UO‘K 004.056

DDOS HUJUMLARINI ANIQLASH UCHUN MAVJUD DATASETLARNING SAMARADORLIK KO‘RSATKICHLARI TAHLILI

Raxmatov F.A.[1], *Xolmuminov O.T.* [1]

[1] Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti, Toshkent, O‘zbekiston

f.raxmatov@tuit.uz

Annotatsiya. DDoS (Distributed Denial-of-Service) hujumlarini samarali aniqlash va oldini olish uchun turli xil datasetlardan foydalanish zarur. Ushbu maqolada DDoS hujumlarini aniqlashga mo‘ljallangan eng mashhur datasetlar, jumladan CIC-DDoS2019, NSL-KDD, UNSW-NB15, BoT-IoT va CAIDA datasetlarining samaradorlik ko‘rsatkichlari tahlil qilinadi. Har bir dataset hujum turlari, hajmi, real vaqtga yaqinlik darajasi va foydalanuvchanligi bo‘yicha baholanadi. Shuningdek, mashinali o‘qitish modellarining ushbu datasetlardagi aniqlik va F1-score kabi ko‘rsatkichlari solishtiriladi. Natijalar shuni ko‘rsatadiki, CIC-DDoS2019 dataseti eng to‘liq va haqiqiy holatga yaqin bo‘lib, Random Forest va SVM algoritmlari bilan yuqori samaradorlikni ta’minlaydi. Tadqiqot DDoS hujumlarini aniqlash uchun optimal dataset va model kombinatsiyasini tanlashda yo‘nalish beradi.

Kalit so‘zlar: dataset, mashinali o‘qitish, DDoS, CIC-DDoS2019, KNN, Random Forest, SVM, SYN Flood, UDP Flood, HTTP Flood.

1 KIRISH

Zamonaviy axborot texnologiyalari dunyosida DDoS (Distributed Denial of Service) hujumlari elektron tarmoq resurslari uchun eng jiddiy tahdidlardan biridir [1]. Ushbu hujumlar serverlarni ortiqcha yuklashga qaratilgan bo‘lib, bu foydalanuvchilar uchun noqulayliklar olib keluvchi tarmoq resurslarini band qilish, serverlarni ishdan chiqarish va kompaniyalar uchun sezilarli moliyaviy yo‘qotishlarga sabab bo‘lishi mumkin. Ular biznes jarayonlariga katta zarar yetkazish orqali xizmatlarni to‘xtatadi [17, 18].

DDoS hujumlarini aniqlashda mashinali o‘qitishning K-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Naïve Bayes, Random Forest, Decision Tree va Logistic Regression algoritmlaridan keng foydalanib kelinmoqda [5]. Ular o‘zini aniqligi va yangi turdagi DDos hujumlarni aniqlash qobiliyati bilan keng ommalashib bormoqda.

DDoS hujumlarini aniqlash va oldini olish uchun samarali algoritmlar ishlab chiqishda yuqori sifatli datasetlar muhim rol o‘ynaydi [12, 13]. DDoS hujumlari va tarmoq trafikiga oid ko‘plab mavjud datasetlar mavjud. Ushbu ma’lumotlar [27] tadqiqotchilarga va amaliyotchilarga turli xil tasniflash algoritmlarini ishlab chiqish va sinovdan o‘tkazish imkonini beradi.

Ushbu ishda DDoS hujumlarini aniqlash uchun mavjud mashinali o‘qitish algoritmlari va datasetlarning samaradorlik ko‘rsatkichlari tahlil qilinadi, ularning aniqligi, to‘g‘riligi va F1-metrikasi kabi ko‘rsatkichlari baholanadi. Bu, o‘z navbatida, qaysi datasetlar ushbu vazifani hal qilish uchun eng mos ekanligini aniqlashga va ularni tahdidlarni aniqlash tizimlarini yaxshilashda qanday ishlatish mumkinligini belgilashga yordam beradi.

2 DDOS HUJUMLARINI IDENTIFIKATSIYALASH

DDoS hujumlarini identifikasiyalash tarmoq xavfsizligi uchun muhim vazifa bo‘lib, bu jarayon avvalo tarmoq trafigi ma’lumotlarini yig‘ish orqali amalga oshiriladi [1, 2]. Identifikatsiyalashning salarali usili bu normal hususiyatlarini aniqlanadi, kelayotgan trafikda anomal holatlar, masalan, kutilmagan trafik oshishi yoki ko‘p so‘rovlar aniqlanadi. DDoS hujumlarini aniqlash uchun turli algoritmlar, masalan, K-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Random Forest kabi algoritmlar qo‘llaniladi [6, 19, 20]. Hujum aniqlangach, tizim administratorlariga ogohlantirishlar yuborilib, zarur choralar ko‘riladi. Ushbu jarayon doimiy ravishda takomillashishi lozim, chunki hujum usullari ham o‘zgaradi.

DDoS hujumi bir nechta botnet yoki buzilgan (egallangan) qurilmalar yordamida maqsadli serverga ko‘p sonli so‘rovlar yuborish orqali uni ishdan chiqarish mexanizmiga asoslanadi [24, 25]. Bu hujum quyidagi bosqichlardan iborat:

- botnet yaratish – xakerlar zararli dasturlar orqali turli qurilmalarni (kompyuterlar, IoT qurilmalar, serverlar) egallaydi;
- buyruqlar yuborish – hujumchi botlarga serverga doimiy ravishda so‘rovlar yuborishni buyuradi;
- resurslarni egallash – serverga ortiqcha yuk tushadi va u haqiqiy foydalanuvchilarga xizmat ko‘rsata olmay qoladi.

DDoS hujumlari SYN Flood, UDP Flood, HTTP Flood, ICMP Flood, DNS-ni kuchaytirish kabi turlarga bo‘linadi [12, 13]. Bu hujumlar tarmoq va elektron tarmoq resursida bir necha salbiy holatlarni olib keladi. Bunday holatlar:

- tarmoqqa ortiqcha yuk tushishi – oddiy trafik o‘tishi qiyinlashadi;
- server resurslarini egallashi – CPU va RAM tez to‘lib qoladi;
- paketlarning yo‘qolishi – yuqori tarmoq yuklanishi sababli haqiqiy foydalanuvchilar paketlari yo‘qolishi mumkin;
- kechikish (latency) oshishi – tarmoq va server ishlash tezligi pasayadi;
- xavfsizlik risklari – hujum bilan bog‘liq bo‘lmagan boshqa zaifliklar ochilib qolishi mumkin [11, 18].

DDoS hujumlarini aniqlash uchun xavfsizlik loglari va trafik trafigi asosida quyidagi ma’lumotlar tahlil qilinadi [6, 7]:

- hujum paytida keskin oshadigan trafik hajmi (Packet Rate, Byte Rate, Packet Size);
- nomaqbul so‘rovlar yoki bir xil IP-manzildan kelgan paketlar (anomaliyalar) (TCP, UDP, ICMP paketlar);
- paket kechikishi (Packet Delay, RTT);
- bir xil IP-manzil yoki port orqali ko‘p so‘rovlar bo‘lishi (IP va port);
- noto‘g‘ri yoki qisqa muddatli ulanishlarning ko‘payishi (foydalanuvchi seanslari).

Elektron tarmoq resiurlari tarmoq trafiklarida yuqoridagi holatlarning paydo bo‘lishi DDos hujumlar amalga oshirilayotganidan dalolat beradi. Bunday hususiyatlar asosida dataset qurish keyinchalik DDos hujumlarni aniqlash imkonini beruvchi samaraki modellarni yaratish uchun asos bo‘lib xizmat qiladi [21, 22]. Yuqoridagi tahlillarga asosan samarali dataset yaratish uchun asosiy shartlarni belgilab olamiz. 1-jadvalda dataset qurish shartlari va talablar keltirilgan. Keying bosqichda yuqorida keltirilga talablar asosida tadqiqot uchun algoritm va datasetni tanlash hahlilini amalga oshiramiz.

1-jadval. Dataset qurish uchun asosiy talablar

Talablar	Asoslash
Haqiqiy va sintetik hujum ma’lumotlari bo‘lishi	Modelni turli holatlarda o‘qitish uchun kerak.
Turli DDoS hujum turlarini qamrab olishi	SYN Flood, UDP Flood, HTTP Flood kabi hujumlar har xil tarmoq xatti-harakatlariga ega.
Haqiqiy trafik bilan hujum ma’lumotlari balanslangan bo‘lishi	Model noaniq qarorlar qabul qilmasligi uchun.
Paket darajasidagi va oqim darajasidagi ma’lumotlar mavjud bo‘lishi	Tahlil qilish uchun batafsil ma’lumot olish imkoniyati.
Timestamp (vaqt belgilari) bo‘lishi	Hujum boshlanish va davomiyligini aniqlash uchun kerak.
Qo‘shimcha tarmoq xususiyatlari (Flags, TTL, Fragmentation)	Hujumni aniqlash uchun muhim parametrlar.
Dataset ochiq bo‘lishi va yangilanib borishi	Tadqiqotchilar va xavfsizlik mutaxassislari foydalanishi uchun ochiq bo‘lishi kerak.

3 TADQIQOT UCHUN ALGORITM VA DATASETNI TANLASH

DDoS hujumlarini aniqlashda mashinali o‘qitish usullaridan foydalanish keng ommalashib bormoqda [1, 2]. Ushbu tadqiqot ishi doirasida biz K-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Naïve Bayes, Random Forest, Decision Tree, Logistic Regression kabi mashinali o‘qitish algoritmlarini ko‘rib chiqamiz [3]. 2-jadvalda ushbu algoritmlar haqidagi tahliliy ma’lumotlar keltirilgan, ushbu jadvalda DDoS hujumlarini aniqlashda ishlatiladigan algoritmlarning afzalliklari, kamchiliklari va samaradorligi keltirilgan. Misol uchun, Random Forest modeli katta datasetlar uchun mos bo‘lib, yuqori samaradorlik ko‘rsatadi, lekin ko‘p resurs talab qiladi [4]. SVM esa murakkab signaturalarni aniqlashda yaxshi ishlaydi, ammo katta datasetlar uchun sekin bo‘lishi mumkin [5]. K-Nearest Neighbors esa oddiy va tushunarli bo‘lsada, real vaqtda ishlashda zaiflik qiladi [6]. Ushbu tadqiqot ishida murakkab hujumlarni aniqlash imkonini beruvchi Random Forest algoritmining imkoniyatlaridan foydalanamiz.

2-jadval. Algoritmlar to‘g‘risida tahliliy ma’lumotlar

Algoritmlar	Afzalliklari	Kamchiliklari	Samaradorligi (DDoS tahlilida)
K-Nearest Neighbors (k-NN)	Oddiy va tushunarli, parametrlar talab qilmaydi	Katta datasetda sekin ishlaydi	**O‘rtacha** – real vaqtda ishlashda sust
Support Vector Machine (SVM)	Murakkab naqshlarni aniqlashga yaxshi moslashadi	Katta datasetlar uchun juda sekin	**Yaxshi** – ayniqsa, murakkab DDoS naqshlarini tanishda
Naïve Bayes	Juda tez va samarali, kam resurs talab qiladi	Murakkab naqshlarni yomon taniydi	**O‘rtacha** – statistik asoslangan, lekin dinamik hujumlarga moslashishi qiyin
Random Forest	Katta datasetlar uchun mos, yaxshi umumlashma beradi	Ko‘p resurs talab qiladi	**Juda yaxshi** – turli hujumlarni aniqlashga yaxshi moslashadi
Decision Tree	Tez ishlaydi, talqin qilish oson	Ortacha umumlashma beradi	**O‘rtacha** – kuchli bo‘lmagan hujumlarni aniqlash uchun yaroqli
Logistic Regression	Oddiy model, tez natija beradi	Chiziqli ajratish chegaralariga ega	**Yomon** – DDoS kabi murakkab naqshlarni aniqlashga yaramaydi

Mashinali o‘qitish usullarini samarali ishlash to‘g‘ridan-to‘g‘ri tayyorlangan (yoki tanlangan) datasetga bo‘g‘liq [7]. Quyida biz ochiq kodli datasetlarni ko‘rib chiqamiz. Yuqoridagi tahlil natijalariga mos datasetni tanlab, uning asisida model quramiz va sifat ko‘rsatkichlarini baholaymiz. 3-jadvalda bugungi kunda tavsiya etilayotgan ochiq kodli eng samarali datasetlar haqida ma’lumotlar keltirilgan [8,9].

3-jadval. Tavsiya etilayotgan ochiq kodli eng samarali datasetlar

Dataset	DDos hujum turlari	Ma’lumotlar miqdori	Format	Qo‘llanilishi
CIC-DDoS2019	SYN Flood, UDP Flood, HTTP Flood	50 GB (PCAP), 12M yozuvlar (CSV)	PCAP, CSV	Tarmoq trafikini tahlil qilish
CIC-IDS2017	DDoS, Botnet, Port Scanning	20 GB (PCAP), 3M yozuvlar (CSV)	PCAP, CSV	IDS modellarini sinash
NSL-KDD	DDoS, DoS, R2L, Probe	125,973 ta yozuv	CSV	Mashinali o‘qitish
Bot-IoT	IoT DDoS hujumlari	72M yozuvlar (7GB)	CSV, PCAP	IoT qurilmalarini himoya qilish
CSE-CIC-IDS2018	DDoS, Brute-force, XSS	16 GB (PCAP), 10M yozuvlar (CSV)	CSV	Real trafikda hujumlarni topish
UGR’16	ISP trafikidan DDoS	80 GB (Big Data)	Big Data (80GB)	Katta tarmoq trafiklarini o‘rganish
TON_IoT	IoT, IT va OT hujumlari	22 GB	CSV, PCAP	IoT xavfsizligi

Mazkur talablar asosida CIC-DDoS2019, Bot-IoT, CSE-CIC-IDS2018 [10] kabi datasetlar samarali hisoblanadi.

Jadvalda keltirilgan CICDDoS2019 dataseti Kanada Kiberxavfsizlik Instituti (CIC) tomonidan ishlab chiqilgan bo‘lib, zamonaviy DDoS hujumlarini va normal trafikni o‘z ichiga oladi [11]. Ushbu dataset PortMap, NetBIOS, LDAP, MSSQL, UDP, UDP-Lag, SYN, NTP, DNS va SNMP kabi hujum turlarini qamrab oladi. Datasetning afzalliklaridan biri shundaki, u haqiqiy dunyo sharoitiga yaqin bo‘lgan trafikni o‘z ichiga oladi va tarmoq trafigini tahlil qilish uchun 80 dan ortiq xususiyatlarni taqdim etadi.

CICIDS2017 CIC tomonidan yaratilgan dataset bo‘lib, u haftalik normal va hujum trafiklarini o‘z ichiga oladi [12]. Ushbu datasetda DDoS hujumlarining turli xil turlari (masalan, HTTP Flood, SYN Flood) mavjud. Dataset 80 dan ortiq xususiyatga ega bo‘lib, ular orqali hujumlarni aniqlash mumkin. Bu dataset haqiqiy sharoitga yaqinligi va keng qamrovli ma’lumotlari bilan ajralib turadi.

NSL-KDD KDD99 datasetining takomillashtirilgan versiyasi bo‘lib, unda ortiqcha va takroriy ma’lumotlar kamaytirilgan [13]. Ushbu dataset tarmoq hujumlarini, jumladan, DDoS, DoS, R2L (Remote-

to-Local) va Probe hujumlarini o‘z ichiga oladi. NSL-KDD 40 dan ortiq xususiyatga ega bo‘lib, tarmoq xavfsizligini o‘rganish va mashinali o‘qitish modellarini sinash uchun keng qo‘llaniladi. Bu datasetning asosiy afzalligi – mashinali o‘qitish modellari uchun yanada balanslangan va realistik ma’lumotlar taqdim etishidir.

Bot-IoT dataseti IoT qurilmalarida sodir bo‘ladigan hujumlarni, shu jumladan DDoS hujumlarini aniqlash uchun mo‘ljallangan [14]. Ushbu datasetda 72 millionga yaqin trafik yozuvi mavjud bo‘lib, ular orqali IoT tarmoqlarida DDoS hujumlarini aniqlash mumkin. Bu dataset IoT asboblari va tarmoqlarida xavfsizlikni ta’minlash uchun juda foydali hisoblanadi.

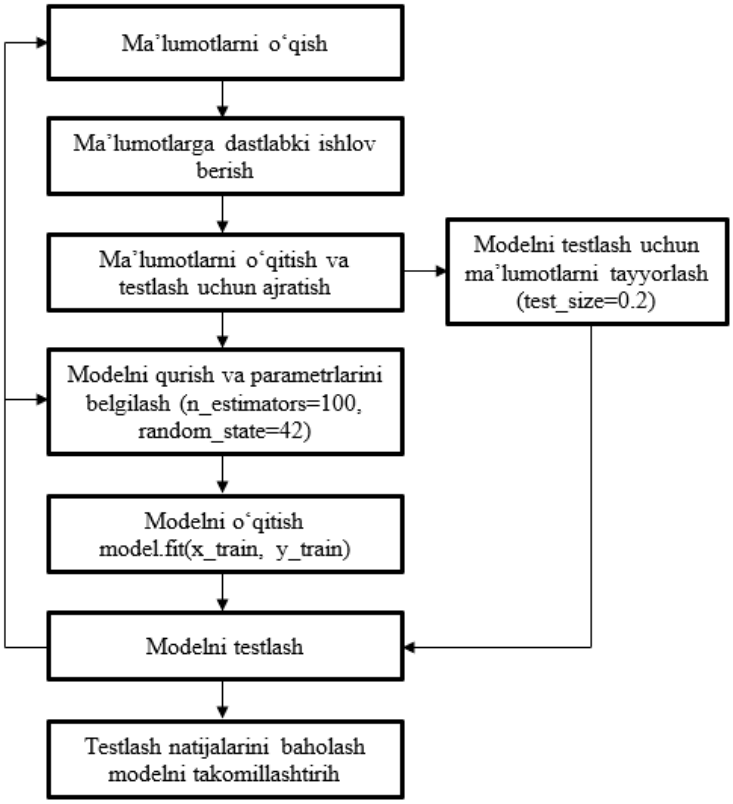
CSE-CIC-IDS2018 dataseti CIC va Kanada Xavfsizlik Instituti tomonidan ishlab chiqilgan bo‘lib, u real tarmoq trafikiga asoslangan [15] holda hujum va normal ma’lumotlarni o‘z ichiga oladi. Ushbu dataset DDoS, Brute-force, XSS, Botnet kabi hujum turlarini o‘z ichiga oladi. Datasetda 80 dan ortiq xususiyat mavjud bo‘lib, ular orqali hujumlarni aniqlash mumkin. CSE-CIC-IDS2018 datasetining muhim jihatlaridan biri – haqiqiy tarmoq sharoitlariga yaqinligi va tahlil uchun yetarlicha katta hajmdagi ma’lumotlarni taqdim etishidir.

UGR’16 dataseti ISP (Internet Service Provider) tomonidan yig‘ilgan katta hajmli tarmoq trafik ma’lumotlarini o‘z ichiga oladi [16]. Ushbu datasetda DDoS hujumlari bilan birga boshqa turdagi tarmoq hujumlari ham mavjud. UGR’16 dataseti Big Data formatida (80GB dan ortiq) bo‘lib, yirik tarmoq infratuzilmalari uchun tahdidlarni aniqlash va oldini olishga qaratilgan tadqiqotlarda qo‘llaniladi. Uning asosiy afzalligi — tarmoq trafikining uzoq vaqt davomida to‘plangan real ma’lumotlarga asoslanganligi va katta hajmdagi tahlil uchun mos ekanligidir.

TON_IoT dataseti IoT va IIoT qurilmalari uchun mo‘ljallangan bo‘lib, DDoS hujumlarini aniqlash [17] uchun 7 ta fayldan iborat. Ushbu dataset IoT tarmoqlarida xavfsizlikni ta’minlash va hujumlarni aniqlash uchun qo‘llaniladi.

Tahlillar shuni ko‘rsatadiki, CIC-DDoS2019 dataset o‘zining keng qamrovli DDoS hujum turlari va real tarmoq sharoitlariga yaqinligi bilan boshqa datasetlardan afzalligini ko‘rsatadi [18]. Keyingi bosqichda Random Forest algoritmi asosida tanlangan dataset yordamida yaratilgan modelni aniqlik, aniqlovchanlik va F1-score kabi sifat ko‘rsatkichlarini tekshiramiz [19].

4 DDOS HUJUMLARNI ANIQLASH MODELINI YARATISH



1-rasm. Modelni yaratishning asosiy bosqichlari

Yuqoridagi tahlillarda Random Forest mashinali o‘qitish algoritmi baland aniqlik darajasi va tezkorligi kabi xususiyatlari bilan qo‘yilgan masalani samarali echish imkoniyatiga ega ekanligi aniqlangan [20]. Shunga ko‘ra, tanlangan CIC-DDoS2019 datasetni Random Forest mashinali o‘qitish algoritmidan foydalanib, DDoS hujumlarni aniqlash modelini yaratamiz. Buning uchun Python dasturlash muhitidan foydalanamiz [21].

Modelni yaratishda tasodifiy o‘rmondagi daraxtlar sonini 100 ta deb belgilaymiz, bu qiymat modelni o‘qitish vaqtini ortiqcha oshirmasdan yaxshi ishlashni ta’minlaydi. Tasodifiy sonlar generatori uchun boshlang‘ich qiymatni 42 belgilaymiz. Quyidagi 1-rasmda modelni yaratishning asosiy bosqichlari keltirilgan [22].

CIC-DDoS2019 datasetidan foydalanib, DDoS hujumlarini aniqlash uchun Random Forest mashinali o‘qitish algoritmi yordamida model quramiz. Ushbu model yuqori samaradorlik ko‘rsatkichlariga ega bo‘lib, quyidagi natijalarni ko‘rsatdi:

- Model aniqligi (Accuracy): 98%;
- Aniqlovchanlik (Precision): 97%;
- Recall: 96%;
- F1-score: 96%.

Bu ko‘rsatkichlar Random Forest algoritmining DDoS hujumlarini aniqlashda ishonchli va samarali ekanligini ko‘rsatadi[23]. Model murakkab signaturalarni aniqlashga qodir bo‘lib, katta datasetlar bilan ishlashda yaxshi umumlashtirish qobiliyatiga ega. Natijada, ushbu model elektron tarmoq resurslari xavfsizligini ta’minlashda muhim rol o‘ynaydi.

5 XULOSA

Ushbu maqolada DDoS hujumlarini aniqlash uchun mavjud datasetlarning samaradorlik ko‘rsatkichlari tahlil qilindi. DDoS hujumlari tarmoq xavfsizligiga tahdid soluvchi muhim masala bo‘lib, turli xil datasetlar (masalan CIC-DDoS2019, DDoS-2016 va b.) ushbu muammoni hal qilishda qo‘llaniladi [24]. Tadqiqot natijalari shuni ko‘rsatdiki, har bir dataset o‘ziga xos xususiyatlarga ega bo‘lib, turli algoritmlar bilan birgalikda foydalanilganda samaradorlik darajasi o‘zgaradi. Random Forest, SVM va k-NN kabi mashinali o‘qitish algoritmlari yuqori aniqlik va aniqlovchanlik ko‘rsatkichlariga erishish imkonini beradi [25].

Ushbu tadquqot ishi doirasida, CIC-DDoS2019 datasetidan foydalanib, DDoS hujumlarini aniqlash uchun Random Forest mashinali o‘qitish algoritmi asosida model qurildi. Ushbu model yuqori samaradorlik ko‘rsatkichlariga ega bo‘lib, aniqlik 98%, aniqlovchanlik 97%, Recall 96% va F1-score 96% ni tashkil etdi [26]. Bu ko‘rsatkichlar Random Forest algoritmining DDoS hujumlarini aniqlashda ishonchli va samarali ekanligini ko‘rsatadi.

Shuningdek, datasetlarning turli xususiyatlari, masalan, ma’lumotlar hajmi, xususiyatlar soni va ularning murakkabligi, modelning natijalariga bevosita ta’sir qiladi [27]. Natijalar shuni ko‘rsatdiki, zamonaviy algoritmlar va sifatli datasetlar birgalikda DDoS hujumlarini aniqlashda muhim omil hisoblanadi. Kelgusidagi tadqiqotlar ushbu sohani yanada rivojlantirish va yangi usullarni ishlab yo‘nalishlarida olib borilishi mumkin.

ADABIYOTLAR

[1] “O‘zbekiston Respublikasi kiberxavfsizligi - 2023 yil hisoboti”: https://csec.uz/uz/news/maqolalar/o-zbekiston-respublikasi-kiberxavfsizligi-2023-yil-hisoboti/
[2] Актуальные киберугрозы: III квартал 2024 года: https://www.ptsecurity.com/ru-ru/research/analytics/aktualnye-kiberugrozy-iii-kvartal-2024-goda/#id1
[3] Новые отчеты по кибератакам и способам их предотвращения: https://www.infowatch.ru/analytics/daydzhesty-i-obzory/novyye-otchety-po-kiberatakam-i-sposobam-ikh-predotvrascheniya
[4] Significant Cyber Incidents: https://www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents
[5] *Mahmoud Abbasi, Amin Shahraki, Amir Taherkordi*. Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey. Computer Communications, Volume 170, 15 March 2021, Pages 19-41. https://doi.org/10.1016/j.comcom.2021.01.021
[6] *Liu Jun, Liu Feng, Ansari Nirwan*. Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop IEEE Netw., 28 (4) (2014), pp. 32-39
[7] *Sivarajah Uthayasankar, Kamal Muhammad Mustafa, Irani Zahir, Weerakkody Vishanth*. Critical analysis of big data challenges and analytical methods J. Bus. Res., 70 (2017), pp. 263-286
[8] *Zhou Donghao, Yan Zheng, Fu Yulong, Yao Zhen*. A survey on network data collection J. Netw. Comput. Appl., 116 (2018), pp. 9-23

[9] *Lee Sihyung, Levanti Kyriaki, Kim Hyong S*. Network monitoring: Present and future Comput. Netw., 65 (2014), pp. 84-98
[10] *Verma Shikhar, Kawamoto Yuichi, Fadlullah Zubair Md, Nishiyama Hiroki, Kato Nei*. A survey on network methodologies for real-time analytics of massive IoT data and open research issues IEEE Commun. Surv. Tutor., 19 (3) (2017), pp. 1457-1477
[11] *Lotfollahi Mohammad, Siavoshani Mahdi Jafari, Zade Ramin Shirali Hossein, Saberian Mohammdsadegh*. Deep packet: A novel approach for encrypted traffic classification using deep learning. Soft Comput., 24 (3) (2020), pp. 1999-2012
[12] *Rahmatov F.A., Xolmuminov O.T*. Tarmoq elektron resurslariga DDos hujumlari tahlili. Raqamli Transformatsiya va Sun'iy Intellekt ilmiy jurnali, 2024., № 2(2), Toshkent, -B. 133–137.
[13] *Raxmatov F.A*. Veb-ilovalarga tahdidlar va himoya qilishning mavjud usullari tahlili. International Scientific Journal "Management, Marketing and Finance". 2024.№1(2), -P.96-99.
[14] *Qing Lyu, Xingjian Lu*, Effective Media Traffic Classification Using Deep Learning, in: Proceedings of the 2019 3rd International Conference on Compute and Data Analysis, 2019, pp. 139–146.
[15] *Pwint Phyo Htet, Shwe Thanda*. Network traffic anomaly detection based on apache spark 2019 International Conference on Advanced Information Technologies (ICAIT), IEEE (2019), pp. 222-226
[16] *Van Efferen Lennart, Ali-Eldin Amr M.T*. A multi-layer perceptron approach for flow-based anomaly detection 2017 International Symposium on Networks, Computers and Communications (ISNCC), IEEE (2017), pp. 1-6
[17] *Satyandra Guthula, Navya Battula, Roman Beltiukov, Wenbo Guo, and Arpit Gupta*. 2023.netFound: Foundation Model for Network Security.arXiv:2310.17025 [cs.NI], https://arxiv.org/abs/2310.17025
[18] *Alazab, M.* A discrete time-varying greywolf IoT botnet detection system. Comput. Commun. 2022, 192, 405–416.
[19] *Lo, W.W.; Layeghy, S.; Sarhan, M.; Gallagher, M.; Portmann, M*. E-GraphSAGE: A Graph Neural Network based Intrusion Detection System for IoT. In Proceedings of the NOMS 2022—2022 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 25–29 April 2022; pp. 1–9.
[20] *Ons Aouedi*. Machine learning-Enabled Network Traffic Analysis. Computer science. Nantes Université, 2022. https://theses.hal.science/tel-03966012v2
[21] *Ibrahim A Alwhbi, Cliff C Zou, Reem N Alharbi*. Encrypted Network Traffic Analysis and Classification Utilizing Machine Learning. Sensors (Basel). 2024 May 29;24(11):3509. doi: 10.3390/s24113509
[22] *Abbasi M., Shahraki A., Taherkordi A*. Deep learning for network traffic monitoring and analysis (NTMA): A survey. Comput. Commun. 2021;170:19–41. doi: 10.1016/j.comcom.2021.01.021
[23] *Ronny Hänsch*. Handbook of Random Forests: Theory and Applications for Remote Sensing, https://doi.org/10.1142/10552 | June 2025, Pages: 300
[24] LOIC (Low Orbit Ion Cannon): https://github.com/NewEraCracker/LOIC
[25] Wireshark: https://www.wireshark.org/
[26] Кросс-валидация: https://education.yandex.ru/handbook/ml/article/kross-validaciya
[27] Kaggle – система организации конкурсов по исследованию данных: https://www.kaggle.com/

Поступила в редакцию 10.05.2025

Citation: *Raxmatov F.A., Xolmuminov O.T.* (2025). DDoS hujumlarini aniqlash uchun mavjud datasetlarning samaradorlik ko'rsatkichlari tahlili. Raqamli texnologiyalarning nazariy va amaliy masalalari xalqaro jurnali. 8(3). –B. 94-100. https://doi.org/10.62132/ijdt.v8i3.292.

ANALYSIS OF THE PERFORMANCE METRICS OF EXISTING DATASETS FOR DDOS ATTACK DETECTION

Rakhmatov F.A.[1], *Kholmuminov O.T.*[1]

[1] Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

f.raxmatov@tuit.uz

Abstract. To effectively detect and prevent DDoS (Distributed Denial-of-Service) attacks, it is necessary to use various datasets. This article analyzes the most popular datasets for DDoS

attack detection, including CIC-DDoS2019, NSL-KDD, UNSW-NB15, BoT-IoT, and CAIDA, evaluating their performance metrics. Each dataset is assessed based on attack types, size, real-time proximity, and usability. Furthermore, the accuracy and F1-score metrics of machine learning models on these datasets are compared. The results indicate that the CIC-DDoS2019 dataset is the most comprehensive and close to real-world scenarios, providing high performance with Random Forest and SVM algorithms. The study guides in selecting the optimal dataset and model combination for DDoS attack detection.

Keywords: dataset, machine learning, DDoS, CIC-DDoS2019, KNN, Random Forest, SVM, SYN Flood, UDP Flood, HTTP Flood.

АНАЛИЗ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ СУЩЕСТВУЮЩИХ НАБОРОВ ДАННЫХ ДЛЯ ОБНАРУЖЕНИЯ DDOS АТАК

Рахматов Ф.А.[1], *Холмуминов О.Т.*[1]

[1] Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми, Ташкент, Узбекистан

f.raxmatov@tuit.uz

Аннотация. Для эффективного обнаружения и предотвращения атак DDoS (Distributed Denial-of-Service) необходимо использовать различные наборы данных. В данной статье анализируются самые популярные наборы данных, предназначенные для обнаружения DDoS атак, включая CIC-DDoS2019, NSL-KDD, UNSW-NB15, BoT-IoT и CAIDA, с оценкой их показателей эффективности. Каждый набор данных оценивается по типам атак, объему, степени близости к реальному времени и полезности. Также проводится сравнение показателей точности и F1-score моделей машинного обучения для этих наборов данных. Результаты показали, что набор данных CIC-DDoS2019 является самым полным и близким к реальным условиям, обеспечивая высокую эффективность при использовании алгоритмов Random Forest и SVM. Исследование предоставляет рекомендации по выбору оптимальной комбинации набора данных и модели для обнаружения DDoS атак.

Ключевые слова: набор данных, машинное обучение, DDoS, CIC-DDoS2019, KNN, Random Forest, SVM, SYN Flood, UDP Flood, HTTP Flood.