UNIVERSITETI I PRISHTINËS
"HASAN PRISHTINA"
UNIVERSITY OF PRISHTINA

Rr.Xhorxh Bush, Ndërtesa e Rektoratit, 10 000 Prishtinë,Republika e Kosovës
Tel: +383 38 244 183 • E-mail: rektorati@uni-pr.edu • www.uni-pr.edu

Fakulteti i Inxhinierisë Elektrike dhe Kompjuterike
Faculty of Electrical and Computer Engineering

Abstrakt i zgjeruar i punimit të diplomës
Niveli Master

Kandidatja	Titulli i temës
Rinesa Bislimi	Zhvillimi i një Sistemi për Monitorim dhe Analizë Cross-Layer të Log-ave me Machine Learning

1. Hyrje dhe kontekst

Zhvillimi i shpejtë i teknologjisë së informacionit dhe rritja e vazhdueshme e kompleksitetit të infrastrukturave moderne kanë sjellë një volum gjithnjë e më të madh të dhënash log, të gjeneruara nga pajisje, sisteme dhe aplikacione të ndryshme. Këto të dhëna përbëjnë një burim thelbësor për monitorimin e gjendjes së sistemeve dhe për identifikimin e kërcënimeve të sigurisë kibernetike, por volumi dhe heterogjeniteti i tyre e bëjnë analizën manuale praktikisht të pamundur.

Një nga qasjet më të rëndësishme që ka dalë në pah në këtë kontekst është **analiza cross-layer e log-eve** (cross-layer analysis). Kjo qasje nënkupton kombinimin dhe korrelacionin e të dhënave log nga disa shtresa të infrastrukturës - rrjeti (network), sistemi (system) dhe aplikacioni (application) - për të zbuluar kërcënime dhe anomali që do të mbeteshin të padukshme nëse secila shtresë analizohet e izoluar.

Ky punim diplome fokusohet në zhvillimin dhe implementimin e një sistemi për monitorim dhe analizë cross-layer të log-eve, i cili integron teknikat e Machine Learning për detektimin e anomalive me një analizë korrelacioni (correlation engine) për lidhjen e ngjarjeve ndërmjet shtresave network, system dhe application.

2. Problematika dhe motivimi

Në mjedise reale të teknologjisë së informacionit, çdo shtresë e infrastrukturës gjeneron sasi të mëdha të dhënash log në formate dhe me semantikë të ndryshme. Monitorimi i sigurisë mbështetet tradicionalisht në analizën e izoluar të secilës shtresë, ku ngjarjet e rrjetit, të sistemit dhe të aplikacionit shqyrtohen veçmas nga njëra-tjetra.

Megjithatë, kjo qasje single-layer rezulton e pamjaftueshme për identifikimin e sulmeve të ndërlikuara dhe shumë-fazore (multi-stage). Një sulmues që vepron nëpër disa shtresa lë gjurmë të pjesshme në secilën prej tyre, të cilat, kur shqyrtohen veçmas, duken si ngjarje rutinë dhe nuk shkaktojnë alarm, ndërsa kuptimi i tyre i plotë shfaqet vetëm kur ato vendosen në një kontekst të përbashkët. Po ashtu, sistemet tradicionale të bazuara në rregulla statike gjenerojnë një numër

të lartë alarmesh false dhe nuk janë në gjendje të përshtaten ndaj kërcënimeve të reja, ndërsa heterogjeniteti i burimeve të log-eve e vështirëson normalizimin dhe korrelacionin e tyre.

Motivimi kryesor i këtij punimi qëndron në adresimin e këtyre sfidave përmes një qasjeje të integruar cross-layer, e cila kombinon teknikat e Machine Learning për detektimin e anomalive me një correlation engine të aftë t'i lidhë ngjarjet ndërmjet shtresave bazuar në dimensione kohore, identitete (si IP adresat) dhe karakteristika semantike, duke i agreguar ato në incidente të vetme.

3. Qëllimi dhe objekti i punimit

Qëllimi i këtij punimi është të dizajnojë, zhvillojë dhe vlerësojë një sistem për monitorim dhe analizë cross-layer të log-eve, i cili kombinon Machine Learning me korrelacionin ndër-shtresor për detektimin e anomalive dhe sulmeve shumëfazore në mjedise reale me volum të lartë të të dhënave.

Objekti i punimit përfshin:

- analizën teorike të Machine Learning në sigurinë kibernetike, sistemeve të detektimit të ndërhyrjeve (IDS) dhe dallimit ndërmjet qasjeve cross-system dhe cross-layer;
- studimin dhe krahasimin e algoritmeve të Machine Learning (Random Forest, XGBoost, LightGBM, SVM, rrjeta neurale dhe Isolation Forest) për zbulimin e anomalive;
- dizajnimin dhe implementimin e një sistemi funksional full-stack që realizon ingestimin, analizën me Machine Learning, korrelacionin cross-layer dhe vizualizimin e incidenteve në kohë reale.

Punimi synon të tregojë se si kombinimi i analizës cross-layer me Machine Learning mund të zbulojë kërcënime që analiza një-shtresore do t'i anashkalonte, duke ofruar një zgjidhje praktike dhe të aplikueshme në mjedise reale.

4. Përmbajtja teorike

Në pjesën teorike, punimi trajton rolin e Machine Learning në sigurinë kibernetike dhe sistemet e detektimit të ndërhyrjeve (IDS), si dhe analizën e log-eve të sistemeve.

Më tej, analizohen dhe krahasohen gjashtë algoritme të Machine Learning - Random Forest, XGBoost, LightGBM, Support Vector Machine, rrjetat neurale (MLP) dhe Isolation Forest - duke u ndalur në bazat e tyre matematikore, përparësitë dhe përshtatshmërinë për secilën shtresë të të dhënave. Trajtohen gjithashtu datasetet benchmark (BGL, CICIDS-2017 dhe TDA) dhe roli i tyre në vlerësimin e modeleve.

Po ashtu, trajtohen procesi i feature engineering për secilën shtresë dhe metrikat e vlerësimit të modeleve (saktësia, preciziteti, recall, F1-score dhe ROC AUC), të cilat interpretohen në kontekstin specifik të sigurisë kibernetike, ku numri i alarmeve false dhe zbulimi i anomalive të rralla kanë rëndësi kritike.

5. Metodologjia dhe implementimi praktik

Në aspektin praktik, punimi paraqet zhvillimin e një platforme funksionale full-stack për monitorim dhe analizë cross-layer. Backend-i është implementuar në Python me framework-un

FastAPI dhe SQLAlchemy, ndërsa frontend-i në Angular 17, me komunikim në kohë reale përmes WebSocket. Sistemi është ndërtuar mbi një arkitekturë të ndarë në shtresa funksionale (ingestim, feature engineering, trajnim, korrelacion dhe vizualizim).

Pipeline-i i analizës cross-layer përfshin:

- ingestimin dhe normalizimin e të dhënave nga tri datasete heterogjene (BGL, CICIDS-2017 dhe TDA) në një skemë të përbashkët;
- nxjerrjen e tipareve (feature engineering) specifike për shtresën e sistemit, të rrjetit dhe të aplikacionit;
- trajnimin dhe vlerësimin e modeleve të Machine Learning për secilën shtresë;
- gjenerimin e anomalive dhe parashikimeve nga modelet e trajnuara;
- korrelacionin cross-layer të ngjarjeve bazuar në kohë, identitete (IP) dhe karakteristika semantike, duke i agreguar në incidente të vetme;
- vlerësimin e rregullave të alarmeve dhe dërgimin e njoftimeve (p.sh. përmes email-it/SMTP);
- vizualizimin e incidenteve, statistikave dhe gjendjes së shtresave në një panel monitorimi në kohë reale.

Platforma përfshin gjithashtu funksionalitete si simulimi i të dhënave në kohë reale, heatmap-i i anomalive, pamja sipas entiteteve, zinxhirët e sulmeve dhe gjenerimi i raporteve etj.

6. Rezultatet dhe diskutimi

Rezultatet e arritura tregojnë se qasja cross-layer e kombinuar me Machine Learning mundëson zbulimin e sulmeve multi-stage dhe anomalive që analiza një-shtresore do t'i anashkalonte. Modelet e trajnuara arrijnë performancë të lartë në detektimin e anomalive për secilën shtresë, ndërsa correlation engine lidh ngjarjet e shpërndara në incidente me kontekst të pasur semantik dhe temporal.

Platforma e zhvilluar ofron dëshmi konkrete mbi vlerën e integrimit të të dhënave shumë-shtresore për monitorimin inteligjent dhe sigurinë kibernetike. Në të njëjtën kohë, punimi diskuton edhe kufizimet e qasjes, si varësia nga cilësia dhe etiketimi i të dhënave, kostoja llogaritëse për datasete me volum të madh, si dhe sfidat e përshtatjes ndaj kërcënimeve që evoluojnë.

7. Përfundime

Si përfundim, ky punim tregon se kombinimi i analizës cross-layer me Machine Learning përbën një qasje bashkëkohore dhe premtuese për monitorimin inteligjent dhe detektimin e anomalive multi-stage. Përmes analizës teorike dhe implementimit praktik, është demonstruar se integrimi i të dhënave nga shtresat network, system dhe application nën një pipeline të vetëm, i shoqëruar me një motor korrelacioni, çon në sisteme më të sakta, më efikase dhe më të vetëdijshme ndaj kontekstit.

Punimi ofron një bazë të qëndrueshme për zhvillime të mëtejshme dhe hap mundësi për hulumtime të ardhshme në drejtim të zgjerimit të shtresave të analizuara, integrimit të teknikave deep learning, korrelacionit në kohë reale në shkallë të gjerë dhe vendosjes së sistemit në mjedise të punës.