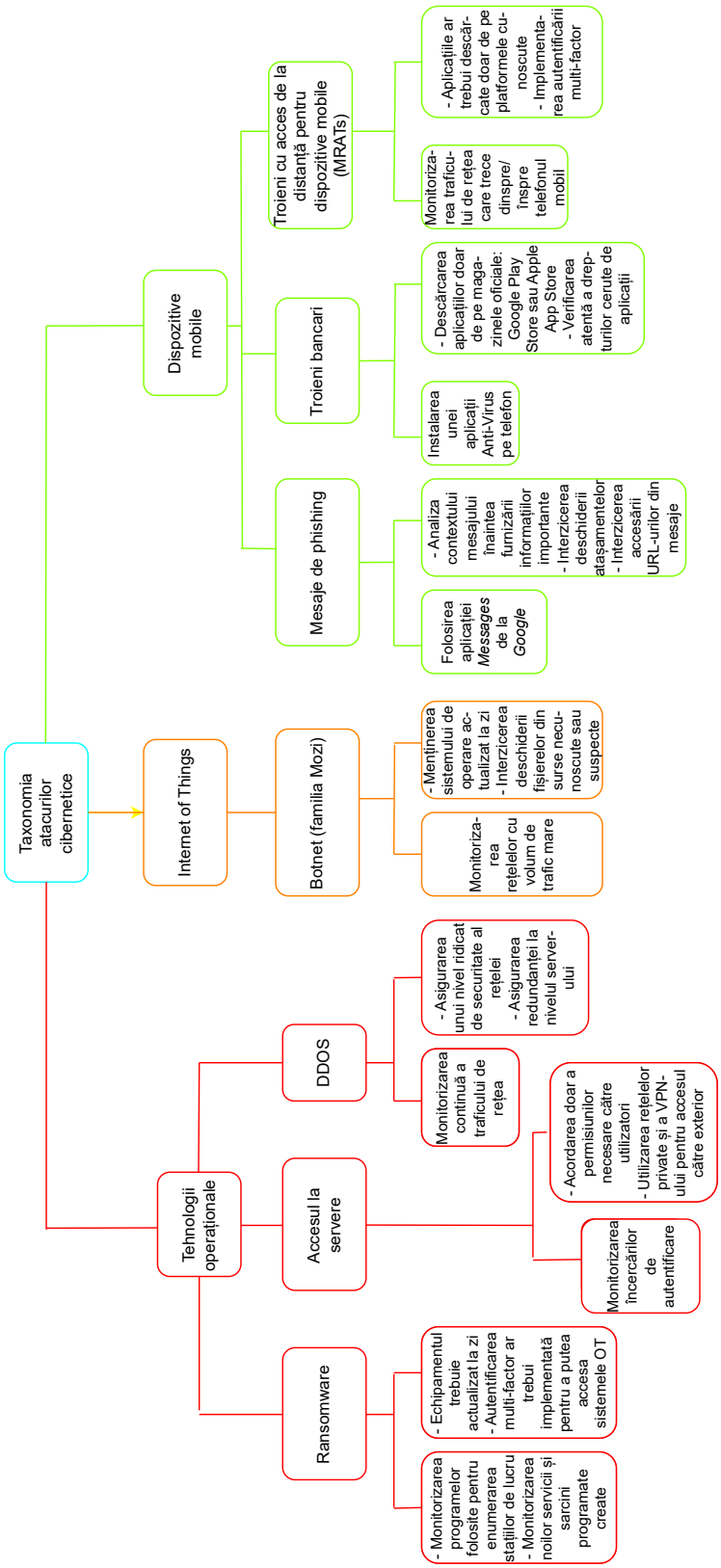

Figure 2.1. Taxonomia atacurilor cibernetice.

2.2 Plaja Curentă a Amenințărilor de Securitate

Această secțiune prezintă cele mai răspândite amenințări cibernetice din epoca modernă. Aceasta include cele mai importante tipuri de atacuri cibernetice, vectori de infecție, tendințe geografice, cele mai active grupări de atacatori și multe alte lucruri de interes.

În figura 2.1, taxonomia începe cu o clasificare scurtă a echipamentelor sau dispozitivelor care au fost targhetate de către adversari. Apoi, cele mai răspândite tipuri de atacuri sunt expuse pentru fiecare dintre aceste platforme. În cele din urmă, strategiile de detecție și contramăsurile care pot fi implementate pentru fiecare tip de atac sunt afișate în partea stângă, respectiv în partea dreaptă.

Cele mai răspândite tipuri de atacuri cibernetice care au fost observate în epoca modernă sunt ransomware, accesul la servere, compromiterea e-mailului de business, furtul de date și credențiale.

Un vector de atac reprezintă o metodă prin care un adversar obține acces neautorizat la un sistem pentru a realiza acțiuni malițioase [23]. Principalii vectori de infecție utilizați de adversari sunt phishing-ul, exploatarea vulnerabilităților, credențialele compromise, atacurile de tip brute force și conexiunea la distanță a desktopului.

Potrivit unui raport publicat de Proofpoint, cercetătorii au identificat în Europa o creștere cu 500% a încercărilor de livrare a programelor malware pe mobil de la începutul până la sfârșitul lunii februarie 2022 [9]. Programele malware de astăzi care vizează terminalele mobile au mai multe capabilități cum ar fi cele de a urmări locația, de a șterge fișierele importante sau de a înregistra atât audio, cât și video. Tabelul 2.1 descrie cele mai comune specii de malware pentru dispozitivele mobile [9].

Table 2.1. Tipuri, funcționalități și răspândire geografică a speciilor de malware pentru telefon, 2022.

	Sistem de operare țintă	**Impersonare aplicații**	**Impersonare financiară**	**Furt de credențiale**	**Acces la cameră și microfon**	**Distribuire prin SMS**	**Escalare de privilegii**	**Locație geografică**
FluBot	Android	Da	Da	Da	Nu	Da	Da	Asia, UK, Europa
TeaBot	Android	Da	Da	Da	Nu	Da	Da	UK, Europa
TangleBot	Android	Nu	Da	Da	Da	Nu	Da	America de Nord
MoqHao	Android	Da	Da	Da	Nu	Da	Nu	Asia, Japonia
BRATA	Android	Da	Da	Da	Nu	Da	Nu	Regatul Unit, Europa, America Latină
TianySpy	Android, iOS	Da	Da	Da	Nu	Nu	Nu	Japonia
KeepSpy	Android	Da	Da	Da	Nu	Nu	Nu	Japonia

Industriile din 2021 care au fost cele mai afectate de atacuri au fost cele de producție, petrol și gaze, transport, utilități, minerit și inginerie. Ransomware este liderul în tipurile de atacuri care au vizat industriile în anul 2021.

Potrivit studiului X-Force Incident Response [15], infractorii cibernetici sunt lideri în sursa atacurilor cibernetice. Ei sunt urmați de actorii statali care au avut ca obiective principale spionajul și supravegherea. Numărul cel mai mic de atacuri a provenit din partea hacktiviștilor.

Majoritatea atacurilor de securitate cibernetică care au avut loc în 2021 au fost direcționate

către Asia, cu un procent de 26% din numărul total. Din această regiune, Japonia a fost cea mai atacată datorită jocurilor olimpice care au avut loc acolo în vara lui 2021. Cea mai răspândită familie de malware care a vizat acest continent a fost ransomware-ul REvil. Industriile de finanțe și asigurări au fost cele mai afectate. Asia este urmată îndeaproape de Europa și America de Nord.

2.3 Concluzii

Acest capitol tratează aspecte cheie semnificative privind amenințările moderne de securitate cibernetică, oferind o privire de ansamblu asupra tipurilor, platformelor celor mai vizate și regiunilor geografice. De asemenea, evidențiază inovațiile apărute în domeniul dezvoltării de software malițios, atribuindu-le uneia dintre cele mai active grupări de atacatori pe care le-am prezentat. Acestea vizează diferite organizații private și publice în funcție de interesele sau cauzele pentru care luptă sau protestează.

3 | Strategii de Detecție și Contramăsuri pentru Atacurile Cibernetice

În acest capitol, discutăm atât despre strategiile tradiționale și moderne de detecție a atacurilor cibernetice, cât și despre cele mai recente contramăsuri. Conținutul acestui capitol se bazează pe publicarea lucrării *Modern Cyber Security Attacks, Detection Strategies, and Countermeasures Procedures* în *2023 24th International Conference on Control Systems and Computer Science (CSCS)*.

3.1 Strategii Tradiționale de Detecție a Atacurilor Cibernetice

Sistemele tradiționale de securitate, cum ar fi firewall-urile, aplicațiile de rețea privată virtuală (VPN), soluțiile antivirus (AV) și sistemele de detecție/prevenție a intruziunilor (IDS/IPS) sunt utilizate pentru a securiza accesul la un anumit dispozitiv sau rețea și a proteja datele sensibile. Acestea se bazează pe reguli sau semnături pentru a detecta amenințările de securitate. Vă prezentăm mai jos care sunt sistemele tradiționale de securitate:

- Firewall - Este un dispozitiv de securitate a rețelei care monitorizează traficul care intră în rețea (inbound) și cel care iese din rețea (outbound). Firewall-ul poate permite trecerea traficului sau îl poate bloca pe baza unor reguli de securitate [8].
- VPN - Este o soluție de securitate care asigură confidențialitatea utilizatorilor pe Internet prin ascunderea adresei IP. De asemenea, securizează conexiunile la Internet atunci când utilizatorii navighează. Când este utilizată o soluție VPN, toate datele în tranzit sunt criptate [22].
- Anti-Virus - Este o soluție de securitate care protejează sistemele de diverse amenințări prin blocarea site-urilor web rău intenționate și prin carantinarea și ștergerea fișierelor sau a programelor periculoase. O soluție Anti-Virus scanează stațiile de lucru pentru detectarea programelor malware și primește periodic actualizări cu noi semnături de viruși pentru a identifica cele mai recente amenințări [3].
- IDS/IPS - Sistemele de detecție a intruziunilor se ocupă cu monitorizarea și analiza traficului pentru identificarea amenințărilor. Soluțiile IPS au aceleași capabilități ca un echipament de tip IDS, doar că, în plus, acestea pot preveni atacurile cibernetice prin întreruperea conexiunilor de rețea sau eliminarea pachetelor [1].

3.2 Strategii Moderne de Detecție a Atacurilor Cibernetice

Sistemele actuale de detecție a atacurilor cibernetice fie efectuează analize la nivelul stației de lucru, fie monitorizează traficul rețelei pentru a capta date legate de atacurile cibernetice [13].

Unele strategii populare de detecție a atacurilor sunt:

- Host Intrusion Detection Systems (HIDS) - acest echipament se instalează la nivelul stației de lucru și monitorizează diferite activități care au loc. Pentru a avea o mai bună vizibilitate asupra operațiunilor în curs, acesta colectează fișiere de jurnal și date de rețea și se uită la apelurile de sistem, procesele executate, arborele de procese și așa mai departe.
- Network Intrusion Detection Systems (NIDS) – acest echipament poate monitoriza traficul de rețea al unei întregi clase care trece printr-o legătură de rețea și pe baza acestuia poate identifica diferite atacuri cibernetice. Un NIDS este instalat la nivelul unei mașini fizice pe care trebuie să fi activat un port în modul promiscuous pentru a intercepta traficul. Cu toate acestea, un NIDS nu poate să ofere protecție împotriva atacurilor care au loc la nivelul stațiilor de lucru, așa cum face un HIDS [7].

În prezent, cele mai importante abordări implementate la nivelul sistemelor care sunt folosite în detecția atacurilor cibernetice sunt semnătura, anomalia și specificația. Ca o simplă comparație, detecția pe bază de semnătură este utilizată pentru amenințările cunoscute, în timp ce detecția bazată pe anomalii este folosită pentru schimbările comportamentale, fiind capabilă să detecteze atacuri noi sau necunoscute.

3.3 Contramăsuri Moderne împotriva Atacurilor Cibernetice

Pentru a îngreuna operațiunile atacatorilor și pentru a le reduce șansele de succes, trebuiesc a fi implementate unele contramăsuri. Mai jos sunt prezentate câteva dintre cele mai importante contramăsuri actuale:

- **Audit de securitate** – se recomandă să se efectueze în mod regulat audituri de securitate pentru a descoperi posibile vulnerabilități, riscuri, configurări greșite, programe sau servicii învechite și așa mai departe, care pot fi exploatate cu ușurință de către un atacator. În companii, această măsură asigură că datele clienților sunt protejate [14]. Mai mult decât atât, o altă opțiune fezabilă ar fi testele de penetrare în care sunt emulate capacitățile specifice unei grupări de atacatori.
- **Conștientizarea utilizatorilor** – toți angajații unei companii ar trebui să cunoască ultimele tehnici de phishing/vishing folosite de atacatori pentru a fura credențiale, informații despre cardul de credit sau orice alt tip de date sensibile. Pentru a preveni astfel de atacuri, angajații ar trebui să participe la diferite cursuri de securitate.
- **Actualizarea aplicațiilor software** – una dintre bunele practici se referă la menținerea aplicațiilor software actualizate. Acest lucru necesită unele politici cu privire instalarea actualizărilor software și a patch-urilor de securitate pentru sistemul de operare, programele de pe calculator, drivere etc. Actualizarea software oferă multe beneficii, cum ar fi remedierea erorilor, repararea vulnerabilităților, îmbunătățirea performanței și lansarea de noi funcționalități. [21].
- **Endpoint Detection and Response** (EDR) – un plus în prevenția și detecția atacurilor cibernetice poate fi adus de soluțiile EDR care monitorizează, identifică, elimină sau carantinează în mod automat fișierele malware. Soluțiile EDR colectează date atât de la nivelul stațiilor de lucru, cât și de la cel al rețelei și oferă capabilități de detecție, prevenție, investigare și răspuns la incidente.

3.4 Concluzii

Cercetarea noastră abordează strategii de detecție și contramăsuri moderne împortiva atacurilor cibernetice care pot ajuta utilizatorii obișnuiți și cercetătorii să se protejeze, să detecteze sau cel puțin să minimizeze riscul unui atac de succes care i-ar putea viza. Lucrarea noastră propune mai multe proceduri populare și eficiente de apărare.

4 | Tehnică de Detecție a Adreselor URL Scurte care sunt Malițioase

În acest capitol, prezentăm un sistem exhaustiv de detecție a adreselor URL scurte care sunt malițioase. Sistemul propus corelează date din cadrul unor platforme populare de Threat Intelligence precum VirusTotal și PhishTank și utilizează diverși algoritmi de învățare automată (ML). Sistemul nostru funcționează pentru orice adresă URL, indiferent de serviciul de scurtare utilizat: public sau privat. Conținutul acestui capitol se bazează pe publicarea lucrării *Malicious Short URLs Detection Technique* în *2023 22nd RoEduNet Conference: Networking in Education and Research (RoEduNet)*. Acest capitol este organizat astfel: mai întâi, prezentăm în secțiunea 4.1 o introducere în domeniul problemei studiate. Apoi, secțiunea 4.2 prezintă detaliile de implementare ale soluției propuse. În secțiunea 4.3 prezentăm o analiză a rezultatelor experimentale. Concluziile referitoare la capabilitățile soluției propuse în detecția adreselor URL scurte care sunt malițioase, încheie acest capitol.

4.1 Introducere

Un serviciu de scurtare a adreselor URL reprezintă un sistem care preia un URL (Uniform Resource Locator) de la un utilizator și îl convertește într-o formă scurtă. Aceasta din urmă reprezintă un alias al celui original și poate fi distribuită către oricine. Când cineva accesează o adresă URL scurtă, în primul rând se realizează o conexiune de rețea către furnizorul de servicii care a scurtat adresa URL, iar ulterior acesta va redirecționa conexiunea către site-ul web original pe care utilizatorul intenționează să îl acceseze. Acest proces are loc în mod automat, fără intervenția utilizatorului. Adresele URL scurte pot fi utilizate atât în scopuri legitime, cât și în cele malițioase. Principalul avantaj pentru adversari atunci când folosesc acest tip de URL este că pot ascunde destinația finală. Prin urmare, victimele nu sunt conștiente de pagina destinație către care sunt redirecționate și nu pot identifica unele caracteristici suspecte care ar fi prezente în adresa URL inițială.

Studiul nostru se concentrează pe dezvoltarea unui sistem exhaustiv de detecție a adreselor URL scurte care sunt malware. Pentru a demonstra eficiența sistemului, vom lua în considerare toate tipurile de adrese URL scurte, indiferent de serviciul de scurtare utilizat: public sau privat. Mai mult, obținem adresa URL finală la care va ajunge utilizatorul urmărind toate conexiunile de tip HTTP Redirect care au loc. Mai mult decât atât, integrăm sistemul nostru cu platforme open-source de Threat Intelligence precum VirusTotal și PhishTank pentru a verifica reputația adresei URL finale. Dacă motoarele de scanare nu identifică adresa URL finală ca fiind malware sau suspectă, atunci aceasta este transmisă modelului nostru de învățare automată propus pentru clasificare. Precizia modelului ajunge până la 97%.

Următoarea secțiune prezintă detaliile de implementare ale soluției propuse.

4.2 Proiectarea și Implementarea Soluției

În situația în care URL-ul nu este detectat de motoarele Anti-Virus, sau nu este prezent în baza de date de phishing, propunem un strat suplimentar de securitate care utilizează algoritmi de Machine Learning (ML). Modelul de ML are o acuratețe de detecție foarte mare, datorită faptului că am propus un set îmbunătățit de caracteristici. Arhitectura sistemului este descrisă în figura 4.1:

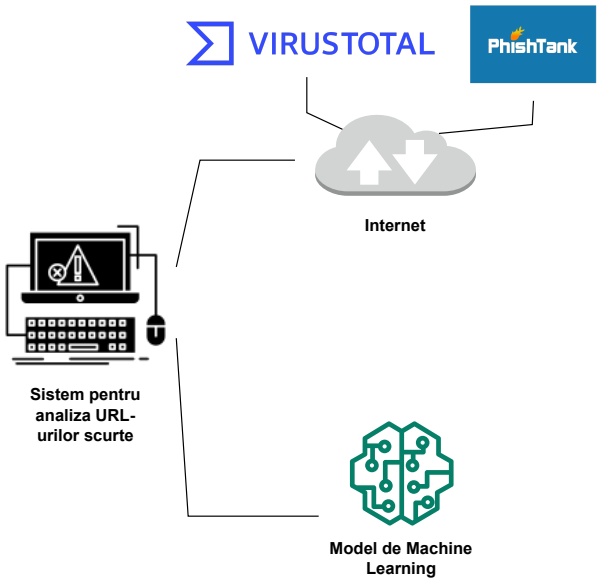

Figure 4.1. Arhitectura generală a sistemului.

Mai jos, prezentăm detaliile fiecăreia dintre abordările de detecție.

Abordarea ce se bazează pe Platforme de Threat Intelligence

Figura 4.2 prezintă abordarea ce utilizează platformele de Threat Intelligence pe care am propus-o în evaluarea de securitate inițială a unui URL scurt. Pentru a prezenta eficiența soluției noastre, am dezvoltat o aplicație în Java folosind IntelliJ IDEA IDE. Acesta așteaptă să recepționeze o adresă URL scurtă pe care mai apoi o redirecționează către modulul următor. Aici, URL-ul final este obținut prin accesarea adresei URL scurte ce a fost recepționată și urmărirea tuturor conexiunilor de tip HTTP Redirect ce vor avea loc ulterior. URL-ul final este trimis prin intermediul API-urilor către platformele open-source de Threat Intelligence (VirusTotal și PhishTank) pentru a fi scanate. Dacă acestea îl consideră ca fiind malware sau suspect, utilizatorul este notificat, iar în caz contrar, adresa URL finală este trimisă pentru clasificare către modelul de Machine Learning (ML).

Abordarea ce se bazează pe Algoritmi de Machine Learning

Această fază are rolul de a asigura un nivel suplimentar de securitate astfel încât să fie luată decizia corectă cu privire la natura unei adrese URL. Înainte de a trimite adresa web spre clasificare, am antrenat modelul de Machine Learning (ML) pe un set de date pe care l-am obținut din [24]. Acesta conține 11430 de instanțe și un set de 87 de caracteristici. L-am adaptat, am adăugat încă trei caracteristici și i-am îmbunătățit acuratețea la 96,7454% utilizând algoritmul Random Forest.

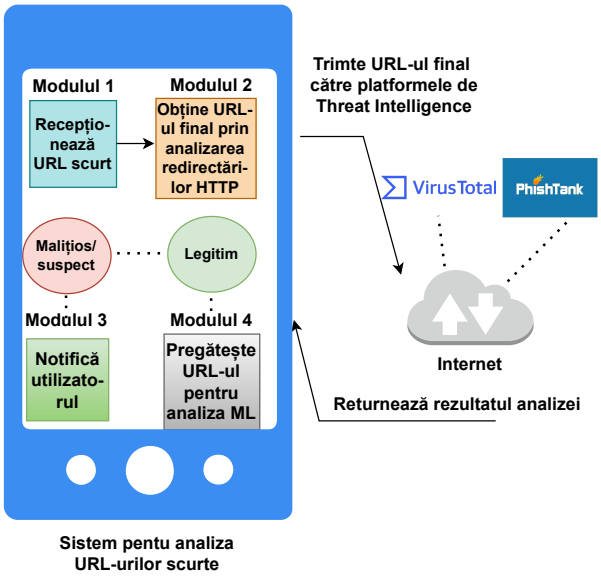

Figure 4.2. Abordarea Threat Intelligence.

Aceste caracteristici fac parte din următoarele trei categorii: proprietățile lexicale ale URL-ului, conținutul site-ului web și specificațiile externe ale domeniului.

Caracteristicile pe care le propunem aparțin proprietăților lexicale ale URL-ului și se referă la numărul de parametri conținuți în URL, dacă acesta din urmă conține sintaxa "&" în loc de simbolul "&" și dacă un fragment este menționat în cadrul URL-ului pentru ca utilizatorul să fie direcționat într-o anumită zonă a paginii web atunci când accesează adresa URL.

Figura 4.3 prezintă abordarea ce folosește algoritmii de machine learning pe care noi o propunem ca un strat suplimentar în vederea verificărilor de securitate, astfel încât să fie luată o decizie corectă.

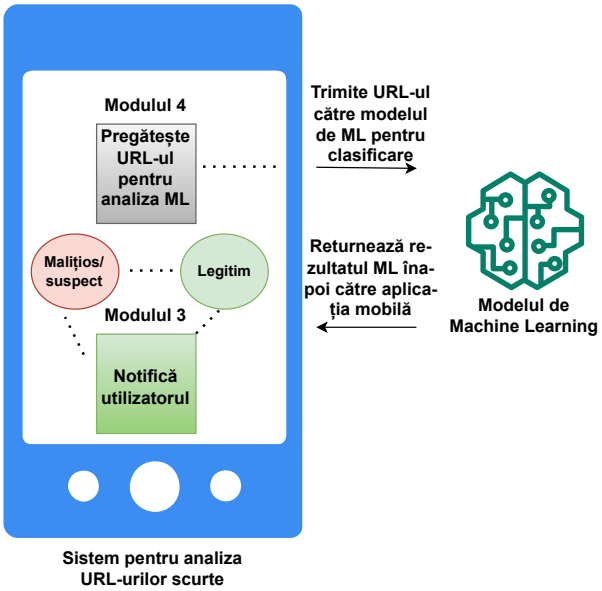

Figure 4.3. Abordarea Machine Learning.

În faza de *Pregătire URL pentru analiza ML*, avem URL-ul deja convertit din forma sa scurtă în cea originală care indică destinația web finală. În continuare, adresa web este trimisă la modelul de ML pentru clasificare. Odată finalizat acest proces, utilizatorul este notificat cu privire la reputația URL-ului: malițios sau legitim.

4.3 Rezultate și Analize Experimentale

În această secțiune, descriem rezultatele obținute atât prin abordarea ce utilizează platformele de Threat Intelligence, cât și prin cea care folosește algoritmii de Machine Learning (ML) în detecția adreselor URL scurte care sunt malware. De asemenea, prezentăm detalii cu privire la procesele de antrenare a modelului de ML și extragere a caracteristicilor. În cele din urmă, comparăm câțiva algoritmi de ML în ceea ce privește acuratețea în detecția adreselor URL scurte ce sunt malițioase.

Analiza Platformelor de Threat Intelligence

În această subsecțiune, discutăm despre eficiența și timpul de detecție a abordării ce utilizează platformele de Threat Intelligence (TI). Am efectuat câteva teste empirice pentru unele adrese URL care au fost scurtate și care aparțin următoarelor categorii: malware, legitime și phishing. Am luat câte trei adrese URL pentru fiecare dintre aceste clase. Mai multe detalii despre URL-urile scurte pe baza cărora am efectuat testele sunt prezentate în Tabelul 4.1.

Procesele de Antrenare și Clasificare ale Modelului de Machine Learning

În această subsecțiune, discutăm diferite metrici de timp referitoare la fazele de antrenare a modelului de Machine Learning (ML), extragere a caracteristicilor și clasificare. Pentru a testa rapiditatea procesului ce combină etapele de extragere a caracteristicilor și clasificare, am ales zece adrese URL pentru fiecare dintre categorii: phishing, legitim și necunoscut. Timpul pentru aceste

Table 4.1. Valori de durată pentru analiza ce folosește platformele de Threat Intelligence.

URL Scurt	URL Final	Reputația URL-ului	Timpul de analiză (s)
https://shorturl.at/fuxAE	https://call.raidstore.org/	Malițios	81.267
https://t.ly/yEC_	https://technology.macosevents.com/	Malițios	2.394
https://rb.gy/w2i1u	https://press.infomapress.com/	Malițios	1.889
https://rb.gy/wtkns	https://stiri.botosani.ro/	Legitim	1.737
https://tinyurl.com/kyc44ft8	https://www.bbc.com/	Legitim	2.717
https://cutt.ly/zwr2HIaS	https://www.amazon.com/b?node=21576558011&ref_=alxcom_lrnmore_btn_23	Legitim	2.864
https://urlis.net/b94wy861	https://sucursalcentroapp.brizy.site/	Phishing	2.681
http://tiny.cc/tzb8vz	https://kuccoiieeiinelovuiie.godaddysites.com/	Phishing	2.499
https://url1.io/s/VpEwo	https://anime-info777.com	Phishing	3.879

două procese combinate poate varia de la 2,8171 (s) la 118,9281 (s), așa cum se poate vedea în Tabelul 4.2, deoarece procesul de extragere a caracteristicilor depinde de factori precum dimensiunea conținutului site-ului web, timpul de răspuns al altor platforme interogate (spre exemplu, Open PageRank pentru popularitatea domeniului, iPTY pentru vârsta domeniului).

Table 4.2. Valori de durată pentru antrenarea modelului, extragerea caracteristicilor și clasificare.

Operațiune	Durată (s)
Antrenarea modelului de ML	8.3454
Extragerea caracteristicilor + clasificare	Min: 2.8171
	Max: 118.9281
	Avg: 12.2929
Clasificare	Min: 2.4097
	Max: 3.3585
	Avg: 2.6234

Acuratețea Algoritmilor de Clasificare Machine Learning

Am analizat mai multe lucrări de specialitate din literatură care abordează problema studiată și am observat că cei mai răspândiți clasificatori de învățare automată (ML) folosiți de alți cercetători sunt JRip, PART, J48 și Random Forest. JRip și PART sunt clasificatori bazați pe reguli, în timp ce J48 și Random Forest se bazează pe arbori de decizie. Nu am ales algoritmi de învățare automoată complecși cum ar fi rețelele neuronale convolutive, deoarece nu avem atât de multe date de clasificat, iar pe de altă parte, aceștia sunt utilizați de obicei pentru clasificarea imaginilor sau a videoclipurilor. Un alt motiv pentru care nu am ales algoritmi de învățare automată complecși este că aceștia necesită multe resurse de calcul, iar noi ne dorim ca soluția noastră de detecție să poată fi implementată pe orice calculator.

Am rulat patru algoritmi de clasificare Machine Learning (ML) pe setul nostru de date: JRip, PART, J48 și Random Forest. Cea mai bună acuratețe a fost obținută folosind Random Forest. Tabelul 4.3 arată acuratețea fiecărui algoritm, înainte și după ce caracteristicile propuse de noi au fost integrate în modelul de ML.

Table 4.3. Acuratețea algoritmilor de ML.

Algoritm de ML	**Acuratețea cu setul de caracteristici propus (%)**	**Acuratețea fără setul de caracteristici propus (%)**
JRip	94.6194	94.5757
PART	95.4243	94.8206
J48	94.7332	94.5932
Random Forest	96.7454	96.6054

4.4 Concluzii

Acest capitol prezintă o soluție combinată care utilizează algoritmi de învățare automată (ML) și diferite platforme open-source de Threat Intelligence pentru a detecta adrese URL scurte care sunt malware. Sistemul este conceput pentru orice utilizator și acoperă diferite scenarii: URL-uri legitime, suspecte și malițioase care sunt scurtate. Rezultatele au arătat că adrese URL scurte care sunt malițioase pot fi identificate, iar utilizatorul este notificat cu privire la legitimitatea URL-urilor.

Sistemul nostru analizează orice URL scurt, indiferent de serviciile de scurtare utilizate, publice sau private.

Acuratețea modelului de ML este de aproape 97% folosind algoritmul Random Forest.

5 | Tehnică Distribuită de Detecție a Adreselor URL Malware în cadrul Atacurilor de tip Smishing

În acest capitol, discutăm despre o tehnică distribuită de detecție a adreselor URL malițioase în cadrul atacurilor de tip smishing. Soluția propusă se bazează pe o arhitectură de tip Cloud-Edge și integrează platforme de Threat Intelligence (VirusTotal, PhishTank) și algoritmi de învățare automată care clasifică adresele URL în funcție de caracteristicile lor. Pentru a demonstra eficiența soluției noastre, implementăm o aplicație Android care detectează adresele URL scurte care sunt malițioase în mesajele SMS și notifică utilizatorul cu privire la legitimitatea acestora. Conținutul acestui capitol se bazează pe publicarea lucrării *Scalable Malicious URL Detection Technique for Smishing Attacks* în *International Journal of Computational Science and Engineering*. Acest capitol este structurat după cum urmează: Secțiunea 5.1 prezintă o introducere în domeniul problemei studiate. În secțiunea 5.2, vă prezentăm soluția propusă. Secțiunea 5.3 prezintă implementarea soluției noastre, demonstrând eficiența acesteia, iar în Secțiunea 5.4, evidențiem configurația folosită și rezultatele experimentale obținute. În cele din urmă, în Secțiunea 5.5, concluzionăm rezultatele soluției și identificăm oportunități viitoare de cercetare.

5.1 Introducere

Cea mai frecventă metodă de a obține acces în cadrul smartphone-urilor și de a targheta cât mai mulți oameni este phishingul prin SMS (denumit și smishing) [17], [16]. Conform raportului 2022 Cyber Attack Trends: Mid-Year de la Check Point [12], FluBot este al doilea în topul familiilor de malware pentru dispozitive mobile la nivel global. Acesta este un malware bancar pentru sistemele Android ce folosește smishing-ul ca vector de infecție. În plus, acesta utilizează același mesaj SMS pentru a-l trimite tuturor persoanelor din agenda de contacte a victimei, ceea ce conduce la o răspândire exponențială.

În această lucrare, prezentăm o soluție distribuită bazată pe modelul Cloud-Edge pentru detecția atacurilor smishing. Soluția propusă diminuează deficiențele sistemelor EDR clasice întrucât noi integrăm soluții moderne, cum ar fi platformele de Threat Intelligence și algoritmii de învățarea automată. Această propunere se bazează pe cercetările noastre anterioare legate de detecția adreselor URL scurte care sunt malware [19]. Acest studiu propune un sistem exhaustiv de detecție a adreselor URL scurte care sunt malițioase prin valorificarea informațiilor din cadrul platformelor populare de Threat Intelligence precum VirusTotal și PhishTank și prin utilizarea diverșilor algoritmi de învățare automată (ML).

5.2 Arhitectura Propusă

În această secțiune, descriem arhitectura pe care am propus-o împreună cu componentele acesteia. Vă prezentăm avantajele ei în ceea ce privește securitatea, scalabilitatea și eficiența pe care utilizatorii finali le-ar obține atunci când o folosesc.

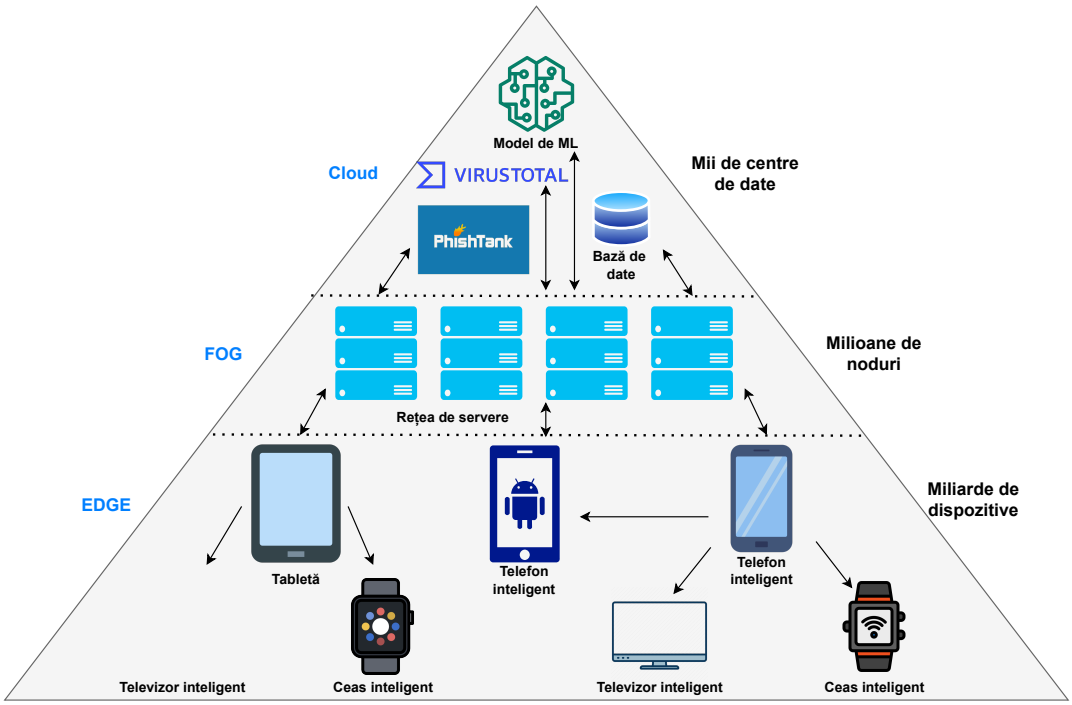

Figure 5.1. Arhitectura Cloud-Edge.

Proiectăm o arhitectură distribuită bazată pe modelul **Cloud-Edge** care urmărește să reducă supraîncărcarea rețelei de comunicare prin gateway-uri de comunicare (de exemplu, set-top box-uri din modelul Cloud-Edge), așa cum este prezentat în Figura 5.1. În stratul **Edge**, avem dispozitive mobile precum smartphone-uri, ceasuri inteligente, tablete și televizoare inteligente. Mai mult decât atât, în stratul **Fog**, propunem o soluție bazată pe set-top box-uri care sunt conectate la stratul Edge și, de asemenea, la stratul Cloud. Obiectivul principal al dispozitivelor Fog este de a primi cereri de verificare URL de la dispozitivele Edge și de a le trimite la stratul **Cloud** unde sunt analizate folosind următoarele două abordări:

- abordarea Threat Intelligence care se bazează pe platforme precum VirusTotal și PhishTank și
- abordarea Machine Learning care utilizează algoritmi de clasificare (de exemplu, JRip, PART, J48 și Random Forest).

În cazul unui rezultat pozitiv din stratul Cloud, set-top box-urile vor trimite notificări către toate dispozitivele mobile conectate la acestea. Folosind această abordare, asigurăm o scădere a nivelului de încărcare a rețelei de comunicație și un timp redus de răspuns la incident în cazul unui atac de tip smishing.

Mai mult decât atât, arhitectura noastră oferă o disponibilitate ridicată (de exemplu, 24/7) a sistemelor în caz de răspuns la incident. Implementăm tehnologia load-balancing astfel încât, dacă un set-top box eșuează, acesta este înlocuit automat cu altul. Un alt aspect important de care ne ocupăm este problema supraîncărcării rețelei. Când există prea multe solicitări către un

set-top box, unele dintre ele vor fi tratate automat de către altul. Aceasta ar aduce performanță și eficiență.

Comunicarea între straturi se realizează prin Internet, în timp ce dispozitivele din Edge pot comunica între ele folosind atât Internet-ul, cât și canale de comunicare ad-hoc bazate pe Bluetooth (de exemplu, Bluetooth low energy). În acest fel, sistemul garantează livrarea notificărilor care indică URL-uri infectate. Prin urmare, în cazul unui incident de securitate, un dispozitiv Edge le poate anunța rapid pe celelalte, astfel încât acestea să poată activa măsurile de protecție.

Fluxul de lucru este prezentat în Figura 5.2.

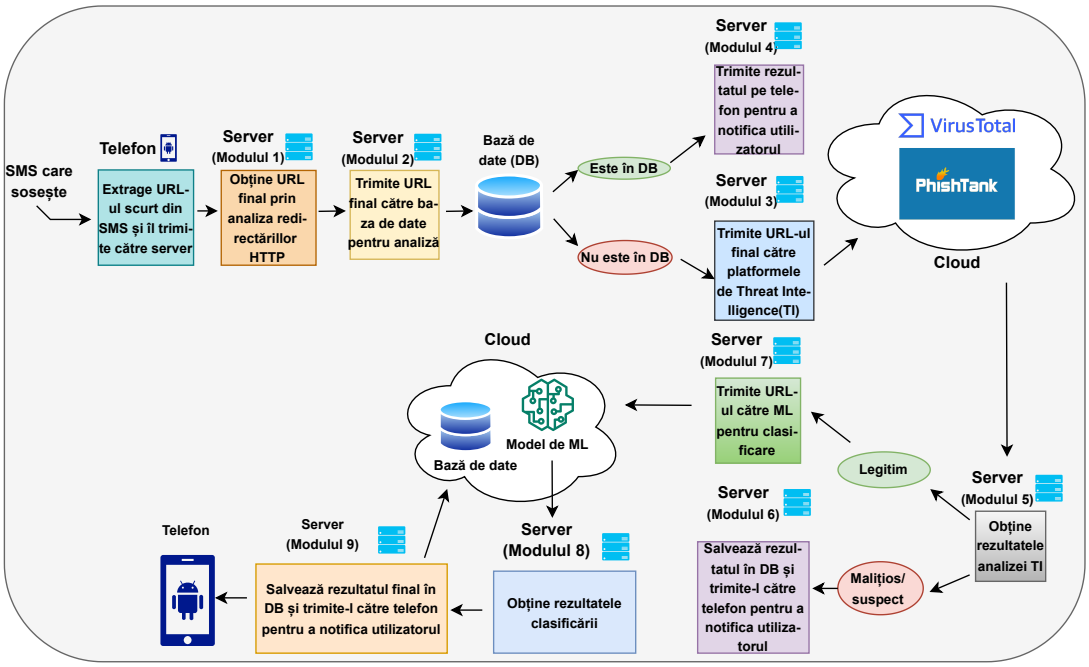

Figure 5.2. Fluxul de lucru al arhitecturii.

5.3 Implementarea Sistemului

În această secțiune, prezentăm detalii referitoare la faza de implementare a sistemului propus. Pentru a demonstra eficiența sistemului nostru, am efectuat mai multe teste cu diferite tipuri de mesaje SMS.

- SMS care conține o adresă URL malware conform datelor de Threat Intelligence - Utilizatorul primește un mesaj SMS cu o adresă URL scurtă. După obținerea URL-ului final, platformele de Threat Intelligence îl raportează ca fiind malware. Figurile 5.3 și 5.4 descriu acești pași.

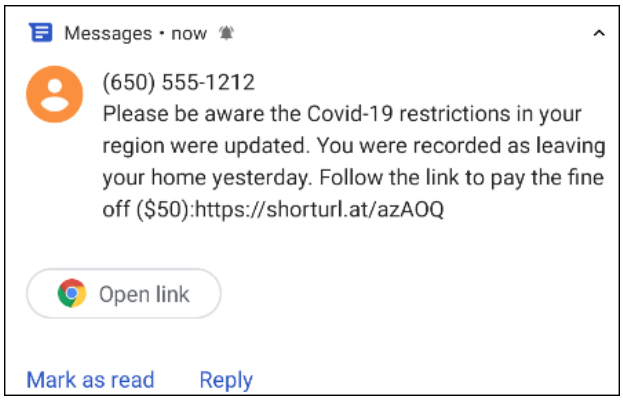

Figure 5.3. Primirea unui SMS ce conține un URL scurt care este malware.

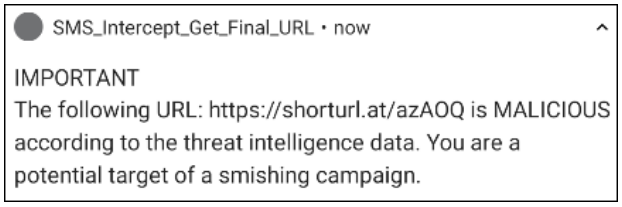

Figure 5.4. SMS clasificat cu succes ca fiind smishing.

- SMS care conține o adresă URL legitimă conform modelului nostru de ML - Utilizatorul primește un mesaj SMS cu o adresă URL scurtă. După obținerea URL-ului final, modelul de Machine Learning (ML) îl clasifică drept legitim. Figurile 5.5 și 5.6 prezintă aceste faze.

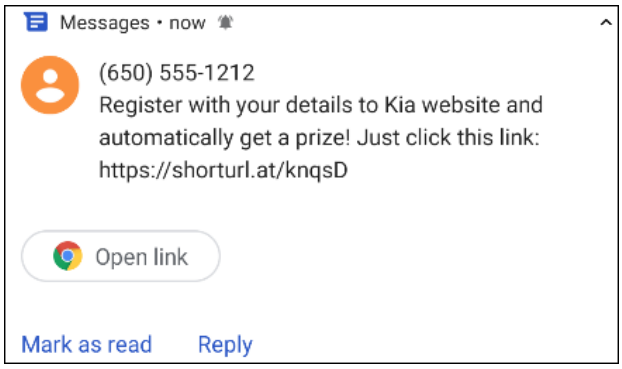

Figure 5.5. Primirea unui SMS ce conține un URL scurt care este legitim.

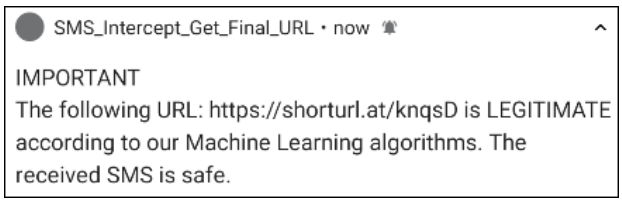

Figure 5.6. SMS clasificat cu succes ca fiind legitim.

- SMS care conține o adresă URL malițioasă conform modelului nostru de ML - Utilizatorul primește un mesaj SMS cu o adresă URL scurtă. După obținerea URL-ului final, modelul de Machine Learning (ML) îl clasifică ca fiind malițios. Figurile 5.7 și 5.8 prezintă acești pași.

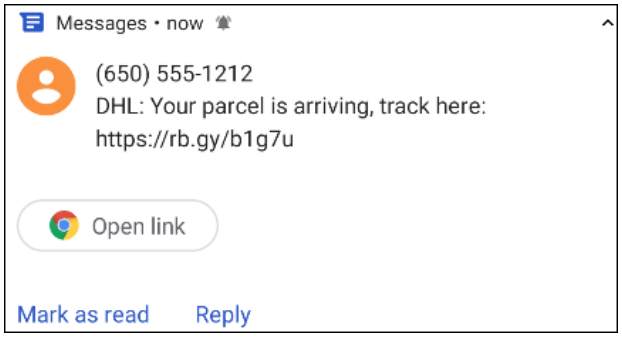

Figure 5.7. Primirea unui SMS ce conține un URL scurt care este malițios.

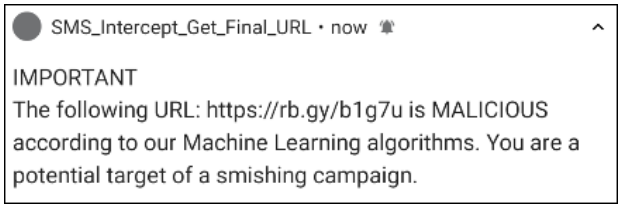

Figure 5.8. SMS clasificat cu succes ca fiind smishing.

5.4 Rezultate și Analize Experimentale

În această secțiune, descriem diferite valori de timp și performanță ale abordărilor de Threat Intelligence și Machine Learning.

Analiza Platformelor de Threat Intelligence

Am efectuat câteva teste empirice și am obținut rezultatele din tabelul 5.1 pentru cazul în care adresele URL nu se aflau în baza de date locală.

Table 5.1. Timpul de analiză al abordării de Threat Intelligence.

URL scurt	URL final	Timpul de analiză al platformelor de T.I. (s)	Reputația URL-ului
https://shorturl.at/fuxAE	https://call.raidstore.org/	81.267	Malițios + Time out primit de la adresa IP
https://t.ly/yEC	https://technology.macosevents.com/	2.394	Malițios + NXDOMAIN
https://rb.gy/w2i1u	https://press.infomapress.com/	1.889	Malițios + NXDOMAIN
https://rb.gy/wtkn	https://stiri.botosani.ro/	1.737	Legitim
https://tinyurl.com/kyc44ft8	https://www.bbc.com/	2.717	Legitim
https://cutt.ly/zwr2HIaS	https://www.amazon.com/b?node=21576558011 &ref_=alxcom_lrnmore_btn_23	2.864	Legitim
https://urlis.net/b94wy861	https://sucursalcentroapp.brizy.site/	2.681	Phishing + Online
http://tiny.cc/tzb8vz	https://kuccoiieeiinelovuiie.godaddysites.com/	2.499	Phishing + Online
https://url1.io/s/VpEwo	https://anime-info777.com/	3.879	Phishing + Online

Antrenarea și Clasificarea Modelelor de Învățare Automată

Pentru testul de viteză al proceselor combinate de extragere a caracteristicilor și clasificare, am ales zece adrese URL ce aparțin următoarelor categorii: phishing, legitim și necunoscut. Tabelul 5.2 prezintă rezultatele pe care le-am obținut. Se poate observa că timpul pentru cele două procese combinate poate varia, de la 2,8171 (s) până la 118,9281 (s), deoarece procesul de extragere a caracteristicilor depinde de factori precum dimensiunea conținutului site-ului web, timpul de răspuns al altor platforme interogate (de exemplu, Open PageRank pentru popularitatea domeniului, iPTY pentru vârsta domeniului).

Table 5.2. Durata proceselor de extragere a caracteristicilor și clasificare pentru diferite tipuri de adrese URL.

Operațiune	**Item**	**Durata pentru extragerea caracteristicilor + clasificare (s)**	**Durata pentru clasificare (s)**
Testarea unor adrese URL VALIDE de phishing	https://vxw2.mengzhan45.top/go/?to=https://pub-c479da8c0e2748d0a34fd7266d91fc30.r2.dev/index.html	7.8176	2.5128
	https://actividadbancaria–giovannyquint.repl.co/	6.7961	2.4780
	https://highlight-himself.toshibanetcam.com/	3.2873	2.4558
	https://dev-asistenonlibcr.pantheonsite.io/	2.8887	2.5539
	https://homeless-hospital.otzo.com/	4.6097	2.5590
	https://joint-knowledge.instanthq.com/	4.3833	2.7674
	https://department-depict.mrface.com/	4.2344	2.6300
	https://japanese-joint.qpoe.com/	4.9541	2.5126
	https://approximately-arab.mrbasic.com/	3.5512	2.5296
	https://administration-adopt.toythieves.com/	3.5949	2.4420
Testarea unor adrese URL POTENȚIALE de phishing	https://ameli-france-connect.com/	3.6744	2.6018
	https://portalemydati.online/	19.7659	2.4354
	http://portalemydati.online/	18.7333	2.5411
	http://infomydati.online/	19.9179	2.6577
	https://pagamenti.staffasestenza.co/8328-1/	2.8171	2.6222
	https://bafkreigfxcytcx7pfkvio4svcgeuwdpvdyey3fybfpckka7cxmhjcr6qbe.ipfs.dweb.link	8.1313	2.6082
	http://updatetan-sp.de/anmelden	4.1362	2.7528
	https://www.gefhuloa.com/pl/pl_dfertz/?uclick=9lpmgmvc&uclickhash=9lpmgmvc-9lpmgmvc-52ft-0-c81z-5m7vfe-5mbzi4-9e712d	3.2091	2.6611
	https://mail.kinepolis.com/optiext/optiextension.dll?ID=6yz6yKcj%2BELFjsP9CS7eIM_Z1YzAB9%2B5w3xcTWiaQ7cRM%2BVkYi5dQ9vpKT9vUN03dFBC0qCK46AQgv0vb_obTAikgCssL	3.5446	2.7708
	https://khatampanjereh.com/wp-includes/knab.php	7.4725	2.6577
Testarea unor adrese URL care NU SUNT de phishing	https://www.cloudflare.com/	8.9051	3.3585
	https://fonts.googleapis.com/css?family=Noticia+Text:400,400i,700,700i	4.5525	2.6238
	https://express.adobe.com	4.5663	3.0263
	http://www.paypal-merchant.com/	22.8981	2.5208
	https://www.mercadopago.com.br/	42.4390	2.4097
	https://onedrive.live.com/about/es-us-signin/	5.7948	2.5991
	http://url.zp.edu.ua/	4.8289	2.5182
	https://bitflyer.com/en-us/	11.8539	2.4980
	https://phishtank.org/phish_detail.php?phish_id=8153631	6.5013	2.9479
	https://quttera.com/	118.9281	2.44898

Acuratețea Algoritmilor de Clasificare Machine Learning

Am rulat următorii algoritmi de clasificare Machine Learning (ML) pe setul nostru de date: JRip, PART, J48 și Random Forest. În timp ce Random Forest depășește rezultatele obținute cu alți

clasificatori, am ales să evidențiem în Tabelul 5.3 metrici de evaluare suplimentare.

Table 5.3. Metrici de evaluare pentru clasificatorul Random Forest.

	Rata de TP	**Rata de FP**	**Precizia**	**Recall**	**F-Measure**	**MCC**	**Regiunea ROC**	**Regiunea PRC**	**Clasa**
	0.969	0.034	0.966	0.969	0.967	0.935	0.994	0.994	legitim
	0.966	0.031	0.969	0.966	0.967	0.935	0.994	0.994	phishing
Medie ponderată	0.967	0.033	0.967	0.967	0.967	0.935	0.994	0.994	

5.5 Concluzii

În acest capitol, descriem un nou model Cloud-Edge pentru detecția adreselor URL scurte care sunt malițioase în cadrul atacurilor de tip smishing. Proiectăm o platformă cu mai multe straturi care reduce supraîncărcarea rețelei de comunicații și garantează o latență scăzută și timp de răspuns rapid, precum și o disponibilitate ridicată prin tehnologia de load balancing. Arhitectura propusă folosește algoritmi de învățare automată (ML) și platforme open-source de Threat Intelligence (TI) pentru a detecta adrese URL scurte care sunt malițioase în cadrul atacurilor de tip smishing.

Sistemul nostru este conceput pentru dispozitive smartphone și acoperă diferite scenarii: URL-uri legitime, suspecte și malițioase care sunt scurtate. Rezultatele au arătat că adresele URL scurte malițioase pot fi identificate, iar SMS-ul corespunzător este clasificat ca smishing. Abordarea noastră are avantajul de a folosi instrumente gratuite și open-source care oferă un sistem eficient de detecție a adreselor URL scurte care sunt malițioase.

Mai mult decât atât, pentru un nivel avansat de detecție, modelul nostru de învățare automată și setul îmbunătățit de caracteristici identifică cu succes adresele URL care sunt malițioase, având o acuratețe de aproape 97% utilizând algoritmul Random Forest.

6 | Soluție de detecție a cyberbullying pentru fișiere multimedia folosind modele bazate pe Deep Learning

În acest capitol, discutăm două soluții pentru detecția cyberbullying-ului în fișiere multimedia folosind modele de tip Deep Learning. Prima soluție efectuează detecția în cadrul GIF-urilor, în timp ce cealaltă realizează aceeași sarcină pentru videoclipurile de pe TikTok. Prima soluție folosește o arhitectură hibridă care cuprinde o rețea neuronală convoluțională (CNN) și trei rețele neuronale recurente (RNNs). Rezultatele obținute dau o acuratețe de 99%. A doua soluție folosește un model bazat pe Transformers care funcționează pe mapări ale caracteristicilor ale rețelei neuronale convoluționale (CNN). Am evaluat eficiența modelului și am observat că am obținut o acuratețe de până la 100%.

Scopul acestei cercetări nu a fost acela de a defini contextul relativ la utilizator, deoarece ceea ce înseamnă cyberbullying pentru noi poate părea un comportament natural, normal pentru altcineva. Soluțiile pe care le-am propus identifică acțiunea care este reprezentată într-un fișier multimedia, iar noi asociem cyberbullying-ul cu violența. Clasificăm un fișier multimedia drept bullying atunci când acțiunea care este reprezentată denotă o formă de violență ce are loc în diferite condiții și care iese din contextele obișnuite sau normale (de exemplu, acțiunea de a împușca sau de a trage cu arma).

Partea de cyberbullying a fost una dintre directivele majore ale Uniunii Europene în programul Horizon în anul 2023 când acesta a fost propus, iar România a adoptat în domeniul de cercetare pe specializarea inteligentă, prevenirea cyberbullying-ului prin diverse metode. În strategia de cercetare pentru dezvoltarea inteligentă a României, la domeniul 6, punctul 6.4, se vorbește de detecția și prevenția cyberbullying-ului și crearea unui Internet mai sigur.

Conținutul acestui capitol se bazează pe publicarea lucrărilor *Bullying Detection Solution for GIFs Using a Deep Learning Approach* și *Cyberbullying Detection on TikTok Using a Deep Learning Approach* în *Information 2024* și textitSci. Bull. Univ. Politeh. Buchar.

Acest capitol este structurat după cum urmează: în Secțiunea 6.1, descriem sursele utilizate pentru a crea seturile de date pentru ambele sisteme și prezentăm datele utilizate în procesele de antrenare și testare. Secțiunea 6.2 evidențiază arhitectura sistemelor propuse. Secțiunea 6.3 prezintă rezultatele clasificării soluțiilor propuse. În cele din urmă, în Secțiunea 6.4, concluzionăm rezultatele soluțiilor propuse și identificăm viitoare oportunități de cercetare.

6.1 Colectarea și Etichetarea Datelor

În această secțiune, prezentăm sursele utilizate pentru colectarea datelor și modul în care acestea din urmă au fost folosite pentru procesele de antrenare și testare. Mai întâi evidențiem aceste

proceduri pentru sistemul utilizat în detecția bullying-ului pe fișiere GIF, iar apoi pentru cel utilizat în detecția cyberbullying-ului în videoclipurile de pe TikTok.

Colectarea și Etichetarea Datelor pentru Soluția de Detecție a Bullying-ului pe GIF-uri

Pentru a crea setul nostru de date, am folosit mai întâi unul ce se intitulează UCF101 (www.crcv.ucf.edu). Am luat videoclipuri ce aparțin doar următoarelor categorii: *Handstand Pushups*, *Pull Ups*, *Rowing* și *Kayaking*. Am combinat categoriile *Handstand Pushups* și *Pull Ups* în *Bodybuilding*, în timp ce pe cele de *Rowing* și *Kayaking* în *Water sports*. Aceste videoclipuri reprezintă activități umane obișnuite și pot fi clasificate ca materiale non-bullying.

În ceea ce privește fișierele media de bullying, am folosit GIPHY Scraper de la Scrapera (https://github.com/DarshanDeshpande/Scrapera) pentru a obține GIF-uri care sunt asociate cu activități de bullying. În aceast studiu, utilizăm fișiere GIF dinamice. Acestea sunt imagini reprezentate sub forma unei animații.

Setul nostru de date cuprinde 512 fișiere media pe care le-am colectat din sursele menționate anterior: setul de date UCF101 și Giphy. Am folosit 80% din date pentru antrenare și 20% pentru testare.

Etichetarea fișierelor media non-bullying a fost realizată în mod automat, deoarece acestea au fost preluate din setul de date UCF101, care conține videoclipuri deja etichetate pentru diferite acțiuni umane. Etichetarea fișierelor media de bullying a fost făcută manual de către autorii acestei lucrări și un expert extern.

Colectarea și Etichetarea Datelor pentru Detecția Cyberbullying-ului în Videoclipurile de pe TikTok

Pentru a crea setul nostru de date, am colectat videoclipuri de pe TikTok ce aparțin următoarelor categorii: *Baschet*, *Fotbal*, *Cântat la violoncel*, *Cântat la chitară*, *Împușcătură* și textitLovitură. Am etichetat ultimele două categorii ca fiind bullying, deoarece au un conținut care poate amenința, îngrozi sau îngrijora pe cineva care-l primește și vizionează un astfel de videoclip. Celelalte categorii au fost etichetate ca non-bullying. Am descărcat manual toate videoclipurile folosind platforma SnapTik (https://snaptik.app/). Toate videoclipurile din setul nostru de date au între 5.5 și 6.5 secunde. Dacă au fost inițial mai mari, le-am împărțit folosind site-ul web Clipchamp (https://app.clipchamp.com/) în fișiere media care să se încadreze în acea gamă de durate temporale.

80% din date au fost folosite pentru antrenare, în timp ce restul au fost utilizate pentru testare.

6.2 Prezentare Generală a Abordării Propuse

În această secțiune, prezentăm arhitectura sistemelor propuse pentru detecția bullying-ului în GIF-uri și videoclipuri de pe TikTok.

Prezentare Generală a Abordării Propuse pentru Soluția de Detecție a Bullying-ului pe GIF-uri

Propunem o soluție de detecție a bullying-ului în cadrul GIF-urilor. Folosim o arhitectură hibridă care cuprinde o rețea neuronală convoluțională (CNN) și trei rețele neuronale recurente (RNNs).

Această arhitectură învață reprezentarea GIF-urilor pentru a le clasifica în una dintre următoarele categorii: bullying și non-bullying.

Codul soluției propuse pentru detecția bullying-ului în cadrul fișierelor de tip GIF folosind o abordare de învățare adâncă se bazează pe următoarea implementare de pe GitHub (https://github.com/keras-team/keras-io/blob/master/examples/vision/video_classification.py).

Tabelul 6.1 compară arhitectura propusă cu una din literatura de specialitate pe care am luat-o de pe GitHub și am îmbunătățit-o.

Table 6.1. Comparație între soluția propusă și cea inițială

	Arhitectură	**Dimensiunea imaginii**	**Numărul de epoci**	**Generarea automată a datelor pentru antrenare și testare**
Soluția inițială de pe GitHub	1 CNN + 1 RNN	224	10	Nu
Soluția propusă și îmbunătățită*	1 CNN + 3 RNNs	169	50	Da

Mai jos discutăm principalele componente ale arhitecturii, apoi prezentăm modul în care aceasta funcționează.

Arhitectura Generală a Sistemului

Figura 6.1 prezintă arhitectura generală a sistemului.

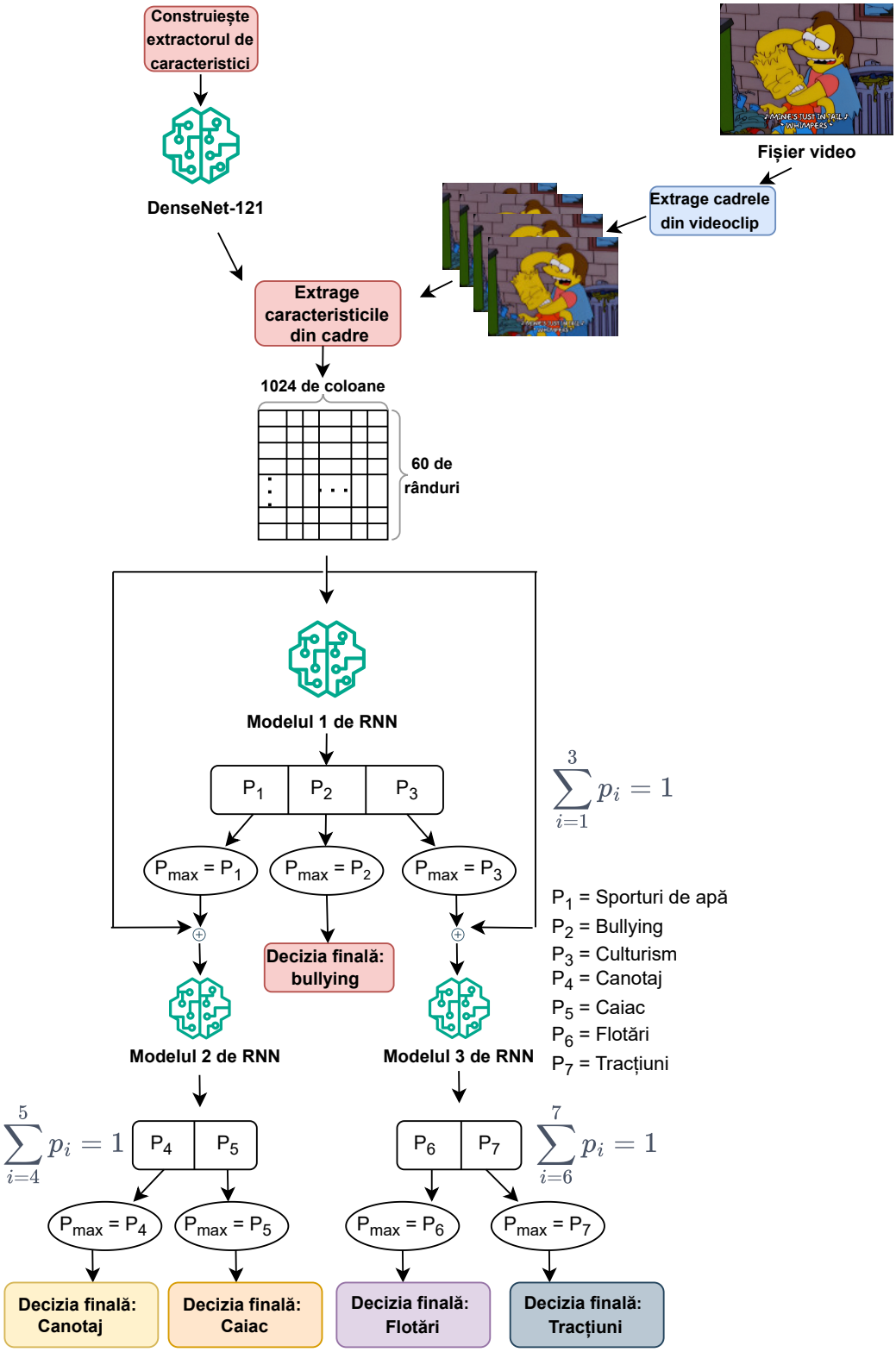

Figure 6.1. Prezentare generală a abordării propuse.

În primul rând, este construită o funcționalitate care extrage caracteristicile pentru a obține trăsăturile din fiecare frame al unui videoclip. Aceasta se bazează pe o rețea DenseNet-121. În al doilea rând, fiecare fișier video din setul de date este descompus în frame-uri care trec prin funcționalitatea de extragere a caracteristicilor astfel încât să fie obținute trăsăturile lor. Acestea sunt stocate într-o matrice de 1024 de coloane și 20 de rânduri. Prima dimensiune, 1024, reprezintă numărul de caracteristici pe care le utilizează, în timp ce cealaltă dimensiune, 20, reprezintă numărul de frame-uri pe care le ia de la fiecare videoclip. Această matrice reprezintă intrarea pentru primul model RNN care a fost antrenat pe întregul set de date care cuprinde videoclipuri din categoriile *bullying*, *Sporturi de apă* și *Culturism*. Dacă cea mai mare probabilitate este P_2, atunci fișierul video evaluat este clasificat ca *bullying*. În caz contrar, dacă cea mai mare probabilitate este P_1, atunci matricea caracteristicilor este utilizată în continuare ca intrare pentru al doilea model RNN care a fost antrenat pe grupul *Sporturi de apă*. În acest caz, rezultatul modelului va fi o matrice care conține două probabilități care corespund categoriilor *Canotaj* și *Caiac*. Pe de altă parte, dacă cea mai mare probabilitate este P_3, atunci matricea de caracteristici este utilizată în continuare ca intrare pentru al treilea model RNN care a fost antrenat pe grupul *Culturism*. Pentru această situație, rezultatul modelului va fi o matrice de două probabilități care corespund categoriilor *Flotări* și *Tracțiuni*.

Prezentare Generală a Abordării Propuse pentru Detecția Cyberbullying-ului în Videoclipurile de pe TikTok

Noi propunem o soluție de detecție a cyberbullying-ului pentru videoclipurile de pe TikTok folosind o abordare bazată pe algoritmi de Deep Learning. Folosim un model bazat pe Transformer care funcționează pe mapări de caracteristici ale rețelei neuronale convoluționale (CNN). Mai mult decât atât, folosim modelul DenseNet121 pre-antrenat pe setul de date ImageNet-1k.

Programul pe care l-am creat pentru clasificarea videoclipurilor de pe TikTok folosind o abordare de învățare adâncă se bazează pe următoarea implementare de pe GitHub (https://github.com/keras-team/keras-io/blob/master/examples/vision/video_transformers.py).

Tabelul 6.2 compară arhitectura propusă cu una din literatura de specialitate pe care am luat-o de pe GitHub și am îmbunătățit-o.

Table 6.2. Comparație între soluția propusă și cea inițială

	Modelul pentru extragerea caracteristicilor	**Modelul de clasificare**	**Numărul maxim de cadre**	**Dimensiunea imaginii**	**Numărul de epoci**	**Generarea automată a datelor pentru antrenare și testare**
Soluția inițială de pe GitHub	DenseNet-121	Transformers-based	20	128	5	Nu
Soluția propusă și îmbunătățită*	DenseNet-121	Transformers-based	30	500	20	Da

Arhitectura generală a sistemului este prezentată în Figura 6.2.

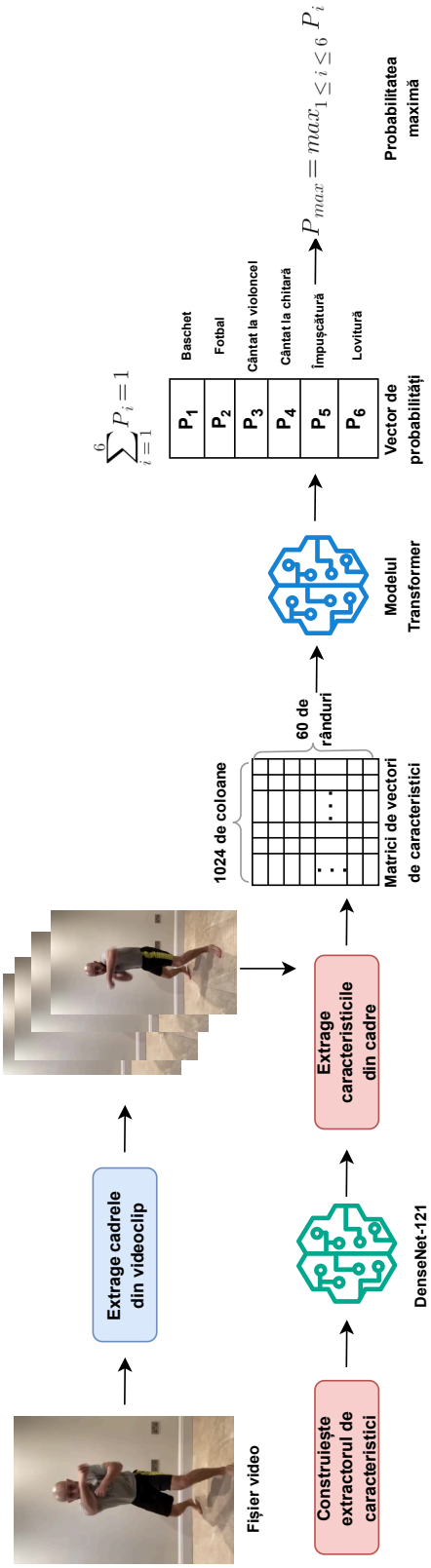

Figure 6.2. Arhitectura generală a sistemului.

După încărcarea frame-urilor din videoclipuri, folosim o rețea neuronală convoluțională (CNN) pentru a extrage caracteristicile. Ca să realizăm această sarcină, am ales o arhitectură de tip DenseNet. În cazul nostru, intrările pentru modelul DenseNet, și anume DenseNet121, sunt reprezentate de imagini de dimensiunea 500x500, iar ieșirea pentru fiecare imagine este un vector de caracteristici cu dimensiunea de 1024. Acesta din urmă va reprezenta intrarea pentru modelul Transformer. Pentru a obține rezultatul de la CNN, nu am inclus și clasificatorul ci doar partea care se ocupă de extragerea caracteristicilor. Numărul asociat modelului indică câte straturi are modelul, în acest caz fiind 121.

După ce am obținut caracteristicile cu ajutorul DenseNet121, am folosit o arhitectură Transformer pentru a crea modelul nostru de detecție pentru bullying.

6.3 Rezultatele Clasificării

În această secțiune, prezentăm rezultatele clasificării obținute de sistemele de detecție a bullying-ului pe care le-am propus. Una dintre platforme vizează fișierele GIF, în timp ce cealaltă se ocupă de videoclipurile de pe TikTok.

Rezultatele Clasificării pentru Soluția de Detecție a Bullying-ului pe GIF-uri

Am comparat soluția propusă care conține un model CNN și trei modele RNN, cu cea care este formată dintr-un model CNN și unul RNN. Tabelul 6.3 evidențiază modul în care soluția pe care am proiectat-o este mult mai eficientă decât cealaltă cu care am comparat-o, în ceea ce privește diferitele valori de performanță. În tabel, am codificat numele fiecărei categorii într-un simbol, după cum urmează: *Bullying* în *(B)*, *Culturism* în *(BB)*, *Sporturi de apă* în *(WS)*, *Caiac* în *(K)*, *Canotaj* în *(R)*, *Tracțiuni* în *(Pull)* și *Flotări* în *(Push)*. Mai mult decât atât, în tabel sunt afișate mai întâi metrici de performanță pentru cele trei modele RNN propuse (adică, Modelul RNN propus nr. 1, Modelul RNN propus nr. 2, Modelul RNN propus nr. 3) și, în final, sunt descrise rezultatele pentru modelul care utilizează un singur model RNN (adică, RNN simplu). Arhitectura care folosește un model CNN și un model RNN are următoarele cinci clase: *Bullying*, *Caiac*, *Canotaj*, *Tracțiuni* și *Flotări*.

Table 6.3. Metrici de performanță pentru modelele RNN.

Modelul RNN nr.	Acuratețe	Precizie	Recall	F1-Score
*Modelul RNN propus nr. 1	99.02%	95.24% (B) 100% (BB) 100% (WS)	100% (B) 97.37% (BB) 100% (WS)	97.56% (B) 98.66% (BB) 100% (WS)
*Modelul RNN propus nr. 2	97.7%	95.65% (K) 100% (R)	100% (K) 95.45% (R)	97.77% (K) 97.67% (R)
*Modelul RNN propus nr. 3	100%	100% (Pull) 100% (Push)	100% (Pull) 100% (Push)	100% (Pull) 100% (Push)
RNN simplu	51.96%	64% (B) 48.39% (K) 51.28% (Pull) 0% (Push) 28.57% (R)	80% (B) 68.18% (K) 100% (Pull) 0% (Push) 10% (R)	71% (B) 56.61% (K) 67.8% (Pull) 0% (Push) 14.82% (R)

Rezultatele Clasificării pentru Detecția Cyberbullying-ului în Videoclipurile de pe TikTok

Am propus un sistem de detecție a cyberbullying-ului în videoclipurile de pe TikTok, folosind o abordare care se bazează pe algoritmi de Deep Learning. Soluția noastră folosește un model bazat pe Transformer care funcționează pe zone de caracteristici ale rețelei neuronale convoluționale (CNN). Tabelul 6.4 descrie experimentele pe care le-am efectuat folosind modelul nostru bazat pe algoritmi de Deep Learning și acuratețea obținută. Câmpul *MAX_SEQ_LENGTH* se referă la numărul maxim de cadre pe care le extragem din fiecare videoclip. Când un număr de cadre din videoclip este mai mic decât valoarea acestui câmp, completăm numărul de cadre cu zerouri. Câmpul *NUM_FEATURES* reprezintă numărul de caracteristici extrase de modelul CNN (adică DenseNet-121) din fiecare videoclip. Câmpul *IMG_SIZE* se referă la dimensiunile matricei ce este extrasă din centrul fiecărui cadru. În cazul nostru, acesta va fi de 500x500 pixeli.

Table 6.4. Rezultatele clasificării

Aplicație Keras	MAX_SEQ_LENGTH	NUM_FEATURES	IMG_SIZE	Nr. epoci	Acuratețe
DenseNet-121	30	1024	500	20	Până la 100%

Unele rezultate pe care le-am obținut în timpul unuia dintre experimentele pe care le-am condus sunt prezentate în tabelul 6.5. Ele evidențiază acuratețea și valorile de pierdere ale fiecărei epoci pentru datele de antrenare și validare. Câmpurile *Pierdere* și *Acuratețe* se referă la antrenare, în timp ce *Pierdere de validare* și *Acuratețe de validare* sunt legate de procesul de validare. Putem spune că acuratețea soluției propuse este de până la 100% deoarece am obținut pentru epoca 14 o acuratețe asupra datelor de validare de 100%.

Table 6.5. Valorile de pierdere și acuratețe pentru fiecare epocă

Nr. epocă	Pierdere	Acuratețe	Pierdere de validare	Acuratețe de validare
1/20	3.4658	0.4251	8.8870	0.0000e+00
2/20	1.3992	0.6232	3.0966	0.0000e+00
3/20	0.5996	0.8213	2.0553	0.1081
4/20	0.2851	0.9058	0.9450	0.6081
5/20	0.2175	0.9251	2.0311	0.3514
6/20	0.2009	0.9469	1.6759	0.4730
7/20	0.1240	0.9638	0.2524	0.8649
8/20	0.0453	0.9831	3.1626	0.3108
9/20	0.0217	0.9928	0.2196	0.9189
10/20	0.0284	0.9879	1.2308	0.6351
11/20	0.0178	0.9952	0.6859	0.7432
12/20	0.0186	0.9976	0.7603	0.7568
13/20	0.0319	0.9928	0.7513	0.7297
14/20	0.0393	0.9903	0.0052	1.0000
15/20	0.1854	0.9444	3.9171	0.4595
16/20	0.0410	0.9831	0.6692	0.8243
17/20	0.0333	0.9879	1.1296	0.7162
18/20	0.0597	0.9734	1.4036	0.6892
19/20	0.0425	0.9831	1.0926	0.7568
20/20	0.0168	0.9928	2.6229	0.5811

Când testăm soluția noastră cu un videoclip al cărui conținut se referă la acțiunea de a împușca, obținem următoarele rezultate remarcabile pe care le-am prezentat în tabelul 6.6. Figura 6.3 arată un cadru din acel videoclip.

Figure 6.3. Videoclip de test cu conținut de bullying (a împușca).

Table 6.6. Rezultatele videoclipului de test (a împușca)

Clasă	Probabilitate
Shoot	100.00%
Basketball	0.00%
Football	0.00%
Kick	0.00%
Playing guitar	0.00%
Playing cello	0.00%

Tabelul 6.7 compară acuratețea arhitecturii propuse de noi cu cea a sistemului din literatură pe care l-am luat de pe GitHub și l-am îmbunătățit.

Table 6.7. Comparație între soluția propusă și cea inițială în ceea ce privește acuratețea

	Acuratețe
Soluția inițială de pe GitHub	67.5%
Soluția propusă și îmbunătățită*	Până la 100%

6.4 Concluzii

În acest capitol discutăm două soluții pentru detecția cyberbullying-ului în fișiere multimedia. Prima se ocupă de fișiere GIF, în timp ce cealaltă tratează videoclipurile de pe TikTok.

În ceea ce privește primul sistem, creăm un set de date cu GIF-uri ce conțin acțiuni care denotă comportamentul de bullying. Cu ajutorul unui instrument de web scrapping, am luat GIF-uri referitoare la bullying de pe platforma GIPHY și le-am filtrat până au rămas cele mai relevante. Acuratețea sistemului propus este de 99%. Acesta poate clasifica GIF-urile în *bullying*, *Culturism* și *Sporturi de apă*. Mai mult decât atât, soluția noastră poate clasifica în continuare fișierele din ultimele două categorii. Primele pot fi clasificate mai departe în *Tracțiuni* sau *Flotări*, în timp ce cele din urmă în *Canotaj* sau *Caiac*.

În ceea ce privește cel de-al doilea sistem, propunem o soluție nouă de detecție a cyberbullying-ului pentru videoclipurile de pe TikTok, folosind un model bazat pe Deep Learning. Creăm un

set de date ce conține videoclipuri de pe TikTok, atât cu conținut de tip bullying, cât și non-bullying. Fișierele video au fost descărcate manual cu ajutorul platformei SnapTik și au fost procesate folosind site-ul Clipchamp. În plus, creăm un sistem care utilizează un model bazat pe Transformer pentru clasificarea videoclipurilor. Acesta funcționează pe zonele de caracteristici generate de rețeaua neuronală convoluțională (CNN). Am evaluat modelul în raport cu setul de date creat și am obținut o acuratețe de până la 100%.

Putem concluziona că soluțiile noastre reprezintă un pas înainte în cercetarea și dezvoltarea sistemelor de securitate, în special pentru diminuarea atacurilor de bullying prin GIF-uri și videoclipuri de pe TikTok.

Posibilele cazuri de utilizare care ar putea beneficia de pe urma propunerii noastre includ bullying-ul din mediul online sau hărțuirea în medii publice sau private, cum ar fi universități, spitale, hoteluri sau corporații. Propunerile noastre aduc un plus valoare utilizatorilor de pe Internet, oferind capabilități avansate de detecție a bullying-ului în GIF-uri și videoclipuri de pe TikTok.

7 | Sisteme de tip Endpoint Detection and Response de generația următoare bazate pe agenți pentru detecția amenințărilor de securitate cibernetică

În acest capitol, discutăm despre sisteme de tip Endpoint Detection and Response de generația următoare bazate pe agenți pentru detecția amenințărilor de securitate cibernetică (NextEDR). Este o platformă inovatoare și interactivă de tip Cloud-Edge-Continuum Endpoint Detection and Response pentru protejarea organizațiilor moderne de atacurile de securitate cibernetică. Proiectăm un Proof-of-Concept bazat pe o soluție interactivă de comunicare de tip (ChatBot) pentru detecția phishing-ului în adrese URL scurte. Soluția noastră este o platformă multistrat centrată pe mobil și bazată pe modelul Cloud-Edge-Continuum.

Acest proiect pe care l-am propus face parte dintr-un program Proof of Concept (programul *Invest National University of Science and Technology POLITEHNICA Bucharest Proof of Concept (Invest PoC)*) și am lucrat la elaborarea unui plan de afaceri.

Conținutul acestui capitol se bazează pe publicarea lucrării *NextEDR - Next Generation Agent-Based EDR Systems for Cybersecurity Threats* în *2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*. Acest capitol este structurat după cum urmează: în Secțiunea 7.1 propunem platforma *NextEDR* și arhitectura sa. Secțiunea 7.2 prezintă rezultatele clasificării obținute. În cele din urmă, în Secțiunea 7.3 descriem concluziile și viitoarele direcții de cercetare.

7.1 Arhitectura NextEDR

În această secțiune, prezentăm arhitectura soluției propuse. Acesta este descrisă în Figura 7.1.

Propunerea noastră include 4 straturi: (i) stratul surselor de date; (ii) stratul Edge; (iii) stratul Cloud; și (iv) stratul Aplicație. În primul rând, stratul surselor de date include informații care provin de la e-mailuri, SMS-uri, aplicații de mesagerie instantă (de exemplu, WhatsApp, Signal etc.) și rețele sociale (de exemplu, Facebook, Instagram etc.) de pe dispozitivele mobile. Stratul Edge extrage adresa URL scurtă din sursele de date, o convertește în forma originală și o trimite la stratul Cloud prin interfața de ChatBot. Acolo aceasta este procesată prin utilizarea a două mecanisme: (1) prin platforme de Threat Intelligence (de exemplu, VirusTotal și PhishTank); (2) prin algoritmi de ML (de exemplu, JRip, PART, J48 și Random Forest). În plus, introducem în stratul Cloud un nou agent conversațional interactiv. Principalele funcționalități comportamentale

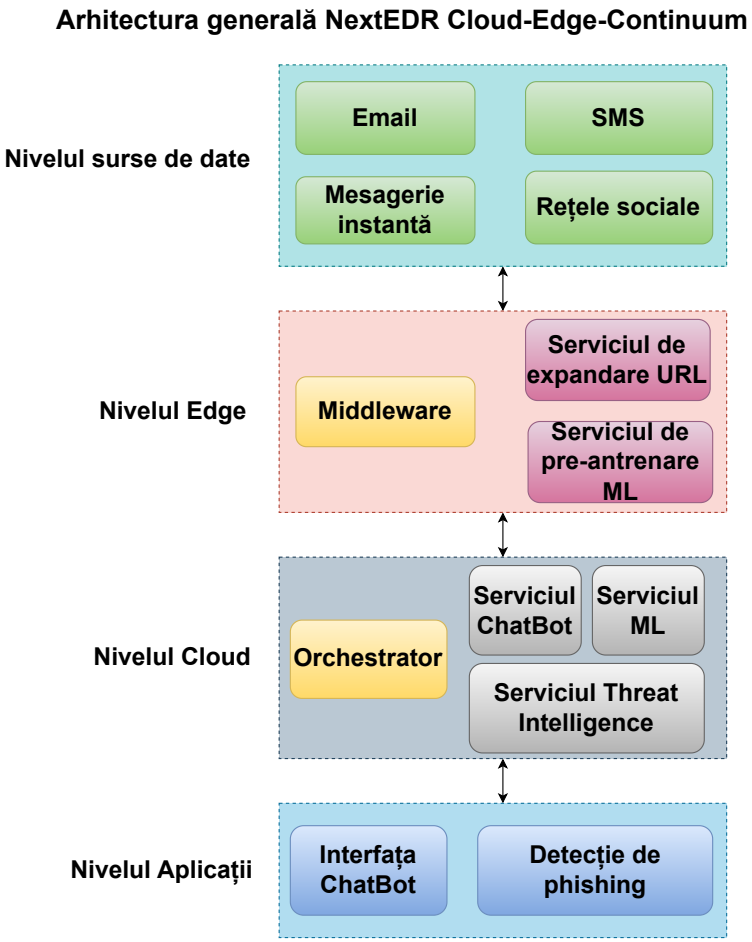

Figure 7.1. Arhitectura generală a *NextEDR*.

ale ChatBot-ului sunt: verificarea URL-ului și antrenarea platformei prin raportarea URL-urilor malițioase. Utilizatorul final folosește serviciul ChatBot pentru a trimite o adresă URL pentru verificare. ChatBot are 3 comportamente: (a) pozitiv; (b) negativ; sau (c) necunoscut. Pentru a facilita raportarea corectă, platforma *NextEDR* implementează mecanisme de recompensă. Mai mult decât atât, în cazul unui rezultat pozitiv, stratul Cloud trimite notificări prin serviciul de alertă către toate dispozitivele mobile din platformă. Acest lucru asigură o reducere a nivelului de încărcare a rețelei de comunicații și un timp redus de răspuns la incident în cazul unui atac de tip phishing.

Întotdeauna, între ce avem în nivelurile de Edge și Cloud, acolo unde se află serviciile de expansiune URL și pre-antrenare a modelului de învățare automată (ML), pot fi aduse modele mai bune. La fel și pe partea de Orchestrator, acolo unde mergem pe partea de servicii de ML și Threat Intelligence, totdeauna vor fi modele mai bune care vor schimba aceste două straturi, rămânând sursele și aplicația nemodificate. Spre exemplu, cineva ar folosi această tehnică pentru a opri automat, pentru copii sau pentru anumite persoane, partea de cyberbullying. Ea funcționează în continuare și putem aduce actualizări software-ului din Middleware și Orchestrator, iar atunci acest lucru poate fi parte din bussiness plan.

Ca direcții viitoare, metoda noastră nu presupune schimbarea nivelului de aplicație sau a surselor de date. Modelele se vor adapta la sursele de date și la aplicație, ceea ce ar fi un avantaj.

7.2 Rezultate și Analize Experimentale

În această secțiune, prezentăm rezultatele obținute. Tabelul 7.1 prezintă timpul de analiză al mecanismului de Threat Intelligence pentru diferite adrese URL scurte.

Table 7.1. Timpul de analiză al mecanismului de Threat Intelligence.

URL Scurt	URL Final	Timpul T.I. (s)
https://shorturl.at/fuxAE	https://call.raidstore.org/	81.267
https://t.ly/yEC	https://technology.macosevents.com/	2.394
https://rb.gy/w2i1u	https://press.infomapress.com/	1.889
https://rb.gy/wtkn	https://stiri.botosani.ro/	1.737
https://tinyurl.com/kyc44ft8	https://www.bbc.com/	2.717
https://cutt.ly/zwr2HIaS	https://www.amazon.com/b?node=21576558011 &ref_=alxcom_lrnmore_btn_23	2.864
https://urlis.net/b94wy861	https://sucursalcentroapp.brizy.site/	2.681
http://tiny.cc/tzb8vz	https://kuccoiieeiinelovuiie.godaddysites.com/	2.499
https://url1.io/s/VpEwo	https://anime-info777.com/	3.879

Pentru testul de viteză al proceselor combinate de extragere a caracteristicilor și clasificare ML, am ales zece adrese URL pentru fiecare din următoarele categorii: phishing, legitim și necunoscut. Tabelul 7.2 prezintă rezultatele pe care le-am obținut. Se poate observa că timpul de clasificare al modelului de ML variază între 2.4420 și 3.3585 secunde.

Table 7.2. Durata clasificării modelului de ML.

Operațiune	**Item**	**Durata clasificării ML (s)**
Testarea unor adrese URL VALIDE de phishing	https://pub-c479da8c0e2748d0a34fd7266d91fc30.r2.dev/index.html	2.5128
	https://actividadbancaria–giovannyquint.repl.co/	2.4780
	https://highlight-himself.toshibanetcam.com/	2.4558
	https://dev-asistenonlibcr.pantheonsite.io/	2.5539
	https://homeless-hospital.otzo.com/	2.5590
	https://joint-knowledge.instanthq.com/	2.7674
	https://department-depict.mrface.com/	2.6300
	https://japanese-joint.qpoe.com/	2.5126
	https://approximately-arab.mrbasic.com/	2.5296
	https://administration-adopt.toythieves.com/	2.4420
Testarea unor adrese URL POTENȚIALE de phishing	https://ameli-france-connect.com/	2.6018
	https://portalemydati.online/	2.4354
	http://portalemydati.online/	2.5411
	http://infomydati.online/	2.6577
	https://pagamenti.staffasestenza.co/8328-1/	2.6222
	https://ckka7cxmhjcr6qbe.ipfs.dweb.link/	2.6082
	http://updatetan-sp.de/anmelden	2.7528
	https://www.gefhuloa.com/pl/pl_dfertz/?uclick=9lpmg9e712d	2.6611
	https://mail.kinepolis.com/optiext/optiextension.dll?ID=6yz6yKcj	2.7708
	https://khatampanjereh.com/wp-includes/knab.php	2.6577
Testarea unor adrese URL care NU SUNT de phishing	https://www.cloudflare.com/	3.3585
	https://fonts.googleapis.com/css?family=Noticia+Text:400,400i,700,700i	2.6238
	https://express.adobe.com/	3.0263
	http://www.paypal-merchant.com/	2.5208
	https://www.mercadopago.com.br/	2.4097
	https://onedrive.live.com/about/es-us-signin/	2.5991
	http://url.zp.edu.ua/	2.5182
	https://bitflyer.com/en-us/	2.4980
	https://phishtank.org/phish_detail.php?phish_id=8153631	2.9479
	https://quttera.com/	2.4490

Implementarea Sistemului

La acest proiect au lucrat patru persoane: eu și conducătorii mei științifici. Eu am ocupat postul de lider de livrare, în timp ce ceilalți au avut funcții de director și mentor academic.

Noi am lansat o versiune beta a sistemului nostru EDR pentru un grup atent selectat de utilizatori finali. Pe parcursul fazei de testare beta, am colectat feedback detaliat prin diverse metode, inclusiv sondaje, interviuri și observații directe ale utilizatorilor. Ne-am concentrat pe înțelegerea experiențelor pe care aceștia le-au avut în urma interacțiunii cu sistemul, pe identificarea oricăror puncte slabe și pe adunarea de sugestii de îmbunătățire.

Am monitorizat îndeaproape valorile cheie ale rezultatelor obținute, cum ar fi acuratețea de detecție, timpii de răspuns, rata de false-positive și gradul de utilizare al sistemului. Aceste date au oferit dovezi cuantificabile ale eficienții sistemului și domeniilor care necesită îmbunătățiri. Pe baza feedback-ului și a valorilor de performanță, am adus îmbunătățiri iterative sistemului. Această abordare agilă ne-a permis să adresăm rapid orice problemă și să perfecționăm produsul

pentru a răspunde mai bine așteptărilor utilizatorilor.

Conform modelului de afaceri al proiectului NextEDR, segmentele de potențiali clienți pentru sistemul nostru includ întreprinderi, companii din domeniul tehnologic, industria comunitară, instituții de învățământ și furnizori de servicii gestionate. Ne-am putea face clienții fericiți și mulțumiți de soluția noastră, oferind protecție îmbunătățită de securitate cibernetică, protecție personalizată și adaptabilă, o soluție rentabilă și scalabilă, precum și accesibilitate și o interfață ușor de utilizat. Figura 7.2 descrie planul de afaceri model canvas pentru platforma NextEDR.

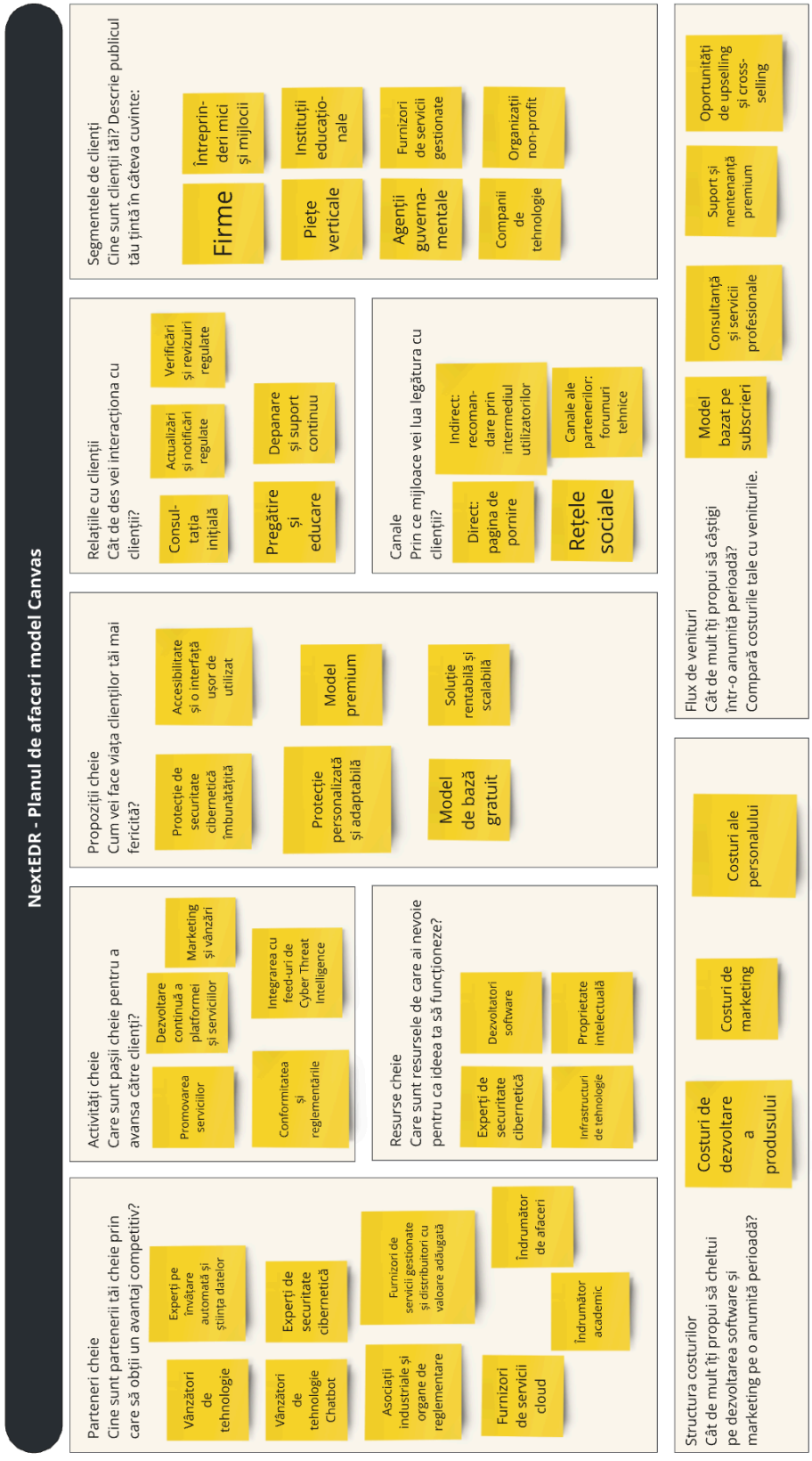

Figure 7.2. NextEDR - Planul de afaceri model Canvas.

Validarea Sistemului pe baza unui Caz Real

Ne-am confruntat cu o situație reală când am primit pe smartphone-urile noastre un mesaj SMS care părea că ar fi fost trimis de către Poșta Română. Mesajul SMS afirma că datele referitoare la adresa de livrare a coletului s-au pierdut și că aceste informații trebuie actualizate accesând o adresă URL care a fost menționată. Mesajul SMS poate fi văzut în Figura 7.3.

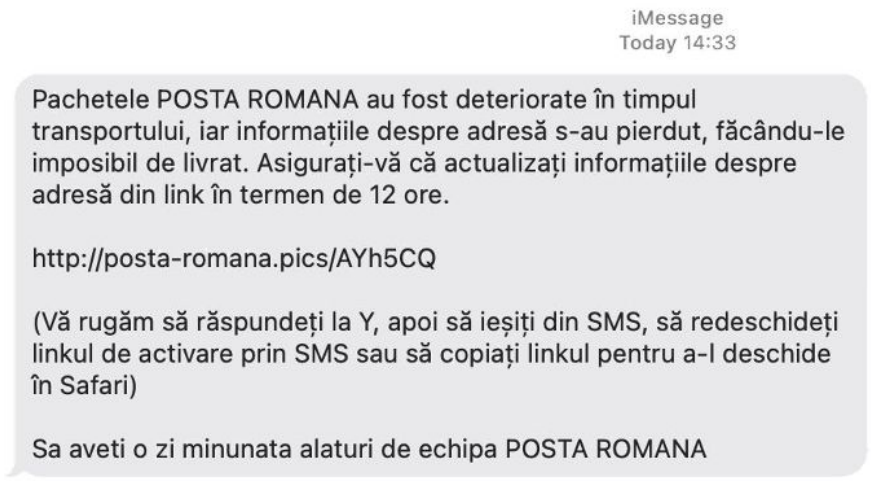

Figure 7.3. Mesaj SMS suspect ce a fost recepționat.

Adresa URL din mesajul SMS părea suspectă și am decis să-i verificăm legitimitatea utilizând sistemul EDR. Figura 7.4 prezintă rezultatele analizei.

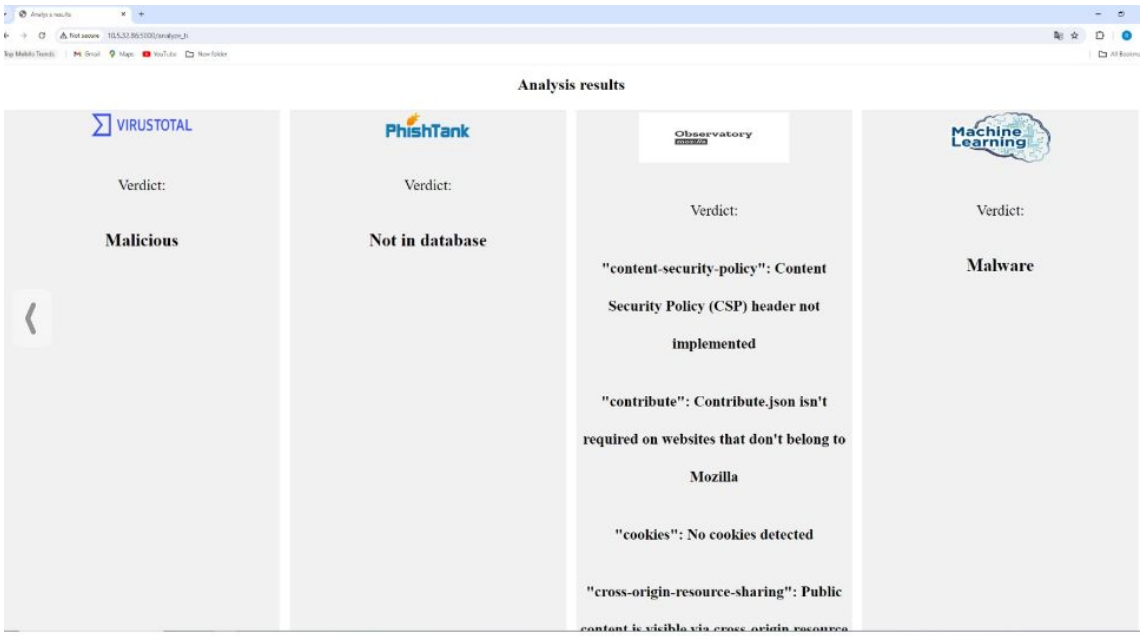

Figure 7.4. Analiza SMS-ului suspect folosind NextEDR.

Se pare că platforma VirusTotal și modelul de Machine Learning au clasificat adresa URL ca

fiind malițioasă. PhishTank nu a identificat adresa URL ca fiind raportată pentru phishing, cel mai probabil pentru că acea campanie era în desfășurare și era foarte recentă. Rezultatele generate de Mozilla Observatory relevă configurările greșite de securitate ale site-ului către care indică URL-ul analizat.

7.3 Concluzii

În această lucrare, propunem o platformă EDR de tip Cloud-Edge-Continuum de generația următoare care avansează la un nivel superior al soluțiilor de securitate cibernetică prin agenți conversaționali inteligenți, numiți ChatBoots.

Această soluție a câștigat premiul I la concursul de proiecte *Invest UNSTPB Proof of Concept*. Dezvoltarea produsului minim viabil (MVP) pentru platforma NextEDR a implicat o rafinare iterativă bazată pe feedback-ul utilizatorilor, acordând prioritate funcțiilor esențiale pentru a oferi o soluție funcțională care să răspundă nevoilor principale ale utilizatorilor în mod eficient.

Având în vedere design-ul bazat pe modelul Cloud-Edge-Continuum, soluția noastră reduce gradul de încărcare al rețelei de comunicații și garantează o latență scăzută și timp de răspuns rapid, precum și o disponibilitate ridicată prin tehnologia de load balancing. Sistemul nostru este conceput pentru dispozitive smartphone și acoperă diferite scenarii: URL-uri legitime, suspecte și malițioase care sunt scurtate.

Rezultatele obținute au arătat că adresele URL scurte care sunt malițioase pot fi identificate, iar SMS-ul corespunzător este clasificat ca smishing. Abordarea noastră are avantajul de a folosi instrumente gratuite și open-source care oferă un sistem eficient de detecție a adreselor URL scurte care sunt malițioase.

8 | Metodologia de Guvernare a Securității Cibernetice pentru Infrastructurile la Scară Largă

În acest capitol, prezentăm un framework de evaluare a riscului de securitate cibernetică în cadrul infrastructurilor la scară largă, împreună cu tehnici de reducere a atacurilor și potențiale contramăsuri. Metodologia pe care o folosim în studiul nostru este una calitativă. Aceasta se bazează pe 66 de proiecte din baza de date CORDIS a Comisiei Europene ce fac referire la guvernanța securității cibernetice în cadrul orașelor inteligente.

Din analiza acestor proiecte, am extras ideile care definesc guvernanța pe domeniul securității cibernetice și diferitele metodologii de prevenție pentru amenințările de securitate cum ar fi cyberbullying-ul, adresele URL scurte care sunt malițioase, atacurile de tip phishing, etc.

Conținutul acestui capitol se bazează pe publicarea lucrării *Cybersecurity Governance in Large-Scale Infrastructures* în *Romanian Journal of Information Technology and Automatic Control*.

Acest capitol este structurat după cum urmează: Secţiunea 8.1 prezintă o introducere în domeniul problemei studiate. În Secțiunea 8.2, prezentăm metodologia propusă, în timp ce principalele amenințări și potențiale contramăsuri sunt descrise în Secțiunea 8.3. În cele din urmă, în Secțiunea 8.4, descriem rezultatele obținute cu ajutorul framework-ului pe care l-am propus și identificăm viitoare oportunități de cercetare.

8.1 Introducere

Multe orașe din întreaga lume riscă să se confrunte cu probleme legate de condițiile de viață, deoarece au probleme importante privind securitatea, scalabilitatea și starea infrastructurilor. Acest lucru se datorează creșterii populației care va ajunge la 9,8 miliarde în 2050 [11]. Ca urmare, mediul urban va întâmpina atât provocări, cât și beneficii. Unele dintre dificultățile cu care s-ar confrunta sunt reprezentate de faptul că sectorul educațional și cel al sănătății vor avea nevoie de noi abordări, economia va avea probleme, consumul de energie va crește, siguranța publică se va confrunta cu noi riscuri, iar posibilitatea unor atacuri cibernetice împotriva orașele este ridicată. Soluția cheie pentru aceste probleme o reprezintă infrastructurile inovatoare, scalabile și eficiente din punct de vedere al costurilor [5].

În această lucrare, propunem un model de evaluare a riscului de securitate cibernetică în infrastructurile la scară largă, împreună cu tehnici de reducere a atacurilor și potențiale contramăsuri. Metodologia pe care o folosim în studiul nostru este una calitativă. Aceasta se bazează pe 66 de proiecte din baza de date CORDIS a Comisiei Europene ce fac referire la guvernanța securității cibernetice în orașele inteligente. Proiectele pe care le-am selectat se concentrează pe cercetare și

inovare și se desfășoară între 2022 și 2027. Framework-ul nostru aduce o contribuție semnificativă comunității științifice, deoarece identifică riscurile de securitate în infrastructurile la scară largă și propune tehnici de reducere a atacurilor împreună cu potențiale contramăsuri.

8.2 Metodologia de Guvernare a Securității Cibernetice

În această secțiune, prezentăm metodologia pe care am propus-o pentru evaluarea riscului de securitate cibernetică în cadrul infrastructurilor la scară largă, împreună cu tehnici de reducere a atacurilor și potențiale contramăsuri.

Pentru a identifica care sunt tendințele în curs de dezvoltare ale erei moderne în care trăim, am început studiul nostru prin evaluarea proiectelor de cercetare ale Comisiei Europene. Acestea sunt stocate în baza de date CORDIS și aparțin diferitelor domenii de activitate, precum medicină, transporturi, construcții, educație și tehnologie. Ele sunt propuse și implementate de universități, centre de cercetare și corporații. Prin filtrarea proiectelor după cele mai importante două teme ale lucrării noastre, "securitate cibernetică" și "infrastructură", am obținut 66 de rezultate. Apoi, am preluat și importat fiecare proiect în NVivo 14, un instrument pentru analiza calitativă a datelor. Am atribuit câte o temă fiecăruia dintre proiecte și le-am grupat pe cele de același tip astfel încât să putem identifica care sunt cele mai răspândite sectoare de activitate ale infrastructurilor la scară largă. Pe baza acestora, identificăm unde ar putea apărea riscurile de securitate și care ar putea fi potențialele tehnici și contramăsuri de apărare. Tabelul 8.1 prezintă pe scurt pașii pe care i-am adoptat în metodologia noastră.

Table 8.1. Metodologia propusă.

Etapă	Descriere
1. Obținerea datelor	Am căutat proiecte de cercetare ale Comisiei Europene în baza de date CORDIS.
2. Filtrarea datelor	Am filtrat proiectele după cuvintele cheie "cybersecurity" și "infrastructură".
3. Importul proiectelor	Am luat fiecare proiect obținut și l-am importat în NVivo 14, un instrument pentru analiza calitativă a datelor.
4. Asignarea temelor	Am analizat fiecare proiect și i-am atribuit o temă.
5. Gruparea temelor	Am grupat temele de același tip și am identificat domeniile principale cărora le aparțin.
6. Riscurile de securitate	Am analizat fiecare zonă identificată dintr-o infrastructură la scară largă și i-am atribuit riscuri de securitate.
7. Diminuarea atacurilor	Am propus tehnici de diminuare și contramăsuri pentru riscurile de securitate identificate.

8.3 Principalele Amenințări și Potențiale Contramăsuri

În această secțiune, prezentăm principalele amenințări și potențialele contramăsuri pentru domeniile din cadrul infrastructurilor la scară largă pe care le-am identificat prin framework-ul de evaluare a riscului de securitate cibernetică propus.

Odată ce am obținut toate proiectele din baza de date CORDIS, le-am importat în NVivo 14, un instrument de analiză calitativă a datelor. Acolo, am procesat fiecare proiect și am atribuit o temă în funcție de obiectivele acestuia. Au fost grupate temele care corespund unei anumite zone dintr-o infrastructură la scară largă.

Prezentăm în tabelul 8.2 amenințările de securitate și potențialele contramăsuri pentru fiecare zonă dintr-o infrastructură la scară largă pe care am identificat-o prin intermediul framework-ului de evaluare a riscurilor de securitate cibernetică propus.

Am precizat în tabelul 8.2, numărul de proiecte din baza de date CORDIS pe care le-am selectat pentru fiecare domeniu. Coloana amenințări se referă la principalele pericole din domeniul securității cibernetice care vizează o anumită zonă. Am analizat mai multe lucrări de cercetare din literatură pentru a le defini.

Table 8.2. Principalele amenințări și potențiale contramăsuri.

Domeniu	Amenințări	Contramăsuri	Acoperire
Tehnologie 16 proiecte	- Acces neautorizat - Atacuri de tip Man-in-the-middle - Furtul datelor de& cercetare și dezvoltare - Exploatarea vulnerabilităților	- Evaluarea vulnerabilităților – 10 proiecte - Mecanisme de autentificare și autorizare – 6 proiecte - Sisteme de detecție a intruziunilor – 10 proiecte - Sisteme de prevenție a pierderii datelor – 5 proiecte	- 1 proiect tratează 4 contramăsuri - 2 proiecte tratează 3 contramăsuri - 8 proiecte tratează 2 contramăsuri - 5 proiecte tratează o contramăsură
Securitate 12 proiecte	- Vulnerabilitățile codului - Bază de date cu semnături malware învechită - Terminarea proceselor Anti-Virus (AV)	- Crearea unui cod mai securizat – 7 proiecte - Menținerea unei baze de date cu semnături malware actualizate – 0 proiecte - Împiedicarea unor programe malware de a termina procesul AV – 0 proiecte - Integrarea cu date de Threat Intelligence – 6 proiecte - Implementarea unui plan de recuperare în caz de dezastru și a unuia pentru continuitatea activității – 7 proiecte	- 1 proiect tratează 3 contramăsuri - 6 proiecte tratează 2 contramăsuri - 5 proiecte tratează o contramăsură
Medical 9 proiecte	- Expunerea la date sensibile - Întreruperea serviciilor - Interceptarea informațiilor sensibile - Trimiterea de informații false - Alterarea datelor pacientilor	- Rețele Wi-Fi securizate pentru a garanta gestionarea în siguranță a informațiilor confidențiale și a datelor personale (e.g., soluții AirTight Networks) – 6 proiecte - Evaluarea riscurilor (e.g., soluții de securitate de la Intel healthcare) – 4 proiecte	- 2 proiecte tratează 2 contramăsuri - 7 proiecte tratează o contramăsură
Energetic 8 proiecte	- Accesul neautorizat - Botnets (e.g., Zeus, Conficker) - Atacuri de tip Denial of service (DoS) și distributed denial of service (DDoS)	- Sisteme de detecție și prevenție a intruziunilor (e.g., Snort) – 3 proiecte - Cyber Threat Intelligence – 2 proiecte - Metodologii de evaluare a riscurilor (e.g., MEHARI, EBIOS) – 8 proiecte	- 2 proiecte tratează 3 contramăsuri - 1 proiect tratează 2 contramăsuri - 5 proiecte tratează o contramăsură
Mediu 7 proiecte	- Atacuri împotriva rețelelor și PLC-urilor - Compromiterea sistemului - Vulnerabilități	- Repararea vulnerabilităților – 2 proiecte - Soluții de securitate pentru monitorizare – 6 proiecte - Water Information Sharing și Analysis Center, American Water Works Association – 1 proiect	- 2 proiecte tratează 2 contramăsuri - 5 proiecte tratează o contramăsură
Infrastructură 4 proiecte	- Atacuri asupra lanțului de furnizori - Comunicații nesecurizate - Autentificare slabă	- Inventarierea sistemelor și evaluarea riscurilor – 3 proiecte - Patch-uri și actualizări de securitate – 3 proiecte - Securitatea lanțului de furnizori – 3 proiecte	- 2 proiecte tratează 3 contramăsuri - 1 proiect tratează 2 contramăsuri - 1 proiect tratează o contramăsură
Transporturi 4 proiecte	- Distrugerea sistemului de frânare - Oprirea motorului - Afișarea de mesaje false la nivelul computer-ului de bord - Bruierea semnalelor GPS	- Folosirea criptografiei (certificate digitale, infrastructuri cu chei publice, criptarea datelor) – 3 proiecte - Soluții pentru detecția anomaliilor – 1 proiect	- 1 proiect tratează 2 contramăsuri - 2 proiecte tratează o contramăsură - 1 proiect (i.e., NextETRUCK) nu acoperă niciuna dintre aceste contramăsuri principale
Educație 3 proiecte	- Încălcarea securității datelor - Compromiterea informațiilor personale - Atacuri de tip ransomware	- Soluții anti-malware și anti-virus – 0 proiecte - Folosirea unor parole puternice – 0 proiecte - Instruire de securitate – 3 proiecte	- 3 proiecte tratează o contramăsură
Cetățeni 2 proiecte	- Cybercrime - Furtul identității	- Instruire de conştientizare – 2 proiecte - Folosirea de parole sigure – 0 proiecte	- 2 proiecte tratează o contramăsură
Guvernanță 1 proiect	- Întreruperea activității infrastructurilor critice - Fraudă fiscală - Fișiere alterate	- Soluții de prevenție a pierderii datelor (e.g., Symantec, Fortinet) – 0 proiecte - Metodologii de evaluare a riscurilor (e.g., MEHARI, EBIOS) – 1 proiect - Analiza amenințărilor din interior – 0 proiecte	- 1 proiect tratează o contramăsură

Coloana *Contramăsuri* propune acțiuni care ar putea fi întreprinse pentru apărarea și prevenirea atacurilor cibernetice. De asemenea, aceasta conține pentru fiecare contramăsură, numărul de proiecte pe care le acoperă. Valoarea decimală reprezintă suma tuturor proiectelor care implementează sau ar trebui să utilizeze acea contramăsură specifică. Un proiect poate necesita sau conține mai multe contramăsuri.

Coloana *Acoperire* afișează o mapare între proiecte și numărul de contramăsuri pe care fie le implementează, fie ar trebui să le folosească.

8.4 Concluzii

În această lucrare, propunem un model de evaluare a riscului de securitate cibernetică în cadrul infrastructurilor la scară largă, împreună cu tehnici de reducere a atacurilor și potențiale contramăsuri. Folosim în studiul nostru o metodologie calitativă, identificând 66 de proiecte de cercetare europene din perioada 2022-2027 referitoare la guvernanța securității cibernetice în infrastructurile la scară largă. Importăm proiectele într-un program numit NVivo pentru analiza calitativă a datelor. Acolo le grupăm în funcție de sectorul de activitate, astfel încât să putem identifica zonele cele mai răspândite. Acestea din urmă sunt împărțite în componentele principale pe baza cărora identificăm amenințările de securitate.

Rezultatele muncii noastre oferă contribuții științifice semnificative prin identificarea riscurilor de securitate în cadrul infrastructurilor la scară largă și prin propunerea de tehnici de reducere a atacurilor împreună cu potențiale contramăsuri pentru aceste provocări. Rezultatele obținute sunt limitate deoarece am căutat proiecte de cercetare în baza de date CORDIS doar după cuvintele cheie "cybersecurity" și "infrastructure". Dacă am fi adăugat și alți termeni în căutare, precum "smart city" și "governance", poate am fi obținut mai multe rezultate. Astfel, domeniile de activitate ar fi fost mai extinse, iar gama de amenințări identificate și contramăsuri propuse ar fi fost mai mare. În acest fel, studiul nostru analizează unele zone și ar putea să nu trateze toate domeniile din care sunt compuse infrastructurile la scară largă.

În ceea ce privește direcțiile viitoare de cercetare, intenționăm să îmbogățim modelul actual prin prezentarea unor cazuri practice de utilizare în care putem descrie încălcări reale ale securității datelor. Ar fi util deoarece acolo putem veni cu tactici și contramăsuri concrete și specifice pentru detecția intruziunilor.

9 | Concluzii și Direcții Viitoare de Cercetare

Această teză se concentrează pe identificarea celor mai răspândite tipuri de atacuri cibernetice din epoca modernă și pe dezvoltarea de strategii de detecție, contramăsuri și soluții de securitate robuste care oferă eficiență, rezultate precise și timp de răspuns rapid.

9.1 Contribuții Originale

1. Am identificat atacurile moderne de securitate cibernetică care descriu plaja de amenințări pentru perioada 2021-2022.
2. Am propus câteva contramăsuri și strategii de detecție contemporane care ajută la apărarea împotriva celor mai răspândite amenințări de securitate.
3. Am propus o tehnică de detecție a adreselor URL scurte care sunt malițioase.
4. Am propus o tehnică scalabilă de detecție a adreselor URL malițioase din cadrul atacurilor de tip smishing.
5. Am propus o soluție de detecție a bullying-ului pentru GIF-uri folosind modele de deep learning.
6. Am propus o soluție pentru detecția cyberbullying-ului în videoclipurile de pe TikTok, folosind o abordare ce se bazează pe un model de deep learning.
7. Am propus *NextEDR* - sisteme de tip EDR de generația următoare bazate pe agenți pentru amenințările de securitate cibernetică.
8. Am propus un framework pentru guvernarea securității cibernetice în cadrul infrastructurilor la scară largă.
9. Am elaborat metodologii de prevenție pentru amenințările ce au fost abordate de soluțiile de detecție pe care le-am propus.

9.2 Lista Publicațiilor și a Proiectelor

Principalele rezultate ale acestei teze au fost prezentate la diferite conferințe și jurnale. Am șapte publicații din care șase ca prim autor și una ca și coautor. Lista publicațiilor mele constă în patru articole în jurnale internaționale (International Journal of Computational Science and Engineering, Information, UPB Sci. Bull., Series C, și Romanian Journal of Information Technology and Automatic Control) și trei lucrări în cadrul unor conferințe internaționale consacrate (2023

24th International Conference on Control Systems and Computer Science (CSCS), 2022 21st RoEduNet Conference: Networking in Education and Research (RoEduNet) și 2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)).

Aș dori să mulțumesc recenzorilor pentru timpul și expertiza lor, comentariile constructive și informațiile valoroase.

Articole științifice publicate:

1. **Stoleriu, Razvan**; Negru, Catalin; Radulescu, Dragos; "Modern Cyber Security Attacks, Detection Strategies, and Countermeasures Procedures", "2023 24th International Conference on Control Systems and Computer Science (CSCS)", pp. 198-205, 2023, IEEE, doi: 10.1109/CSCS59211.2023.00039.

2. **Stoleriu, Razvan**; Negru, Catalin; Mocanu, Bogdan-Costel; Pop, Florin; "Malicious Short URLs Detection Technique", "2023 22nd RoEduNet Conference: Networking in Education and Research (RoEduNet)", pp. 1-6, 2023, IEEE, doi: 10.1109/RoEduNet60162.2023.10274913.

3. **Stoleriu, Razvan**; Negru, Catalin; Mocanu, Bogdan-Costel; Constantinescu, Emil-Andrei; Mocanu, Alexandra-Elena; Pop, Florin; "Scalable Malicious URL Detection Technique for Smishing Attacks", "International Journal of Computational Science and Engineering", Inderscience Publishers (IEL).

4. Mocanu, Bogdan-Costel; **Stoleriu, Razvan**; Mocanu, Alexandra-Elena; Negru, Catalin; Dragotoiu, Elena-Gabriela; Moisescu, Mihnea-Alexandru; Pop, Florin; "NextEDR-Next generation agent-based EDR systems for cybersecurity threats", "2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)", pp. 183-190, 2024, IEEE, doi: 10.1109/PDP62718.2024.00033.

5. **Stoleriu, Razvan**; Nascu, Andrei; Anghel, Ana Magdalena; Pop, Florin; "Bullying Detection Solution for GIFs Using a Deep Learning Approach", "Information", 15, 2024, MDPI, doi: https://www.mdpi.com/2078-2489/15/8/446.

6. **Stoleriu, Razvan**; Nascu, Andrei; Pop, Florin; "Cyberbullying Detection on TikTok Using a Deep Learning Approach", "UPB Sci. Bull., Series C", 2024. - **Lucrare acceptată**

7. **Stoleriu, Razvan**; Petre, Ionut; Pop, Florin; "Cybersecurity Governance in Large-scale Infrastructures", "Romanian Journal of Information Technology and Automatic Control", 2025. - **Lucrare acceptată (ISI)**

În perioada de doctorat, am fost membru al unui proiect pentru care sunt recunoscător, oferindu-mi perspectiva de a construi cazuri de utilizare din lumea reală pentru teza mea și, de asemenea, oportunitatea de a interacționa cu diferiți cercetători. Proiectul din care am făcut parte a câștigat premiul I la concursul de proiecte *Invest UNSTPB Proof of Concept*. Acest proiect este:

- INVEST UPB Proof of Concept – Faza 2, pentru proiectul NextEDR - Sisteme EDR de generația următoare bazate pe agenți pentru amenințările de securitate cibernetică, finanțat prin proiectul DECIP: Dezvoltarea Capacității Instituționale a Universității POLITEHNICA din București, Contract PFE 22, Perioada: 24 Noiembrie – 10 Iunie 2024 (7 luni), Director: Asist. Bogdan-Costel Mocanu.

Bibliography

[1] Asmaa Shaker Ashoor and Sharad Gore. Difference between intrusion detection system (ids) and intrusion prevention system (ips). In *Advances in Network Security and Applications: 4th International Conference, CNSA 2011, Chennai, India, July 15-17, 2011 4*, pages 497–501. Springer, 2011.

[2] Andreea Bendovschi. Cyber-attacks–trends, patterns and security countermeasures. *Procedia Economics and Finance*, 28:24–31, 2015.

[3] Sarika Choudhary, Ritika Saroha, and Mrs Sonal Beniwal. How anti-virus software works?? *International Journal*, 3(4):483–484, 2013.

[4] Prithviraj Dasgupta, Joseph B Collins, and Ranjeev Mittu. *Adversary-Aware Learning Techniques and Trends in Cybersecurity*. Springer, 2021.

[5] Rida Khatoun and Sherali Zeadally. Cybersecurity and privacy solutions in smart cities. *IEEE Communications Magazine*, 55(3):51–59, 2017.

[6] Sean Lawson. Beyond cyber-doom: Cyberattack scenarios and the evidence of history. *Mercatus Center at George Mason University*, 2011.

[7] Richard Lippmann, Seth Webster, and Douglas Stetson. The effect of identifying vulnerabilities and patching software on the utility of network intrusion detection. In *Recent Advances in Intrusion Detection: 5th International Symposium, RAID 2002 Zurich, Switzerland, October 16–18, 2002 Proceedings 5*, pages 307–326. Springer, 2002.

[8] Michael R Lyu and Lorrien KY Lau. Firewall security: Policies, testing and performance evaluation. In *Proceedings 24th Annual International Computer Software and Applications Conference. COMPSAC2000*, pages 116–121. IEEE, 2000.

[9] Adam McNeil and W Stuart Jones. Mobile malware is surging in europe: A look at the biggest threats, 2022.

[10] Bogdan-Costel Mocanu, Razvan Stoleriu, Alexandra-Elena Mocanu, Cătălin Negru, Elena-Gabriela Drăgotoiu, Mihnea-Alexandru Moisescu, and Florin Pop. Nextedr-next generation agent-based edr systems for cybersecurity threats. In *2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pages 183–190. IEEE, 2024.

[11] United Nations. World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100 | united nations. https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100, 2017. [Online; accessed 18-January-2024].

[12] CHECK POINT. 2022 cyber attacks trends: Mid-year report. https://resources.checkpoint.com, 2022. [Online; accessed 06-October-2023].

[13] Jamal Raiyn et al. A survey of cyber attack detection strategies. *International Journal of Security and Its Applications*, 8(1):247–256, 2014.

[14] Jungwoo Ryoo, Syed Rizvi, William Aiken, and John Kissell. Cloud security auditing: challenges and emerging approaches. *IEEE Security & Privacy*, 12(6):68–74, 2013.

[15] IBM Security. X-force threat intelligence index 2022. https://www.ibm.com/downloads/cas/ADLMYLAZ, 2022. [Online; accessed 10-April-2024].

[16] A Seetharaman, Nitin Patwa, Veena Jadhav, AS Saravanan, and Dhivya Sangeeth. Impact of factors influencing cyber threats on autonomous vehicles. *Applied Artificial Intelligence*, 35(2):105–132, 2021.

[17] Ioannis Stellios, Kostas Mokos, and Panayiotis Kotzanikolaou. Assessing smart light enabled cyber-physical attack paths on urban infrastructures and services. *Connection Science*, 34(1):1401–1429, 2022.

[18] Razvan Stoleriu, Andrei Nascu, Ana Magdalena Anghel, and Florin Pop. Bullying detection solution for gifs using a deep learning approach. *Information*, 15(8):446, 2024.

[19] Razvan Stoleriu, Catalin Negru, Bogdan-Costel Mocanu, and Florin Pop. Malicious short urls detection technique. In *2023 22nd RoEduNet Conference: Networking in Education and Research (RoEduNet)*, pages 1–6. IEEE, 2023.

[20] Răzvan Stoleriu, Cătălin Negru, and Dragos Rădulescu. Modern cyber security attacks, detection strategies, and countermeasures procedures. In *2023 24th International Conference on Control Systems and Computer Science (CSCS)*, pages 198–205. IEEE, 2023.

[21] Kami Vaniea and Yasmeen Rashidi. Tales of software updates: The process of updating software. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 3215–3226, 2016.

[22] Zhensheng Zhang, Ya-Qin Zhang, Xiaowen Chu, and Bo Li. An overview of virtual private network (vpn): Ip vpn and optical vpn. *Photonic network communications*, 7:213–225, 2004.

[23] Aaron Zimba. Malware-free intrusion: a novel approach to ransomware infection vectors. *International Journal of Computer Science and Information Security*, 15(2):317, 2017.

[24] AlMaha Abu Zuraiq and Mouhammd Alkasassbeh. Phishing detection approaches. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, pages 1–6. IEEE, 2019.