TRAFIKNI TASNIFLASHNING RIVOJLANISH TENDENSIYALARI VA SHIFRLANGAN TRAFIK MUAMMOLARI

Hamdamov Utkir Raxmatullayevich
Muhammad Al-Xorazmiy nomidagi Toshkent axborot texnologiyalari, Infokommunikatsiya injinering kafedrasi professori
E-mail: utkir.hamdamov@mail.ru
Tojiyeva Feruza Qobiljon qizi
Muhammad Al-Xorazmiy nomidagi Toshkent axborot texnologiyalari, Infokommunikatsiya injinering kafedrasi doktoranti
E-mail: 11feruza@gmail.com

Annotatsiya: Maqolada tarmoq trafigini tahlil qilish va sinflashtirishning nazariy hamda metodologik asoslari yoritilgan. Internet xizmatlarining kengayib borishi natijasida trafik hajmining ortishi, shifrlangan oqimlarning ko‘payishi va ilovalar turining xilma-xilligi tarmoq boshqaruvini murakkablashtirmoqda. An’anaviy portga va paket tarkibiga asoslangan (DPI) usullar shifrlangan trafikda o‘z samaradorligini yo‘qotgani bois, mashinaviy o‘qitish va statistik yondashuvlar muqobil yechim sifatida taklif etilgan. Tadqiqotda trafikni oqim (flow) darajasida tahlil qilish, atributlarni ajratish va tanlash hamda sinflashtirish modellari uchun ma’lumotlarni tayyorlash bosqichlari metodik jihatdan bayon etilgan. Ushbu metodologiya shifrlangan trafik sharoitida ham oqimlarini aniqlash imkonini beradi hamda keyingi bosqichlarda amaliy modellashtirish uchun nazariy asos vazifasini bajaradi.

Kalit so‘zlar: tarmoq trafigi, sinflashtirish, mashinaviy o‘qitish, statistik atributlar, shifrlangan trafik, DPI, oqim asosidagi tahlil.

Аннотация: В статье рассмотрены теоретические и методологические основы анализа и классификации сетевого трафика. Рост объёмов интернет-трафика, увеличение доли зашифрованных потоков и разнообразие типов приложений усложняют управление современными телекоммуникационными сетями. Традиционные методы, основанные на номерах портов и анализе содержимого пакетов (DPI), теряют эффективность в условиях шифрования трафика, поэтому в качестве альтернативы предлагаются статистические и машинно-обучающиеся подходы. В исследовании методически описаны этапы анализа трафика на уровне потоков (flow), извлечения и отбора атрибутов, а также подготовки данных для моделей классификации. Предложенная методология позволяет выявлять типы трафика даже в условиях шифрования и служит теоретической основой для последующего этапа практического моделирования.

Ключевые слова: сетевой трафик, классификация, машинное обучение, статистические атрибуты, зашифрованный трафик, DPI, потоковый анализ.

ISSN 3030-3907

Abstract: This article presents the theoretical and methodological foundations of network traffic analysis and classification. The rapid expansion of Internet services has led to increased traffic volumes, a growing share of encrypted flows, and greater diversity of application types, all of which complicate modern network management. Traditional methods based on port numbers and Deep Packet Inspection (DPI) have become less effective under traffic encryption; therefore, statistical and machine learning approaches are proposed as alternative solutions. The study methodically describes the stages of flow-level traffic analysis, feature extraction and selection, as well as dataset preparation for classification models. The proposed methodology enables the identification of traffic types even in encrypted environments and serves as a theoretical basis for subsequent stages of practical modeling.

Keywords: network traffic, classification, machine learning, statistical features, encrypted traffic, DPI, flow-based analysis.

Kirish

Jahonda internet tarmog‘i taqdim etayotgan xizmatlarning turli xilligi, butun dunyo bo‘ylab biznes va tashkilotlarning yangi rivojlanish bosqichiga o‘tishda yangi imkoniyatlarni yaratmoqda. Internet xizmatlariga bo’lgan ehtiyoj kundan-kun ortib bormoqda. Ijtimoiy tarmoqlarda ma’lumotlar almashinuvidan tortib, global tarmoqlarda kerakli ma’lumotlarni qidirishdagi barcha amallar internet tarmog‘i bilan bog‘liqdir. AQSH, Xitoy, Avstraliya, Janubiy Koreya, Hindiston, Isroil, Kanada va boshqa mamlakatlar kabi rivojlangan mamlakatlarda tarmoq trafigini sinflashtirish va filtrlash samaradorligini oshirishga alohida e’tibor qaratilmoqda. Youtube, Netflix va MegaUpload kabi ilovalar internet tarmoq tuzilishini butunlay o’zgartirdi. Bog‘liq ravishda internet foydalanuvchilarining soni ham keskin ortib bormoqda. ITU (International Telecommunication Union) ning 2024 yilgi statistik tadqiqotlariga ko‘ra, internet foydalanuvchilarining soni 5.5 milliard, ya’ni dunyo aholisining 68 foizini tashkil qiladi. Bu ko‘rsatkich 2023 yilga nisbatan 3 foizga yuqorilagan. Bu ko‘rsatkichlar global raqamli tengsizlikni kamaytirish bilan birga, mavjud tarmoqlarda trafikni samarali boshqarish, xizmat sifati (QoS)ni kafolatlash va xavfsizlikni kuchaytirish vazifalarini dolzarb etadi.

Internet xizmatlarining kengayishi tarmoqda bir vaqtning o‘zida turli sinfdagi oqimlar ilova, video/striming va yuklab olish oqimlari harakatlanishiga olib keladi. Tashkiliy siyosatga rioya qilish, xavfsizlikni kuchaytirish va xizmat sifati (QoS)ni kafolatlash uchun bu oqimlarni aniqlash va to‘g‘ri sinflarga ajratish talab etiladi. Amaliyotda paket tarkibini o‘qishga tayanadigan DPI (Deep Packet Inspection) keng qo‘llanadi; biroq trafikning katta qismi TLS/SSL (hamda QUIC/HTTP/3) orqali shifrlangani sababli payloadga kira olmaslik, yuqori

ISSN 3030-3907

hisoblash tannarxi va maxfiylik cheklovlari DPI yondashuvining qo‘llanishini keskin cheklaydi.

Ilovalar oqimlari tashkilot siyosatiga zid bo‘lishi, ishlab chiqarish samaradorligiga salbiy ta’sir ko‘rsatishi yoki xavfsizlik xavflarini yuzaga keltirishi mumkin. Masalan, ish vaqtida ko‘ngilochar striminglar va masshtabli yuklab olish oqimlari tarmoq o‘tkazuvchanligini band etadi; ayrim xizmatlar esa ma’lumotlarni yashirin uzatish (data exfiltration), C2-kanallar bilan aloqa yoki tarmoqni skaner qilish kabi xatti-harakatlarni maskalashi mumkin. Bunday oqimlar istalmagan yoki potensial xavfli oqimlar toifasiga kiritilib, ularni aniq ajratish siyosatga muvofiqlik, xavfsizlik va QoS boshqaruvi uchun zarurdir. Bundan tashqari, billing (hisob-kitob) uchun ham tasniflash muhim: trafikni ilova yoki xizmatlar bo‘yicha ajratish to‘lovlarni to‘g‘ri hisoblashga yordam beradi, limitdan oshish yoki noto‘g‘ri foydalanish holatlarini erta ko‘rsatadi, bo‘limlar bo‘yicha xarajatni adolatli taqsimlashni osonlashtiradi va daromad yo‘qotilishining oldini oladi. Trafik real vaqt rejimida kuzatilsa, billingning aniqligi va shaffofligi yanada oshadi.

Bugungi kunda bu muammolarni hal qilishga yo‘naltirilgan ko‘plab tijoriy va ochiq kodli yechimlar mavjud. Ularning asosiy qismi DPIga tayangan bo‘lib, paket mazmunini tahlil qiladi; biroq keng tarqalgan TLS/SSL/QUIC shifrlanishi sharoitida mazkur yondashuvlar resurs talabchan, maxfiylik nuqtayi nazaridan noqulay va ko‘p hollarda amaliy samarasini yo‘qotadi.

Internet-trafik ko‘p oqimlarning ustma-ust dinamik superpozitsiyasidan iborat bo‘lib, ular turli protokollar tomonidan generatsiya qilinadi. Umumiy maqsad: oqimni ilova/xizmat turiga sinflashtirish. Bunda tasnif sifatini baholash uchun miqdoriy ko‘rsatkichlar qo‘llanadi: to‘g‘ri aniqlash ulushi (accuracy), noto‘g‘ri ijobiy va salbiy qarorlar (false positive/false negative, ortiqcha bloklash), shuningdek amaliy jihatdan vaqt ko‘rsatgichlari qo‘llaniladi.

Mashinaviy o‘qitish (ML) yondashuvlari ushbu vazifada ayniqsa istiqbolli: ular o‘zgaruvchan internet muhitiga moslashish va trafik xususiyatlarini hisobga olish imkonini beradi.

ADABIYOTLAR TAHLILI

Moore va Papagiannaki [1] tadqiqotchilar tomonidan portga asoslangan yondashuvning bir qancha kamchiliklari ko‘rildi. Ular Genome - biologiya sohasidagi ilmiy tadqiqot kompusi tarmog‘idan o‘tayotgan traffikni o‘nta parameter asosida tasniflashgan. Natijada umumiy trafikdan 30-70% to‘g‘ri aniqlanmagan. Hamda, IP pog‘onasida shifirlanish TCP va UDP headerlarini adashtirilishiga olib kelgan va ularni aniq port raqamlarini aniqlash imkoni bo‘lmagan.

ISSN 3030-3907

Sen S., Spatschek O., Wang D.[2] taqdiqotchilar tomonidan P2P trafiklari uchun portga asoslangan yondashuvni samaradorligini tekshirib chiqilgan. Natijada tasniflagan traffik (VPN va VPN emas tarmoq traffiklari) 30% aniqlikni ko‘rsatdi.

Foydali yuklama asosida trafikni sinflashtirish va filtrlashda aniqlik ko‘rsatgichlarini oshirishda bir qancha izlanishlar olib borolgan. Sen S., Spatschek O., Wang D [3] tadqiqotchilari tomonidan beshta P2P ilovalari (BitTorrent, eDonkey, Gnutella, Kazaa va Direct Connec) uchun protokol imzosi yig‘ildi. Yig‘ilgan imzolar asosida P2P trafigini dastlabki 10 ta paketlarini tahlil qilgan holda, samarali va aniq aniqlay oladigan online filtr ishlab chiqilgan. Ushbu texnika yordamida portga asoslangan yondashuvga nisbatan P2P ilovasi (Kazaa) 3 barobar ko‘p trafikni aniqlay olgan.

A.W.Moore,K.Papagiannaki [1] tadqiqotlarida, birinchi K bayt oqim yuklamasi, protokollar imzosi va protokolarni sematik tahlili kabi bir qancha usullar yordamida tarmoq trafiklarni tasniflashgan.

Williams [4] tarmoq trafigini sinflashtirishda Bayes Net, C4.5, AdaBoost C4.5, NBTree, AdaBoost NBD va Nearest Neighbor algoritmlarini ishlash samaradorligini taqqoslagan. Shuningdek, tadqiqotda CFS (Correlation-based Feature Selection), Wrapper va Consistency kabi atribut tanlash algoritmlarini ham tahlil qilingan. Tadqiqotda National Laboratory for Applied Network Research (NLANR) tomonidan taqdim etilgan tarmoq trafiklari to‘plamidan foydalanilgan bo‘lib, HTTP, SMTP kabi protokollarni aniqlash uchun model qurilgan. Ushbu protokollarni aniqlashda mualliflar paket uzunligi va paketlar orasidagi vaqt (inter-arrival time) kabi statistik atributlarni samarali deb isbotlangan.

Jenefa[5] C5.0 algoritmini C4.5, SVM va Naïve Bayes (NB) o‘quv algoritmlari bilan taqqoslashgan. Ular tomonidan ajratilgan atributlar: paket tezligi (packet rate), ma’lumot uzatish tezligi (data rate), paketlar orasidagi vaqt statistikasi (minimum, o‘rtacha, maksimal va standart og‘ish). Eksperimental baholash laboratoriya sharoitida to‘plangan ma’lumotlar asosida olib borilgan bo‘lib, bu to‘plam 17 ta turli xil ilova trafigini o‘z ichiga olgan. Tadqiqot natijalari shuni ko‘rsatdiki, C5.0 algoritmi ko‘plab sinflar uchun 99% aniqlik (precision) ko‘rsatkichlariga erishgan va qolgan algoritmlardan ustun bo‘lgan.

Dias [6] real vaqt rejimidagi (real-time) ilovalarni, ayniqsa video tarmoq trafigini tasniflash uchun mashinali o‘qitishga asoslangan maxsus frameworkni taklif qilishgan. Ushbu framework 13 ta statistik atribut asosida o‘qitilgan bo‘lib, ular orasida quyidagilar mavjud: paket yetib kelish vaqti (packet arrival time), o‘nlik qiymatlarning o‘rtacha qiymati (average of the decimal value), IP datagramma uzunligining o‘rtacha qiymati (average IP datagram length). Ma’lumotlar to‘plami mualliflarning laboratoriyalari sharoitida to‘plangan bo‘lib,

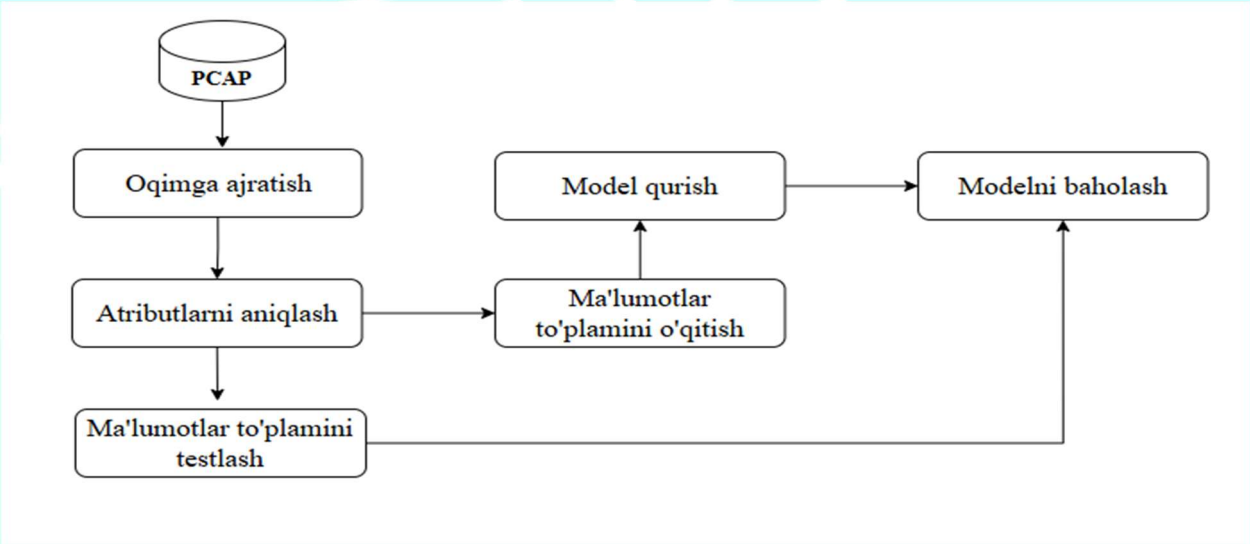
ISSN 3030-3907

YouTube va Netflix videolari, shuningdek yuklab olingan fayllar trafigini o‘z ichiga olgan. Taklif etilgan algoritm 98.88% aniqlik (accuracy)ga erishgan va real vaqtli ilovalarni tasniflashda yuqori samaradorlik ko‘rsatgan. Shuningdek, ushbu model Naïve Bayes (NB) algoritmi bilan solishtirilganda, modelni qurish va sinflashtirish bosqichlari uchun kamroq vaqt talab qilgan.

Alshammari [7] o‘z tadqiqotlarida tarmoq oqimlarini tasniflashda uchta mashinali o‘qitish algoritmi: AdaBoost, C4.5 va Genetik Dasturlash (Genetic Programming, GP) algoritmlarining samaradorligini tahlil qilganlar. Eksperimentda oqim davomiyligi, paket uzunligi, hamda paketlar orasidagi yetib kelish intervali kabi statistik atributlar, ikkala oqim yo‘nalish bo‘yicha hisobga olingan.

TADQIQOT METODOLOGIYASI

Modelni samarali ishlashi uchun trafikni tuzish, atributlarni ajratish va ularni o‘qitish/testlash uchun mos formatga keltirish zarur. Quyidagi diagrammada (1-rasmda) ushbu jarayonning bosqichma-bosqich ishlash tartibi keltirilgan. Trafik PCAP formatida yig‘ilib, oqimlarga ajratiladi. Har bir oqimdan statistik atributlar ajratiladi, ular asosida o‘qitish va test to‘plamlari shakllantiriladi. Shundan so‘ng model qurilib, baholanadi.



1-rasm.Trafikni sinflashtirish modeli qurish va baholash bosqichlari

Bu jarayon natijasida yaratilgan model real vaqt rejimida yoki oflayn tarzda trafik turlarini aniqlashda qo‘llanilishi mumkin. Baholash natijalariga qarab model takomillashtiriladi yoki moslashtirilgan variantlar ishlab chiqiladi.

Tarmoq trafigini mashinali o‘qitish yordamida tasniflashda quyidagi bosqichlar ketma-ketligi amalga oshirildi: ma’lumotlarni yig‘ish, atributlarni aniqlash, modelni o‘qitish, hamda tasniflash va baholash bosqichlari.

Ma’lumotlarni yig‘ish (PCAP) - bu tarmoq trafigini tasniflash tizimining dastlabki va eng muhim tarkibiy bosqichidir. Ushbu bosqichning asosiy vazifasi tarmoq infratuzilmasida kechayotgan real ma’lumotlar almashinuvi jarayonini aks

ISSN 3030-3907

ettiruvchi xom trafik (ya'ni IP-paketlar)ni aniq, to'liq va tizimli shaklda yig'ish, saqlash hamda keyingi ishlov berish uchun tayyorlashdan iborat.

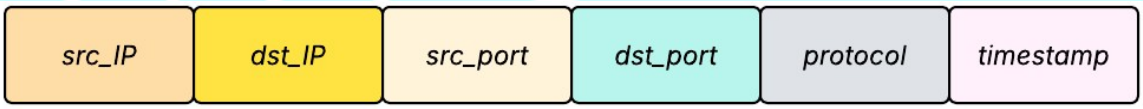
Ma'lumotlarni yig'ish ikkita asosiy yondashuvga tayanadi:

1. Xususiy (real) ma'lumotlarni yig'ish - bu yondashuvda trafik na'munalari kompyuter, IoT qurilmasi yoki mobil telefoni orqali tanlangan ilovalarni ishlatilganda, maxsus qurilma va dastur yordamida yozib oladi.
2. Ommaviy (ochiq) ma'lumotlar to'plamidan foydalanish - bu yondashuvda oldin boshqa tadqiqotchilar tomonidan yig'ilgan va ommaga taqdim etilgan ma'lumotlar bazalaridan (masalan, ISCX VPN-nonVPN dataset, USTC-TFC dataset, CIC-IDS dataset va boshqalar) foydalaniladi.

Yig'ilgan trafik odatda PCAP (Packet Capture) formatida saqlanadi. Bu fayllarda har bir paketning sarlavhasi, vaqti, uzunligi, protokol turi va boshqa muhim atributlari mavjud bo'ladi.

Yig'ilgan xom trafik keyingi bosqichda oqimlarga (flow) ajratiladi. Oqim bu manba IP, manzil IP, manba porti, manzil porti, protokol va vaqt oralig'i asosida aniqlanadigan ketma-ket paketlar yig'indisidir. Har bir oqim alohida trafik sessiyasini ifodalaydi.

Oqim ko'rinishida ifodalash tarmoqdagi har bir ilova tomonidan yaratilgan aloqani ajratib olish imkonini beradi va u orqali statistik atributlar aniqlanishi osonlashadi. Odatda quyidagi shaklda oqimlar aniqlanadi:



2-rasm. Sessiyani aniqlash uchun ishlatiladigan asosiy atributlar: manba IP (src_IP), manzil IP (dst_IP), manba porti (src_port) va manzil porti (dst_port)

Bu bosqichda har bir oqim uchun statistik ko'rsatkichlar (paket soni, o'rtacha paket uzunligi, sessiya davomiyligi, inter-paket vaqtlar, umumiy trafik hajmi va boshqalar) hisoblab chiqiladi. Ushbu atributlar keyinchalik model qurish uchun asosiy kirish ma'lumotlari bo'lib xizmat qiladi.

Tarmoq trafigini sinflashtirish tizimlarida atributlarni aniqlash - sinflashtirishning umumiy samaradorligiga bevosita ta'sir ko'rsatadigan muhim bosqich hisoblanadi. Ushbu bosqichda har bir oqimdan ajratib olingan statistik xususiyatlar hisoblab chiqiladi. Bu ko'rsatkichlar yig'ilgan trafik oqimlarining xarakterini tavsiflab, trafik turlarini farqlashda asosiy rol o'ynaydi. Har bir trafik turining o'ziga xos xususiyatlari bor va ularni ajratib olish uchun ba'zi atributlar ko'proq, ba'zilari esa kamroq foydali bo'ladi.

Atributlarni ajratib olish (feature extraction) - bu har bir ilovaga xos statistik belgilarni aniqlash orqali turli trafik turlarini farqlash imkonini beradi. Atributlarni

ISSN 3030-3907

kamaytirish (feature reduction) esa keraksiz yoki takrorlanuvchi atributlarni olib tashlab, model aniqligini oshiradi va hisoblash yuklamasini kamaytiradi.

Bugungi kunda atribut tanlashda quyidagi metodlar keng qo‘llaniladi:

Wrapper usullari - tasniflagichni turli atributlar to‘plamida sinab ko‘rib, eng samarali to‘plamni aniqlaydi.

Filter usullari - atributlar orasidagi statistik munosabatlarga tayanadi va tasniflagichni hisobga olmaydi.

CFS (Correlation-based Feature Selection) [8] - har bir atributning label bilan korrelyatsiyasi va atributlararo o‘zaro bog‘liqlikni baholaydi.

CON (Consistency-based subset search) [9] - sinfga bog‘liqlikda barqarorlikka asoslanib, ishonchli atributlarni aniqlaydi.

Gain Ratio (GR) - atributlarning klassga bo‘lgan ta’sirini baholaydi va Information Gain (IG) asosida yaxshilangan reyting beradi [10].

Information Gain (IG) - atributlarning entropiyani kamaytirish qobiliyatini baholaydi, ya’ni sinf farqliligini aniqlovchi eng muhim atributlarni topadi.

Ma’lumotlar to‘plamini tayyorlash (Datasets preparation):

Atributlar ajratilib, ularni tanlash bosqichi yakunlangach, sinflashtirish uchun zarur bo‘lgan trafik namunalaridan iborat yakuniy ma’lumotlar to‘plami shakllantiriladi. Ushbu to‘plam tarmoq trafigini sinflashtirish modelini qurish va testlash uchun asos bo‘ladi. Bu bosqichda odatda ikki yondashuvdan foydalaniladi:

1. Alohida o‘qitish (training) va test (testing) to‘plamlaridan foydalanish;
2. N-qatlamli (N-fold) kross-validatsiya.

Alohida o‘qitish va test to‘plamlaridan foydalanilganda, ikkita mustaqil ma’lumotlar to‘plami kerak bo‘ladi. O‘qitish bosqichi uchun katta hajmdagi ma’lumotlar talab qilinadi, chunki bu modelning aniqligi va barqarorligi uchun muhim. Bu bosqichda mashinaviy o‘qitish algoritmiga statistik atributlar bilan birga, yorliqlangan namunalar taqdim etiladi. Algoritm ushbu ma’lumotlar asosida trafik sinflari o‘rtasidagi naqsh va qonuniyatlarni o‘rganadi.

Test bosqichi - ilgari ko‘rilmagan namunalarni qurilgan model orqali sinflarga ajratish jarayonidir. Test to‘plami ham yetarli hajmda bo‘lishi kerak, aks holda baholash natijalari ishonchsiz bo‘lishi mumkin.

Shuningdek, bitta to‘plamdan o‘qitish va testlash uchun foydalanish mumkin emas, chunki u modelning umumlashtirish (generalization) qobiliyatini noto‘g‘ri baholaydi. Bunday holatlarda N-qatlamli kross-validatsiya usuli qo‘llaniladi.

Bunda:

- Ma’lumotlar to‘plami N ta teng qismga bo‘linadi;
- Har safar N−1 qism o‘qitish, 1 qism esa test uchun olinadi;
- Bu jarayon N marta takrorlanadi, har safar boshqa qatlam test sifatida olinadi.

ISSN 3030-3907

O‘rganilgan tadqiqotlarga ko‘ra, N = 10 bo‘lgan holatlarda sinflashtirish natijalari yetarli aniqlikka ega bo‘ladi.

Modelni qurish (Model Building). Avvalgi bosqichda shakllantirilgan o‘qitish (training) to‘plami asosida, tarmoq trafigini ularning ilova turlari yoki protokollariga ko‘ra avtomatik ravishda tasniflay oladigan model yaratiladi.

Tadqiqot talablari asosida ushbu dissertatsiyada tajribalar sifatini baholash uchun bir nechta ko‘rsatkichlardan foydalaniladi. Tarmoq trafigini tasniflash bo‘yicha ilmiy adabiyotlarda eng ko‘p qo‘llaniladigan ko‘rsatkich aniqlik (accuracy) hisoblanadi.

Aniqlik ko‘rsatkichi to‘g‘ri tasniflangan oqimlar sonini hisoblash orqali aniqlanadi. Non-binar tasniflash vaziyatida aniqlik ko‘rsatkichining aniq ta’rifi quyidagicha ifodalanadi:

$$Accuracy = \frac{\sum_{i=1}^{N}(TP)}{\sum_{i=1}^{N}(TP)+\sum_{i=1}^{N}(FP)} \quad (1)$$

N: Kategoriyalar soni (misol uchun ilovalar guruhlari);

TP (True Positives): ma’lum bir kategoriya uchun to‘g‘ri aniqlangan oqimlar soni.

FP (False Positives): ma’lum bir kategoriya uchun noto‘g‘ri aniqlangan oqimlar soni.

Odatda qo‘llaniladigan boshqa ko‘rsatkichlar: bu aniqlik (precision) (ya’ni, ijobiy bashorat aniqligi) va to‘liq qamrov (recall) (ya’ni, sezgirlik) kategoriyalar bo‘yicha hisoblanadi.

Precision quyidagicha aniqlanadi:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (2)$$

Bu ko‘rsatkich ma’lum bir kategoriyaga tegishli deb tasniflangan namunalar ichida haqiqatan ham to‘g‘ri tasniflanganlari ulushini bildiradi. Shu bilan birga, ma’lum bir kategoriyaning recall ko‘rsatkichi quyidagicha aniqlanadi:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (3)$$

bu yerda FN (False Negatives) - aslida o‘rganilayotgan kategoriyaga mansub bo‘lsa-da, boshqa kategoriyaga tasniflangan oqimlar soni. Ushbu ko‘rsatkich ma’lum bir kategoriyaga mansub bo‘lgan barcha namunalar ichida to‘g‘ri tasniflanganlar ulushini bildiradi.

Shunga qaramay, ushbu ko‘rsatkichlar tarmoq trafigini tasniflash adabiyotida eng ko‘p ishlatiladigan ko‘rsatkichlar bo‘lsa-da, ularning ayrim cheklovlari mavjud. Shu sababli, olingan natijalarni tasdiqlash uchun qo‘shimcha ko‘rsatkichlarni hisoblash zarur bo‘ladi.

Kappa koeffitsienti ko‘rsatkichi yanada ishonchli deb hisoblanadi, chunki u tasodifiy to‘g‘ri tasniflash holatlarini ham hisobga oladi. Kappa koeffitsientini [11] quyidagi formula asosida hisoblandi:

ISSN 3030-3907

$$k = \frac{P_0 - P_c}{1 - P_c} \quad (4)$$

$$P_0 = \frac{\sum_{i=1}^{N}(TP)}{\sum_{i=1}^{N}(TP) + \sum_{i=1}^{N}(FP)} \quad (5)$$

$$P_e = \sum_{i=1}^{N}(P_{i1} \times P_{i2}) \quad (6)$$

P_{i1}: birinchi kuzatuvchi uchun i-kategoriyaning paydo bo‘lish ulushi.

P_{i2}: ikkinchi kuzatuvchi uchun i-kategoriyaning paydo bo‘lish ulushi.

Kappa koeffitsientining qiymati agar tasniflash asosan tasodifiy moslik natijasida yuzaga kelgan bo‘lsa, 0 ga yaqin bo‘ladi. Aksincha, agar tasniflash usuli kuchli farqlash qobiliyatiga ega bo‘lsa, qiymatlar 1 ga yaqinlashadi.

Yana bir muhim ko‘rsatkich bu - to‘liqlik (completeness). Ushbu ko‘rsatkich ma’lumotlar to‘plamidagi tasniflanmagan trafik miqdorini o‘lchaydi. Ya’ni, tasniflashdan so‘ng noma’lum (unknown) sifatida qoladigan trafikning foizini bildiradi.

Ground-truthni shakllantirish tarmoq trafigini tasniflashni baholashning eng muhim bosqichlaridan biridir. Chunki butun tasniflash jarayoni dastlabki yorliqlash (labeling)ni olish uchun qo‘llanilgan usulning ishonchliligiga tayanadi. Adabiyotlarda uchta asosiy yondashuv qo‘llaniladi.

Eng aniq usul bu - ma’lumotlar to‘plamini inson tomonidan qo‘lda yorliqlashdir. Biroq oqimlar bo‘yicha qo‘lda yorliqlash juda mashaqqatli jarayon bo‘lib, millionlab oqimlarni o‘z ichiga olgan joriy ma’lumotlar to‘plamida amalda qo‘llash qiyin.

MUHOKAMA VA NATIJALAR

Tadqiqot davomida tarmoq trafigini tahlil qilishning dolzarb muammolari va ularni hal etishda qo‘llanilayotgan asosiy yondashuvlar o‘rganildi. Trafikni real vaqt rejimida kuzatish, tahlil qilish va sinflashtirish imkonini beruvchi zamonaviy texnologiyalar, xususan, chuqur o‘rganish (Deep Learning) modellari yordamida aniqlikni oshirishga erishilmoqda.

O‘rganilgan ilmiy manbalar shuni ko‘rsatadiki, an’anaviy portga yoki paket tarkibiga (DPI) asoslangan usullar shifrlangan trafikni aniqlashda tobora samarasiz bo‘lib bormoqda. Buning sababi internet xizmatlarining ko‘pchiligi TLS/SSL yoki QUIC kabi protokollar orqali shifrlanayotgani, bu esa paketlarning foydali yuklamasini tahlil qilish imkonini cheklaydi. Shu bois, transformatorlarga asoslangan modellar, avtomatik atribut ajratish (feature extraction) va ko‘p bosqichli o‘rganish texnikalari real vaqtda trafikni yuqori aniqlikda tasniflash imkoniyatini bermoqda.

Ushbu maqola doirasida tarmoq trafigi bilan bog‘liq asosiy muammolar trafik hajmining keskin ortishi, shifrlangan oqimlarning keng tarqalishi va trafik

ISSN 3030-3907

xususiyatlarining o‘zgaruvchanligi alohida e’tiborga olindi. Tahlillar shuni ko‘rsatadiki, bu muammolarga qarshi texnologik yechimlar shakllanayotgan bo‘lsa-da, ularni real tarmoqlarda amaliy qo‘llash, moslashtirish va hisoblash samaradorligini oshirish masalalari hali ham dolzarb bo‘lib qolmoqda.

Shunday qilib, tarmoq trafigini statistik atributlar va mashinaviy o‘qitish metodlariga asoslangan holda sinflashtirish yondashuvi, ayniqsa shifrlangan trafik sharoitida, eng istiqbolli va amaliy jihatdan qulay yo‘nalishlardan biri ekanligi aniqlanadi. Ushbu nazariy asoslar keyingi bosqichlarda amaliy modellashtirish va sinov ishlari uchun poydevor vazifasini bajaradi.

XULOSA VA TAKLIFLAR

Tarmoq trafigini tahlil qilish va sinflashtirish bugungi kunda yuqori sifatli, uzluksiz va xavfsiz Internet xizmatlarini ta’minlashda muhim ahamiyat kasb etadi. Trafik oqimlarini aniqlash, ulardagi ilova turlarini farqlash hamda shifrlangan trafikni samarali tahlil qilish zamonaviy telekommunikatsiya tarmoqlarining barqaror ishlashi uchun zarurdir.

Trafikni boshqarish jarayonida to‘plangan ma’lumotlarni oqim darajasida (flow-based) tahlil qilish usullari eng samarali hisoblanadi. Ayniqsa, statistik atributlarga tayanadigan yondashuvlar (paket uzunligi, inter-paket vaqt, sessiya davomiyligi va oqim tezligi kabi ko‘rsatkichlar) trafikni shifrlangan sharoitda ham farqlash imkonini beradi. Chuqur paket tahlili (DPI) yondashuvi aniq natijalar berishi mumkin, biroq TLS/SSL/QUIC kabi protokollar orqali shifrlangan trafikda uning qo‘llanilishi chegaralangan. Shu sababli, mashinaviy o‘qitish va chuqur o‘rganish (Deep Learning) modellaridan foydalanish tarmoq trafigini real vaqt rejimida yuqori aniqlik bilan sinflashtirishga imkon beradi.

Ushbu tadqiqotda keltirilgan nazariy tahlillar va metodologik yondashuvlar tarmoq operatorlari va xizmat ko‘rsatuvchi provayderlar uchun trafikni samarali nazorat qilish, xizmat sifati (QoS)ni saqlash, va anomaliyalarni erta aniqlash imkonini beradi. Trafikni sinflashtirishning statistik va mashinaviy o‘qitish asosidagi modellarini joriy etish orqali tarmoq resurslaridan foydalanishni optimallashtirish hamda xizmatlarni ustuvorlik darajasi bo‘yicha boshqarish imkoniyati yaratiladi.

Takliflar:

1. Statistik atributlar asosidagi avtomatlashtirilgan trafik tahlilini mashinaviy o‘qitish modellariga integratsiya qilish real vaqt rejimida trafikni aniqlash va bashorat qilish samaradorligini oshiradi.
2. Shifrlangan trafik uchun DPI va NetFlow ma’lumotlariga asoslangan chuqur o‘rganish modellarini joriy etish oqim xatti-harakatlarini tahlil qilish orqali aniqlikni oshirishga xizmat qiladi.

ISSN 3030-3907

3. Trafik sinflashtirish natijalarini filtrlash va prioritetlash tizimlariga bog‘lash orqali xizmat sifati (QoS) va xavfsizlik siyosatlarini avtomatik boshqarish imkonini yaratish zarur.

4. Atribut tanlash va optimallashtirish algoritmlarini (CFS, Information Gain, Wrapper) qo‘llash modelning hisoblash yuklamasini kamaytiradi va aniqlikni oshiradi.

5. Trafikni vizualizatsiya va monitoring qilish uchun qulay interfeyslar yaratish administratorlar uchun qaror qabul qilishni tezlashtiradi va tizim holatini intuitiv nazorat qilishni ta’minlaydi.

Shunday qilib, raqamli infratuzilmaning kengayib borishi bilan birga, tarmoq trafigini tahlil qilish va sinflashtirish bo‘yicha moslashuvchan, avtomatlashtirilgan va intellektual yechimlarni ishlab chiqish dolzarb vazifalardan biri bo‘lib qolmoqda. Kelgusida bunday yondashuvlarni amaliy tarmoqlarda joriy etish orqali trafikni boshqarish, xizmat sifati va xavfsizlik darajasini sezilarli darajada oshirish mumkin bo‘ladi.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. Moore, A. W., & Papagiannaki, K. (2005). *Toward the accurate identification of network applications*. Passive and Active Network Measurement, 41–54. Springer.

2. Sen, S., Spatschek, O., & Wang, D. (2004). *Accurate, scalable in-network identification of P2P traffic using application signatures*. Proceedings of the 13th International Conference on World Wide Web (WWW), 512–521.

3. Sen, S., Spatschek, O., & Wang, D. (2004). *Analyzing peer-to-peer traffic across large networks*. *IEEE/ACM Transactions on Networking*, 12(2), 219–232.

4. Williams, N., Zander, S., & Armitage, G. (2006). *A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification*. *ACM SIGCOMM Computer Communication Review*, 36(5), 5–16.

5. Jenefa, J., & Kannan, A. (2019). *Comparative analysis of C5.0, C4.5, SVM, and Naïve Bayes algorithms for application traffic classification*. *International Journal of Computer Applications*, 182(12), 1–8.

6. Dias, M. A., Oliveira, L. S., & De Carvalho, A. C. P. L. F. (2021). *A machine learning-based framework for real-time multimedia traffic classification*. *Journal of Network and Computer Applications*, 175, 102935.

7. Alshammari, R., & Zincir-Heywood, A. N. (2009). *A flow-based approach for VPN traffic detection using machine learning*. *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 1–8.

8. Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, Hamilton, New Zealand.

ISSN 3030-3907

9. Liu, H., & Setiono, R. (1996). *A probabilistic approach to feature selection – A filter solution*. *Proceedings of the 13th International Conference on Machine Learning (ICML)*, 319–327.
10. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers.
11. Cohen, J. (1960). *A coefficient of agreement for nominal scales*. *Educational and Psychological Measurement*, 20(1), 37–46.

ISSN 3030-3907