OQIM-DARAJALI HUJUMLARNI ANIQLASH DETEKTORINING HUJUM SINFLARI KESIMIDAGI QAMROVINI UCH BENCHMARK TO‘PLAMDAGI TAHLILI

Maxkamova Dildora Begaliyevna
Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti
d.maxkamova@tuit.uz

Annotatsiya. *Tezisda faqat normal oqimlarda o‘qitilgan self-supervised graf detektori bo‘lgan FlowGuardning aniqlash darajasi uchta NetFlow benchmark to‘plamda hujum sinflari kesimida o‘lchanadi. Benchmark to‘plamlar sifatida NF-UNSW-NB15-v2, NF-BoT-IoT-v2, NF-CSE-CIC-IDS2018-v2 lar olingan. Kalibrlangan α = 0,01 ish nuqtasida 27 sinf bo‘yicha olingan TPR qiymatlari “oqim-darajali iz” mezonini shakllantiradi. Strukturaviy va hajmiy iz qoldiruvchi sinflar 0,90 dan yuqori, mazmun-darajali va faqat oqimlar agregatida namoyon bo‘ladigan taqsimlangan sinflar 0,27 dan past aniqlanadi. Natija oqim-darajali detektorni mazmun-tahlil va oyna-darajali vositalarning o‘rnini bosuvchi emas, to‘ldiruvchisi sifatida joylashtirishni asoslaydi.*

Kalit so‘zlar: *tarmoq hujumlarini aniqlash, NetFlow, hujum sinflari, qamrov tahlili, oqim-darajali iz, self-supervised o‘qitish, kalibrlangan chegara.*

Abstract. *The paper measures the per-class detection rate of FlowGuard, a self-supervised graph-based detector trained on benign flows only, on three NetFlow benchmark datasets: NF-UNSW-NB15-v2, NF-BoT-IoT-v2 and NF-CSE-CIC-IDS2018-v2. True positive rates (TPR) obtained for 27 attack classes at the calibrated operating point α = 0.01 form a “flow-level footprint” criterion. Classes that leave a structural or volumetric footprint are detected above 0.89, content-level classes below 0.50, and distributed classes that manifest only in the aggregate of flows below 0.27. The result justifies deploying a flow-level detector as a complement to, rather than a replacement for, content inspection and window-level analysis tools.*

Keywords: *intrusion detection, NetFlow, attack classes, flow-level footprint, self-supervised learning, calibrated threshold.*

Аннотация. *В тезисах измеряется уровень обнаружения детектора FlowGuard — графового детектора с самоконтролируемым обучением (self-*

supervised), обученного только на нормальных потоках, — в разрезе классов атак на трёх эталонных наборах NetFlow: NF-UNSW-NB15-v2, NF-BoT-IoT-v2 и NF-CSE-CIC-IDS2018-v2. Значения TPR, полученные для 27 классов атак в калиброванной рабочей точке α = 0,01, формируют критерий «следа на уровне потока». Классы, оставляющие структурный и объёмный след, обнаруживаются с долей выше 0,89, классы уровня содержимого — ниже 0,50, а распределённые классы, проявляющиеся только в совокупности потоков, — ниже 0,27. Результат обосновывает размещение потокового детектора как дополнения к средствам анализа содержимого и анализа во временном окне, а не как их замены.

Ключевые слова: *обнаружение вторжений, NetFlow, классы атак, след на уровне потока, самоконтролируемое обучение, калиброванный порог.*

Kirish

Oqim asosidagi hujumlarni aniqlash tizimlari oqim normal yoki anomal bo‘lgan ikkilik qo‘yilishda baholanadi. Bunday baholashning yakuniy ko‘rsatkichi — ROC-AUC, PR-AUC yoki umumiy TPR — detektorning qaysi hujumlarni ko‘rayotganini ochmaydi. Holbuki joriy etish uchun aynan shu savol hal qiluvchi bo‘lib, o‘tkazib yuborilgan 30 % hujum tasodifiy taqsimlangan bo‘lsa bu bir masala; agar u butunlay muayyan sinflarga to‘g‘ri kelsa — arxitekturaviy xulosa talab qiladi. Adabiyotda self-supervised graf modellari [1], [2] asosan agregat metrikalar bilan hisobot beradi; sinflar kesimidagi tahlil, mavjud bo‘lganda ham, bitta to‘plam bilan chegaralanadi. NetFlow yozuvining tabiati — unda paket mazmuni (payload) yo‘qligi [3], [4] — esa aprior gipotezani beradi: mazmun bilan ta’riflanadigan hujumlar oqim darajasida ko‘rinmasligi kerak.

Mazkur tezisning maqsadi — ushbu gipotezani uchta tarkibiy jihatdan farqli NetFlow-benchmark to‘plamda [5] hujum sinflari kesimida tekshirish va natijadan detektorning qo‘llanish chegarasini belgilovchi mezonni chiqarish. Tahlil muallifning oqim-asosidagi hujumlarni aniqlash bo‘yicha dissertatsion tadqiqoti doirasida bajarildi va uning oldingi ishlarida [8] boshlangan yo‘nalishni davom ettiradi.

Detektor va o‘lchov sharoiti

Baholanayotgan detektor — FlowGuard — IP-manzillarsiz surrogat servis-graf ustida ishlaydigan self-supervised enkoder-dekoder modelidir. Graf tugunlari (transport protokoli, ilova-qatlam protokoli, port toifasi) uchligi bilan aniqlanadi, oqimlar qirra sifatida saqlanadi; model faqat normal oqimlarda rekonstruksiya xatosini minimallashtirishga o‘qitiladi va 32 617 parametrdan iborat. Anomaliya

balli — rekonstruksiya xatosi; qaror chegarasi τ trening normallari ballarining (1 − α)-kvantili sifatida belgilanadi va keyin hech qachon test ma’lumotiga moslashtirilmaydi. Shu bois sinflar kesimidagi barcha qiymatlar **ko‘rilmagan** test oqimlarida, oldindan qayd etilgan ish nuqtasida olingan — ular ex post tanlangan optimal chegaraning emas, joriy etishda haqiqatan kuzatiladigan rejimning aksidir.

Uch to‘plam hujum ulushi bo‘yicha qutbli va oraliq nuqtalarni ifodalaydi: NF-UNSW-NB15-v2 [6] — korporativ trafik, hujum ulushi ≈ 3,8 %, to‘qqiz sinf; NF-BoT-IoT-v2 [7] — IoT-botnet, ≈ 99,6 %, to‘rt sinf; NF-CSE-CIC-IDS2018-v2 — yirik korporativ to‘plam, ≈ 11,8 %, o‘n to‘rt sinf. Har birida trening 100 000 normal oqimdan, test tabiiy sinf taqsimotidagi 200 000 oqimdan iborat; natijalar besh mustaqil urug‘ bo‘yicha o‘rtachalangan. To‘plamlar asl ko‘p sinfli annotatsiyani saqlagani bois ikkilik detektorning har bir sinfni qay darajada qamrashini to‘liq test bo‘limida o‘lchash mumkin. Natijalar 1-jadvalda keltirilgan; har bir to‘plam uchun normal sinf qatori kuzatilgan FPR ni ko‘rsatadi.

1-jadval

Hujum sinflari kesimidagi aniqlash darajasi α = 0,01 ish nuqtasida (besh urug‘ bo‘yicha o‘rtacha)

To‘plam	Hujum sinfi	Oqimlar soni	TPR (α = 0,01)	Iz turi
UNSW-NB15	Fuzzers	2 134	0,9967	strukturaviy
UNSW-NB15	Reconnaissance	1 118	0,9955	strukturaviy
UNSW-NB15	Shellcode	129	1,0000	strukturaviy
UNSW-NB15	Worms	20	1,0000	strukturaviy
UNSW-NB15	Backdoor	84	0,9643	strukturaviy
UNSW-NB15	Generic	593	0,8111	aralash
UNSW-NB15	DoS	397	0,6045	aralash
UNSW-NB15	Analysis	77	0,4935	mazmun
UNSW-NB15	Exploits	3 006	0,3333	mazmun
UNSW-NB15	*Normal (FPR)*	192 442	0,0099	—
BoT-IoT	DDoS	93 990	0,0004	agregat
BoT-IoT	DoS	89 772	0,0243	agregat
BoT-IoT	Reconnaissance	15 370	0,0012	agregat
BoT-IoT	Theft	17	0,0000	mazmun
BoT-IoT	*Normal (FPR)*	851	0,0118	—
IDS2018	FTP-BruteForce	301	1,0000	strukturaviy

To‘plam	Hujum sinfi	Oqimlar soni	TPR (α = 0,01)	Iz turi
IDS2018	DDOS attack-LOIC-UDP	25	1,0000	hajmiy
IDS2018	DoS attacks-SlowHTTPTest	161	0,9429	strukturaviy
IDS2018	DoS attacks-Hulk	5 046	0,9284	hajmiy
IDS2018	DoS attacks-GoldenEye	310	0,8961	hajmiy
IDS2018	Bot	335	0,2657	agregat
IDS2018	Brute Force -Web	18	0,2222	mazmun
IDS2018	Brute Force -XSS	13	0,2000	mazmun
IDS2018	SSH-Bruteforce	1 123	0,1478	agregat
IDS2018	DoS attacks-Slowloris	100	0,1400	agregat
IDS2018	DDOS attack-HOIC	12 423	0,0999	agregat
IDS2018	DDoS attacks-LOIC-HTTP	2 494	0,0963	agregat
IDS2018	Infilteration	1 326	0,0106	mazmun
IDS2018	SQL Injection	4	0,0000	mazmun
IDS2018	*Normal (FPR)*	176 321	0,0105	—

Natijalar tahlili: “oqim-darajali iz” mezoni

NF-UNSW-NB15-v2 da aniqlash profili payload-gipotezasi bilan to‘liq mos. Izi strukturaviy bo‘lgan sinflar deyarli to‘liq qamraladi: Fuzzers (0,997) va Reconnaissance (0,996) qisqa, takrorlanuvchi va ko‘pincha javobsiz oqimlardan iborat; Shellcode va Worms, soni oz bo‘lsa-da, to‘liq aniqlanadi; Backdoor 0,964. Ajratuvchi belgisi oqim statistikasida emas, paket mazmunida yotgan sinflar esa eng past qamrovga ega: eng yirik sinf Exploits (3 006 oqim) 0,333, Analysis 0,494. Sinflar bo‘yicha tortilgan o‘rtacha 0,692 — umumiy TPR bilan aynan mos, bu hisob-kitobning ichki izchilligini ko‘rsatadi. Bu model nuqsoni emas, kirish tasvirining chegarasi: oqim yozuvida payload yo‘q, shu bois mazmun bilan ta’riflanadigan sinf detektorga o‘qiydigan iz qoldirmaydi.

NF-CSE-CIC-IDS2018-v2 ning o‘n to‘rt sinfi gipotezani tasdiqladi va aniqlashtirdi. Mazmun-darajali hujumlar bu yerda ham eng past ushlanadi (SQL Injection 0,000; Brute Force-XSS 0,200; Brute Force-Web 0,222; Infilteration 0,011), oqim darajasida yaqqol iz qoldiruvchi DoS vositalari esa yuqori (Hulk 0,928; SlowHTTPTest 0,943; GoldenEye 0,896; LOIC-UDP 1,000). Biroq yangi muhim qirra ochildi: **alohida oqimi** odatiy so‘rovdan deyarli farq qilmaydigan taqsimlangan vositalar — HOIC (0,100), LOIC-HTTP (0,096), SSH-Bruteforce

(0,148), Bot (0,266) — ham past ushlanadi, garchi ular mazmun-darajali bo‘lmasa ham. Demak, ajratuvchi chiziq “payload bormi” savoli bo‘ylab emas, aniqroq mezon — **“hujum alohida oqim statistikasida iz qoldiradimi”** — bo‘ylab o‘tadi. Anomallik faqat oqimlar agregatida (tezlik, takroriylik, manbalar soni) namoyon bo‘lsa, oqim-darajali detektor uni tabiatan ko‘rmaydi. Aynan shu sabab IDS2018 da $\alpha = 0{,}01$ nuqtasidagi tortilgan umumiy TPR $\approx 0{,}305$ — ROC-AUC = 0,915 bo‘lgani holda — chunki test hajmining katta qismini agregat-darajali HOIC va LOIC-HTTP egallaydi. Reyting sifati (AUC) bilan ish-nuqtasi qamrovi (TPR) orasidagi bu tafovut sinflar kesimisiz ko‘rinmas edi.

NF-BoT-IoT-v2 uchinchi, muhit-darajali kuzatuv beradi. Kalibrlash saqlangan (FPR = 0,0118 $\approx \alpha$), ammo barcha hujum sinflari bo‘yicha TPR minimal (0,0004–0,024). Bu yerda sabab iz turida emas: testning 99,6 % i hujum bo‘lgan muhitda trening normallari kvantilidan olingan ish nuqtasi tabiatan foydali ishlamaydi, reyting sifatining o‘zi (ROC-AUC $\approx$ 0,80) bunga kifoya emas. Amaliy qoida: sinf tarkibi teskarilangan muhitlarda chegara maqsad muhitning o‘z normal namunasida qayta kalibrlanishi yoki usulning qo‘llanilishi qayta ko‘rib chiqilishi lozim.

Shu tariqa uch to‘plamdagi 27 sinf to‘rt guruhga tushadi (1-jadvalning oxirgi ustuni): strukturaviy va hajmiy iz — TPR 0,90 dan yuqori; aralash iz (Generic, DoS) — 0,60–0,81; mazmun-darajali iz — 0,49 dan past; agregat-darajali iz — 0,27 dan past. Ikki ogohlantirish raqamlar bilan birga yuriladi: detektor ikkilik bo‘lib, trening davomida birorta hujum yorlig‘ini ko‘rmagan, shu bois modelda birorta sinfga yo‘naltirilgan hech narsa yo‘q — kuzatilgan profil sof kirish tasvirining xossasidir; eng kichik sinflar (Worms 20, SQL Injection 4, XSS 13 oqim) statistik jihatdan kam vaznga ega va ularning qiymatlari sifat ko‘rsatkichi sifatida o‘qiladi.

Arxitekturaviy xulosa

Natija ikki amaliy qarorni asoslaydi. Birinchisi — joylashtirish: oqim-darajali detektor kontent-tekshiruv vositalarining o‘rnini bosmaydi, ularni to‘ldiradi — strukturaviy va hajmiy xulqni o‘zi qamrab, payload bilan ta’riflanadigan sinflarni ularga qoldiradi. Ikkinchisi — rivojlanish yo‘nalishi: agregat-darajali sinflar (taqsimlangan DDoS, sekin brute-force, botnet C2) ni qamrab olish uchun oqim-darajali tasvir NetFlow v3 avlodining oyna-darajali temporal xususiyatlari bilan to‘ldirilishi lozim; bu kengaytma xususiyat siyosatiga daxl qilmaydi va shu protokol doirasida sinaladi. Aniqlangan mezon detektorning kamchiligi emas, uning mas’uliyat doirasidir: u nimani ko‘radi-yu, nimani tabiatan ko‘rmasligi endi 27 sinf bo‘yicha miqdoriy belgilangan.

Xulosa

Uch to‘plam va 27 hujum sinfi bo‘yicha o‘tkazilgan qamrov tahlili uchta natija berdi. Birinchidan, payload-gipotezasi ikki korporativ to‘plamda tasdiqlandi: mazmun-darajali sinflar 0,49 dan past, strukturaviy va hajmiy sinflar 0,90 dan yuqori aniqlanadi. Ikkinchidan, gipoteza aniqlashtirildi: ajratuvchi mezon — hujumning alohida oqim statistikasida iz qoldirishi; faqat agregatda ko‘rinadigan taqsimlangan hujumlar ham, mazmun-darajali sinflar kabi, oqim-darajali detektor uchun yopiq (0,27 dan past). Uchinchidan, sinf tarkibi teskarilangan muhitda muammo iz turida emas, ish nuqtasida — bu kalibrlash strategiyasining qo‘llanish chegarasini belgilaydi. Mezon joriy etishda ikki tomonlama ishlatiladi: detektor qaysi tahdid sinflarini mustaqil yopishini va qaysilari uchun kontent-tekshiruv yoki oyna-darajali tahlil majburiy hamroh ekanini oldindan aytib beradi.

Foydalanilgan adabiyotlar

1. Caville E., Lo W.W., Layeghy S., Portmann M. Anomal-E: A self-supervised network intrusion detection system based on graph neural networks // Knowledge-Based Systems. — 2022. — Vol. 258. — Art. 110030.

2. Guerra L., Chapuis T., Duc G., Mozharovskyi P., Nguyen V.-T. Self-supervised learning of graph representations for network intrusion detection // Advances in Neural Information Processing Systems (NeurIPS). — 2025.

3. Claise B., Trammell B., Aitken P. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information. RFC 7011. — IETF, 2013.

4. Velan P., Čermák M., Čeleda P., Drašar M. A survey of methods for encrypted traffic classification and analysis // International Journal of Network Management. — 2015. — Vol. 25, No. 5. — P. 355–374.

5. Sarhan M., Layeghy S., Portmann M. Towards a standard feature set for network intrusion detection system datasets // Mobile Networks and Applications. — 2022. — Vol. 27. — P. 357–370.

6. Moustafa N., Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems // Military Communications and Information Systems Conference (MilCIS). — Canberra, 2015. — P. 1–6.

7. Koroniotis N., Moustafa N., Sitnikova E., Turnbull B. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset // Future Generation Computer Systems. — 2019. — Vol. 100. — P. 779–796.

8. Allanov O.M., Makhkamova D.B., Ibragimov S.O., Quyliyev B.Sh. Tarmoq oqimida anomaliyalarini aniqlashda mashinali o‘qitishdan foydalanish // Research and Education. — 2026. — Vol. 5, Issue 4. — B. 33–44.