IP forgalom analízis

Kún Gergely

Óbudai Egyetem, Kandó Kálmán Villamosmérnöki Kar, Híradástechnika Intézet
Budapest, 1084 Tavaszmező utca 17.
kun.gergely@kvk.uni-obuda.hu

Összefoglaló: Adathálózatokban a közös erőforrások védelmének, a túlterhelés elkerülésének érdekében sok esetben találkozhatunk az előfizetői, felhasználói forgalmakat érintő forgalomszabályozási technikákkal. Ezek hatékony megvalósításához a hálózati forgalom sokrétű és sokszor több lépcsős vizsgálatára van szükség. A forgalmak osztályozásának célja, hogy a felhasználók által futtatott alkalmazásokat azonosítsuk, és ez alapján pontos és mérvadó információkhoz jussunk, melyek felhasználhatók a hálózat tervezéséhez, optimalizálásához, kapacitás becsléshez, számlázási csomagok kialakításához és biztonsági monitorozáshoz.

Keywords: forgalom osztályzás, alkalmazás azonosítás, IP hálózatok, forgalom menedzsment

1 Bevezetés

Az internet használat széles körű terjedése, az adathálózatok forgalmának növekedése egyre inkább fókuszba helyezi a kiszolgáló hálózatok által nyújtott minőségi kérdéseket, hatékonyságot és a minél többféle hálózati szolgáltatások elérhetőségét. A forgalom menedzsment hatékony és ésszerű megvalósításához elengedhetelen fontosságú napjainkban a hálózati forgalmak osztályozása, és az abból származó információk ismerete és megfelelő hasznosítása, amit jelenthet a felhasználók sávszélességének (bitsebességének) szabályozása, teljesítmény optimalizáció, bizonyos forgalmak priorizálása (pl. VoIP), vagy káros (pl. vírusok) forgalmak felismerése és blokkolása a felhasználók megóvása céljából.

A szélessávú internet hozzáféréseket nyújtó szolgáltatók hálózatán keresztülhaladó forgalom összetétele meglehetősen heterogén: a klasszikus elektronikus levélküldő és böngésző forgalmakon kívül számos népszerű alkalmazás kommunikációja zajlik. Forgalom menedzsment nélkül a felhasználók forgalmai az aggregált linkeken azonos prioritással rendelkeznek – mindenki az úgynevezett legjobb szándékú (best effort) csomagkezelést kapja a hálózat részéről – mindössze csak a maximális hozzáférési sebesség korlátozott az előfizetői csomagok függvényében.

Szolgáltatói szinten, az előfizetői forgalmak osztályozása, felismerése egy automatizált rendszerrel mind az előfizetői szokásokba való belelátást, mind a hálózat működését, mind a globális felhasználói elégedettséget előnyösen érintheti. A különböző alkalmazások forgalmai viszont különböző jellegű átviteli követelményeket támasztanak a hálózattal szemben: egy interaktív párbeszéd esetén a rövid válaszidő alapvető fontosságú, egy program letöltésénél a kellően nagy sávszélesség a fontos, míg egy böngésző forgalom normális válaszidővel igen rugalmasan alkalmazkodhat a hálózat forgalmi helyzetének ingadozásaihoz.

A best effort alapú, azonos prioritással rendelkező folyamok továbbításának hátránya, hogy az átviteli mód gyengeségeit kihasználó, agresszívebb hálózati kommunikációt folytató programok (tipikusan P2P technológián alapuló tartalommegosztó alkalmazások) a közös csatornán dominánsan képesek jelen lenni, ezzel hosszabb ideig, jelentősen ronthatják a többi felhasználó által tapasztalt minőségi paramétereket (QoS - Quality of Service), illetve a felhasználók szubjektív szolgáltatás-minőség megítélését (QoE - Quality of Experience) [1]. A szolgáltatók alapvető kötelessége és érdeke, hogy az előfizetők számára a megfelelő QoS jellemzőket folyamatosan teljesítsék, így például a fenti alkalmazások forgalmait lehetőségekhez mérten azonosítsák és szükség esetén korlátozzák.

A P2P technológia sokoldalú felhasználhatósága miatt számos olyan alkalmazás működik ilyen alapon, melyek fejfájást okoznak a szolgáltatóknak: fájlmegosztók, hang és videó hívásokat lehetővé tevő VoIP alapú alkalmazások (pl. Skype), chatprogramok. A fájletöltések, vagyis a fájlmegosztó alkalmazások forgalmai, a hálózat terheltségét befolyásolják negatívan, sok esetben okoznak problémát mind a gerinc, mind a hozzáférési szegmensben. A kommunikációs alkalmazások a szolgáltatók alapszolgáltatásaival konkurálnak, sokszor jobb minőséggel és kedvezőbb árral (akár ingyenességükkel) csábítják el a felhasználókat. A mobil eszközök és okostelefonok elterjedésének köszönhetően, napjainkban már nemcsak a vezetékes, hanem a mobil szolgáltatók hálózati forgalmát is jelentős mértékben meghatározzák ezek az alkalmazások. A szolgáltatóknak így szükséges ezen alkalmazások minél pontosabb feltérképezése, naprakész nyilvántartása, adott esetben egyes típusok korlátozása, vagy teljes letiltása, üzletpolitikájuknak megfelelően.

A forgalom korlátozásokra válaszképpen a P2P alkalmazások forgalmi jellege változik időről-időre a legdinamikusabban, és nehéz lépést tartani az újabb és újabb trükkökkel, amikkel ezek az alkalmazások igyekszenek eltűnni a szokványos forgalmak között, hogy továbbra is minél szélesebb kör, minél többféle médián férhessen hozzá az általuk nyújtott szolgáltatásokhoz, tartalmakhoz.

A forgalom menedzsmentben a hálózati forgalmak csoportosítása – az alkalmazásokhoz rendelhető portszámok, a csomagok tartalma vagy statisztikai jellemzőik alapján – már jó ideje alkalmazott megoldások, de az újabb és újabb alkalmazások felismerése a régebben bevált módszerekkel sok esetben nem lehetséges, emiatt újabb megoldások után kell kutatni.

Az osztályozó algoritmusoknál a valós idejűség és a helyes működés megvalósítása a kihívás. Néhány forgalom típus felismerése eléggé nyilvánvaló – pl. DNS lekérdezések – de napjainkra már a hálózati forgalmak nagy részét titkosított csatornák, nem szabványosított kódolású, protokollú forgalmak alkotják, ezek helyes felismerése jóval nehezebb feladat. [19, 20]

2 Forgalom osztályozási eljárások

Forgalom osztályozáshoz a hálózati entitások közötti forgalmakat adatfolyamok (streamek) szintjén érdemes kezelni. Egy-egy adatfolyam olyan IP csomagok összessége, aminek azonosítására egy ötelemű adatcsoportot definiálhatunk:

- forrás és cél IP cím,
- forrás és cél port szám,
- protokoll.

Ezek alapján a vizsgálati ponton áthaladó adatfolyamok protokollvizsgálat szempontjából adott TCP vagy UDP kliensek között továbbított adatok összességét jelentik. Az alkalmazások azonosítása szempontjából a fent definiált adatfolyamokból legalább egy, de legtöbbször több különálló szál tartozik egy-egy alkalmazás forgalmai közé, ezeket az összetartozásokat a vizsgálat során fel kell ismerni és megfelelően kezelni.

Fontos megjegyezni, hogy az alkalmazásokhoz rendelhető forgalom mindig kétirányú. Az alkalmazás jellegétől függően a forgalom lehet például szimmetrikus és aszimmetrikus. A vizsgálatok pontosságát jelentősen befolyásolja, ha nincs hozzáférés a kapcsolatok két irányához, mivel azok jellege együttesen hordoz fontos, azonosításhoz szükséges információkat. Gerinchálózati szinten gyakran előfordulhat olyan útválasztási eset, hogy az oda-vissza irány nem egy nyomvonalon valósul meg, emiatt általában az ilyen monitorozó eszközöket a hálózat hozzáférési pontjaihoz érdemes telepíteni.

2.1 Port alapú azonosítás

IP alapú hálózatokban a felhasználók forgalmainak többsége TCP és UDP szegmensekben kerül továbbításra, mely protokollok fejlécükben portszámokat használnak az alkalmazási rétegbeli végpontjaik azonosítására. Egy vagy több portszám egy irányban egy alkalmazáshoz rendelhető, értékük 0 és 65535 között lehet.

Az internet hőskorában az IANA (Internet Assigned Numbers Authority [2]) által rögzített és kötelező érvényűen használt, úgynevezett „jól ismert” (well-known) portszámok alapján a forgalmak osztályzása annak idején könnyen megvalósítható

volt. A klasszikusnak mondható forgalmak egy részének (pl. SMTP, POP3, HTTP, FTP, Telnet) azonosítására jelenleg is alkalmazható, bár egyre kisebb mértékben.

Előnyként jegyezhetjük meg, hogy valós idejű megvalósítása korábban sem ütközött problémába, hiszen csupán a szállítási rétegbeli fejléc információk kinyerésére és az ott található portszám azonosítására volt szükség.

Napjainkra viszont az Internetes kommunikációt használó alkalmazás típusok száma jelentősen megnőtt, az általuk használt vagy használható portszámok nem kerültek rögzítésre, így számos alkalmazás esetében a portszám használat teljesen véletlenszerű. Még ha azonos alapértelmezett portszámot is használ egy program, általában szabadon megváltoztatható annak értéke a felhasználó részéről. Sokféle ismeretlen eredetű forgalom található meg a szokásos portokon is, (pl tűzfalon átjutás céljából), így a forgalmak jelentős részének felismerésére más módszerek alkalmazása szükséges.

2.2 Csomagtartalmon alapuló azonosítás

A forgalmak azonosításának egyik legkézenfekvőbb megoldása, ha a csomagtartalmakat vizsgáljuk meg, más nevén mély csomagvizsgálatnak vetjük alá az adatgrammokat (Deep Packet Inspection - DPI) [3, 21]. Ez sok esetben megbízható eredményt szolgáltat, mivel az alkalmazás neve, vagy valamilyen azonosítója, egyedi jellemzője gyakran szerepel egy-egy adatfolyam csomagjaiban, de pl. az egyre nagyobb számú titkosított forgalom esetén ez a lehetőség sem működik. Másik problémája ennek a módszernek, hogy a begyűjtött adatok feldolgozása legtöbb esetben rendkívül számításigényes, így sokszor a valós idejű működés kivitelezhetetlen.

Problémája még az eljárásnak, hogy a kiválasztott minta mennyire specifikus: ha túl általános jellemző alapján osztályzunk, túl sok hibásan osztályzott forgalmat kapunk eredményül, míg specifikusabb minta esetén a nem felismert alkalmazások száma lesz magas.

A mély csomagvizsgálat jellemzően nyílt forráskódú protokollok esetén működik elvárt pontossággal, mivel ebben az esetben könnyű egyértelmű azonosítókat találni a továbbított csomagokban. Általános esetben meglehetősen sok valódi forgalom vizsgálata alapján lehet csak kellően jó mintát találni, ráadásul ezeket a mintákat folyamatosan ellenőrizni is szükséges, mivel pl. verzió váltások esetén lehet, hogy épp az azonosítást jelentő adatokat, mezőket érinti változás.

A felhasználói forgalmak titkosításának rohamos terjedése a mély csomagvizsgálat eszközeit, lehetőségeit érzékenyen érinti, mivel ezeken a forgalmakon nem alkalmazhatóak ezen eljárások.

2.2.1 Minta alapú forgalomosztályozás

A minta alapú forgalomosztályozás esetén a megvizsgált forgalomban speciális bájt mintákat keresünk, melyeket az egyes alkalmazástípusokhoz előre definiáltunk. Például egy web forgalom esetén "GET" vagy „POST" mintát. Ezzel a módszerrel jelentősen csökkenthetjük az eltárolni és feldolgozni szükséges adatok mennyiségét, hiszen a példában szereplő mintát is megtaláljuk a http protokoll fejlécében, a felhasználói adatokhoz nem is szükséges hozzáférnünk.

2.2.2 Dinamikus minta alapú forgalom osztályozás

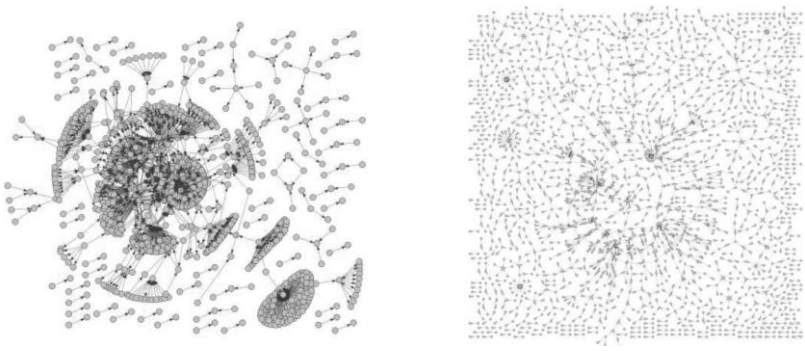
Az előzőleg bemutatott módszerrel a forgalom vizsgálata előre definiált, nem változó minták után kutatott a csomagokban. A módszer dinamikus változata azt a lehetőséget ragadja meg, hogy számos protokollban adott mezőkben azonos típusú, de különböző értékek szerepelhetnek, vagyis a minták, amiket az algoritmus keres nem fixek, hanem az adott forgalom állapotától függően különbözőek, pl. időbélyeg értékek, sorozatszámok stb. Ezeket tipikusan reguláris kifejezésekkel tudjuk megkeresni a csomagokban.

2.2.3 Személyes adatok védelmi kérdései

A mély csomagvizsgálat alkalmazása sok ország adatvédelmi szabályozásai alapján számos kérdést is felvet, mivel a csomagok tartalmának és már akár a csomag fejlécének kezelése és feldolgozása is olyan személyes információnak számíthat, ami az adott felhasználó azonosítására is alkalmas lehet. Adott esetben maga az IP cím is azonosíthat egy-egy személyt. Hogy maga a technika ilyen helyzetben is alkalmazható legyen, a problémás fejléc mezőket anonimizálni szükséges, pl. úgynevezett egyirányú titkosítási művelettel (pl. MD5 [4]). Magát a csomag által szállított információt viszont nem lehet ilyen módon leképezni, így ez a vizsgálati módszer olyan környezetben, ahol a felhasználók személyes adatainak védelme alapvető fontosságú, nem alkalmazható.

2.3 Kapcsolati mintákon alapuló eljárások

A kapcsolati mintákon alapuló eljárások az egyes végpontok közötti kapcsolatok feltérképezésén alapul. Sok alkalmazás esetében jellemző kapcsolati térképről beszélhetünk: pl. email szolgáltatás esetén tipikusan sok kliens kapcsolódik viszonylag kevés számú szerverhez, vagy ennek ellentétére példa a P2P alkalmazások, ahol a kapcsolatok szerkezete teljesen decentralizált.



a) Email traffic b) P2P traffic

1. ábra.
Kapcsolati minták [5]

A [6,7] munkákban bemutatott rendszer volt az első ezen a területen, mely meglehetősen nagyszámú alkalmazást tudott íly módon azonosítani. A végpontok közötti kapcsolatok feltérképezésén túl, a cél- és forrás IP címek és portok alapján a kommunikációban résztvevő végpontok szerepét is megállapította.

Ez a módszer rezisztens a forgalmak titkosításával szemben, mivel csak olyan fejléc információkat használ, amelyeket nem érint a titkosítás. Mint a fenti ábra is mutatja ennek a módszernek a leghasználhatóbb és leglátványosabb kimenete a kapcsolatok grafikus ábrázolása. Ezeken az ábrákon a láthatóan dedikált szereppel rendelkező végpontok alkalmazásai felmérhetők, eltárolhatók és a későbbiekben felhasználhatók közvetlen azonosításra.

Vigyázni kell azonban, mert hálózati topológia változások esetén jelentős változások léphetnek fel a kapcsolati ábrán, mely befolyásolhatja a helyes döntést. Továbbá jellemzően egy-egy szerver több alkalmazást is futtat, így a jellemző minták keveredése, szintén hibákat okozhat.

Az eljárás másik fontos tulajdonsága, hogy egy-egy folyamra egyáltalán nem alkalmazható a módszer, csak az adott hálózati szegmens összes forgalmának vizsgálatára és jellemzésére.

Az eljárásnak meglehetősen hosszú időre van szüksége ahhoz, hogy a kellő alaposságú feltérképezést elvégezze. Emiatt a számítási és tárolási igénye meglehetősen nagy, főként ha magasabb szintű gráf metrikákat akarunk a begyűjtött adatokra alkalmazni.

2.4 Mérhető paramétereken alapuló forgalom osztályozás

Ez a módszer a csomagok felhasználói adattartalmát nem érinti, mivel csak az adatfolyamok statisztikai jellemzői alapján kerülnek az egyes alkalmazások azonosításra. Legtöbb esetben titkosított adatfolyamokon is alkalmazható. Az adatfolyamok a monitorozott linken az őket azonosító öt adat alapján külön-külön kerülnek vizsgálatra. E mellett adategységenként időbélyegek eltárolására is szükség van az időbeli statisztikai jellemzőinek számításához. A vizsgálat a protokoll stack különböző rétegein végezhető.

A korábbi eljárásokkal összehasonlítva a statisztikai mértékeken alapuló vizsgálatok kevesebb tárolóhely igénnyel rendelkeznek, és a szükséges számítási erőforrások is sokkal jobban skálázhatók. A forgalmi adatok lementésével pedig off-line kiértékelésre is lehetőség adódik.

2.4.1 Mérhető jellemzők

A folyamok azonosításán túlmenően a csomagok fejlécéből a csomag mérete is kinyerhető. A csomaghoz rendelt időbélyeg segítségével, és a különböző protokoll rétegbeli információk megfelelő feldolgozásával számos adatfolyam jellemző paraméter számolható. Hálózati szinten, az IP csomagokra, a csomagméret, az érkezési idő és időközök, szállítási szinten az adatfolyamok mérete, időbeli hossza, adatfolyamok érkezési időközei stb. számíthatók.

A [8] forrás szerzői ezek alapján 248 különböző jellemzőt gyűjtöttek össze, melyekből a forgalmak azonosításához 22-t véltek a legfontosabbnak.

- Előre irányú paraméterek csomagokra:
  - méret: min, max, átlag, szórás
  - érkezési időközök (inter-arrival times): min, max, átlag, szórás
- Vissza irányú paraméterek csomagokra:
  - méret: min, max, átlag, szórás
  - érkezési időközök (inter-arrival times): min, max, átlag, szórás
- Folyamparaméterek
  - Adatfolyam időbeli hossza
  - Előre irányban: csomagok száma, bájtok száma
  - Vissza irányban: csomagok száma, bájtok száma
- Protokoll

[9]-ben a szerzők felülvizsgálva a paramétereket, és azok használatával a számítható jellemzőket a fenti 22 paraméterből 7-et hagytak meg, melyek használatával az előzőekhez képest nem tapasztaltak jelentős pontatlanságot:

- Protokoll
- Metrikák az előre irányú csomagokra: min, max, átlag, szórás
- Csomagszám előre irányban egy adatfolyamban

- Vissza irányú adatfolyamban a minimális csomagméret

2.5 Gépi tanuláson alapuló módszerek

A statisztikákon alapuló módszerek esetén jól alkalmazhatóak a gépi tanulás eljárásai (ML – machine learning), mellyel a forgalmak automatikus csoportosítása válik lehetővé a kiszemelt hálózati paraméterek alapján. Például figyelhetjük a csomagok, vagy adatfolyamok méreteit és beérkezési időközeiket. Az eljárások ezen paraméterek alapján a hasonló jellemzőkkel bíró forgalmakat egy csoportba fogják sorolni.

A gépi tanulási módszerek két csoportja ismert:

- Felügyelet nélküli [10, 21], más nevén klaszterezés. Ezek az algoritmusok az előre kiválasztott paraméterek szerint csoportokba, klaszterekbe, rendezik a forgalmakat. A klasztereket előre nem definiáljuk, azok automatikusan alakulnak ki a feldolgozott paraméterek alapján. Ez az eljárás főként nem ismert forgalom típusok esetén alkalmazható.
- Felügyelt: [11,12,13,21] az egyes klasztereket a vizsgálat előtt definiáljuk, melyeket mesterségesen generált, „betanító forgalmak" segítségével határozunk meg. Ez után az algoritmus a forgalmakat ezekhez az előre meghatározott csoportokhoz fogja rendelni. Hátránya e módszernek, hogy az időközben megjelenő új típusú forgalmakat a rendszer a már definiált csoportokhoz fogja rendelni helytelenül.

Vizsgált jellemzők számának csökkentése fontos lépés az algoritmus valós idejű alkalmazásának esetében. Sok esetben felesleges az összes elérhető jellemző alapján alakítani a klasztereket, érdemes a kialakítás jellemzőit jobban meghatározó paramétereket kiválasztani. Ezzel az eljárás számítási kapacitás igénye is kézben tartható. Pl. adatfolyam alapú eljárások esetén sok esetben előfordul, hogy csak azok a csomagok kerülnek vizsgálat alá, melyek valóban tartalmaznak felhasználói adatokat. Így például TCP vezérlő szegmensek (SYN, ACK) kikerülhetnek a vizsgálati halmazból.

A jellemzők csökkentésére természetesen automatizált lehetőség is adódik: ebben az esetben egy alkalmas algoritmus feladata, hogy a meghatározott klaszterek kiválasztásában nyomós szerepet játszó paramétereket megkeresse. [14]

A fentiekkel ellentétben sok esetben találkozhatunk a manuális beavatkozással is. [15]-ben egy valós idejű eljárás kerül bemutatásra, mely TCP folyamok azonosítására alkalmas és felügyelet nélküli klaszteringet használ. Az eljárás a TCP folyamok csupán első néhány csomagjának méretét használja az azonosításhoz. Az ötlet azon alapul, hogy számos alkalmazás esetében jellemző, hogy a kapcsolatfelvétel elején közel egyforma üzeneteket továbbítanak, melyek kellően egyediek lehetnek az alkalmazások megkülönböztetéséhez.

Ez az algoritmus az alábbi feltételek mellett működik:

- Mind a két irány forgalmát vizsgálni kell.
- Az azonosítási fázis előtt az algoritmust be kell tanítani mesterséges minták alapján az alkalmazások felismerésére.
- Az algoritmus valós idejű működésekor a fejlécek feldolgozásakor a csomagméreteknek elérhetőnek kell lenni.

[16]-ban egy TCP és UDP folyamokat azonosító eljárás található. A metódus előnye, hogy a folyam kezdete után az azonosítást minél hamarabb igyekszik befejezni, jellemzően még jóval az adatfolyam vége előtt. Az eljárás számításigénye igen alacsony, így valós időben meglehetősen gyorsan tud eredménnyel szolgálni, továbbá feltételezi, hogy mind a két irányhoz hozzáfér.

Az ilyen, viselkedés jellegű, vizsgálatok esetén a kiválasztott metrikáknak az alábbi feltételeknek kell teljesülni:

- a metrikáknak kellően átfogónak kell lenni, hogy az adott forgalom típus semmilyen fontos jellemzője ne maradjon ki,
- a metrikának általánosíthatónak kell lenni, azaz a vizsgált viselkedésnek nem egy adott csomagra kell igaznak lenni, hanem az adafolyamban lévő csomagok teljes halmazára,
- a valós idejű implementálás miatt a számításigénynek alacsonyan kell maradni.

2.6 Forgalmi jellemzők mérései

Forgalmi jellemzők méréseinél egy olyan vizsgálati környezetet feltételezünk, ami egy internet szolgáltató hozzáférési hálózatához hasonlít. A cél, hogy a felhasználók kétirányú aggregált forgalmának vizsgálatával képet kapjunk a használt alkalmazások megoszlásáról úgy, hogy a linken haladó forgalmat csupán passzívan monitorozhatjuk. A monitorozó készüléket közvetlenül egy aggregációs pont után kell elhelyezni, rendszerint annak egyik olyan kimenetére csatlakoztatható, amire a vizsgált linken haladó kétirányú forgalom tükrözve van. Így a forgalom figyelése az adott szegmens összes felhasználóját érinti.

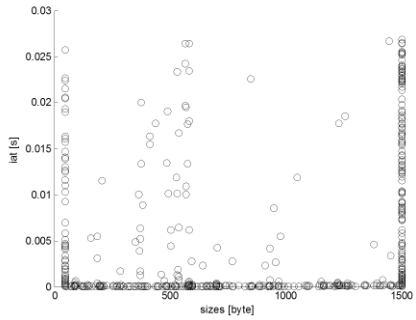
A monitorozó eszköz a vizsgált linket passzívan figyeli. A monitorozó alkalmazás feladata, hogy regisztrálja az áthaladó csomagokat mindkét irányban: elmentse az érkezési idejüket, a cél és forrás IP címüket és portszámaikat, a csomagok méretét, és lehetőség szerint a használt szállítási, vagy felsőbb rétegbeli protokollját is – technikailag az áthaladó csomagok első 100-200 bájtját. Kisebb forgalmú linkeken erre alkalmas szoftver lehet a WireShark, vagy tcpdump program is. A vizsgáló algoritmus ezután a csomagokat az irányok (fel- és letöltési irány) az IP címek és a portszámok szerint különválogatja. Ezzel rendelkezésünkre állnak a vizsgálathoz szükséges szeparált adatfolyamok, melyekhez az őket generáló alkalmazások

típusait, vagy a konkrét alkalmazásokat kell majd rendelni. A vizsgálatokhoz az eddigi gyakorlat szerint kb. 2-3 perc hosszú minták kerültek rögzítésre.

A szeparált adatfolyamokat ez után meg kell vizsgálni, hogy elvégezhető-e rajtuk az IANA által rögzített portszámok szerinti alkalmazáshoz rendelés. Így jelentős számítási igény felszabadítható, mivel a statisztikai vizsgálatoknak nem kell alávetni a sikeresen azonosított forgalmi mintákat. Jelen mérések célja azonban nem egy teljes osztályozó–azonosító eljárás bemutatása, hanem a statisztikai vizsgálat eredményességének alátámasztása, így jelen esetben az összes regisztrált forgalom vizsgálata megtörténik statisztikailag is.

A különválasztott folyamok statisztikai vizsgálata következik tahát, mely a csomagméret, és a csomagok beérkezése között eltelt idő (IAT – inter-arrival time) alapján történik. A két adat segítségével egy kétdimenziós síkban – ahol a vízszintes tengelyen a csomagméretet a függőlegesen a csomagközi időt ábrázoljuk – elhelyezhető egy pont, ami a csomag két paraméterét egyszerre reprezentálja. A regisztrált forgalmi mintában minden csomaghoz rendelhető így két szám, minek eredményeként képezhető egy ponthalmaz a kétdimenziós síkban. Erre mutat egy példát a 2. ábra, ahol egy letöltés irányú normális böngésző forgalom csomagjainak paramétereit lehet látni.

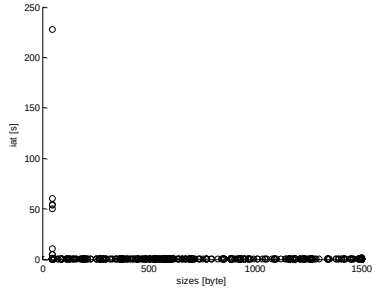
Az ábrán látható, hogy a forgalomban nagyon sok nagyméretű (1500 bájt) csomag van jelen, változó csomagközi idővel. Ezen kívül jellemző kis mérettel (kb. 40 bájt) találunk még sok csomagot, melyek a forgalomban található nyugtázó csomagok lesznek. Ezen túlmenően kis követési időközzel számos köztes méretű csomag is jelen van a forgalomban.



2. ábra.
Böngésző forgalom, letöltés

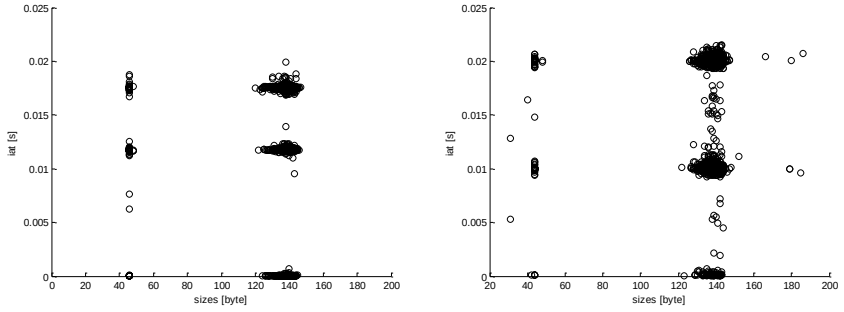
A fenti ábra egy úgynevezett csonkolt ábra, mivel a szélsőséges értékek itt nem kerültek ábrázolásra. Az összes regisztrált csomag ábrázolási képét a néhány extrém paraméterrel (általában nagy IAT) rendelkező csomag nagymértékben eltorzítja, mely a forgalmi minta olyan szakaszán keletkezik mikor tényleges forgalom nincs a linken. Ez látható a lenti ábrán. Ezeket a szakaszokat vagy a regisztrált

forgalomból kell kiszűrni a vizsgálat előtt, vagy itt a konstellációs ábra előállításakor kell megfelelően kezelni. Jelen bemutatott eredmények esetében a szélsőséges értékek a konstellációs ábra megrajzolása előtt kerültek eldobásra, oly módon, hogy az IAT értékek szerint sorba állított adatok 10 %-a került kiszűrésre. Lehetőség még egy konkrét időhatár meghúzása is, de ezt az alkalmazások jellegéhez igazítani kell, ezzel a későbbiekben lehet a rendszer kiegészítve.



3. ábra.
Böngésző forgalom csonkolás nélkül

A következő ábrákon a mérési adatok 90%-a alapján előállított konstellációs ábrák láthatók, interaktív valósidejű hangátvitel forgalmi minták alapján Skype használatával.



4. ábra.
Skype fel- és letöltés irányú forgalmak

A Skype forgalom jellemzői meglehetősen tipikusak, mellyel többen is foglalkoztak már [17,18]. Jellemzően kétféle csomagméretet használ, elsősorban vezérlési információk továbbítására a kisebb, 40 bájt körüli csomagokat, míg a hasznos adat (digitalizált hang) átvitelére pedig 120-140 bájt körüli méreteket, A csomagok követési idői 0.01 és 0.02 másodperc körül csomósodnak. (Az ebből számított adatsebesség kb. 52 és 104 kbit/s-ra adódik).

3 Összegzés

Jelen cikk egy összefogalalót adott az adathálózatokon alkalmazható forgalomosztályozási technikákról, melyekkel mindenki találkozhat, vagy megtapasztalhat akár felhasználói, akár szolgáltatói oldalról. A klasszikus technikák napjainkban már csak részben alkalmazhatók, míg az újabbak bizonytalanságaikkal okozhatnak meglepetéseket. Számos külső tényező is beleszól a forgalom elemzésbe, mint pl. a személyes adatok védelme, mely adott esetben számos megvalósítási lehetőséget zár ki az eszköztárból. A folyamatos kutatásoknak, fejlesztéseknek köszönhetően a forgalom felismerési technikák egyre kifinomultabbak, de a biztos működésnek alapvető feltétele, hogy az ismert módszerek vegyesen kerüljenek alkalmazásra úgy, hogy a szükséges erőforrások gazdaságossága megmaradjon.

References

[1] Pál Varga, Gergely Kún, Gábor Sey, István Moldován, and Péter Gelencsér: "Correlating User Perception and Measurable Network Properties: Experimenting with QoE", *MMNS/MANWEEK2006, 9th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services*, 2006, Dublin, Ireland

[2] Internet Assigned Numbers Authority (IANA), http://www.iana.org/

[3] S. Sen, J. Wang, "Analyzing peer-to-peer traffic across large networks", *Proc. Second Annual ACM Internet Measurement Workshop*, 2002, Franciaország

[4] R. Rivest, "The MD5 Message-Digest Algorithm", *IETF, RFC1321*, 1992 április, http://www.ietf.org/rfc/rfc1321.txt

[5] M. Iliofotou, P. Pappu, M. Faloutsos, M. Mitzenmacher, S. Singh, G. Varghese: Network Traffic Analysis using Traffic Dispersion Graphs (TDGs): *Techniques and Hardware Implementation, Technical Report*, 2007, http://www.cs.ucr.edu/~marios/Papers/UCR-CS-2007-05001.pdf

[6] T. Karagiannis, A. Broido, M. Faloutsos, and K. Claffy, Transport Layer Identification of P2P Traffic, *Proc. IMC (Taormina, Sicily, Italy)*, October 2004.

[7] T. Karagiannis, et al., "BLINC: Multilevel traffic classification in the dark", *ACM SIGCOMM*, 2005.

[8] A. W. Moore and D. Zuev, "Internet Traffic Classification Using Bayesian Analysis Techniques", *Proc. SIGMETRICS (Banff, Alberta, Canada)*, June 2005

[9] N. Williams, S. Zander, and G. Armitage. "A preliminary performance comparison of five machine learning algorithms for practical ip traffic flow classification", *SIGCOMM Comput. Commun*. Rev. 36 (2006), no. 5, 5–16.

[10] J. Erman, A. Mahanti, and M. F. Arlitt, "Internet traffic identification using machine learning", *GLOBECOM*, 2006.

[11] A. W. Moore and D. Zuev, "Internet Traffic Classification Using Bayesian Analysis Techniques", *Proc. SIGMETRICS* (Banff, Alberta, Canada), June 2005

[12] A. McGregor, M. Hall, P. Lorier, and A. Brunskill, "Flow Clustering Using Machine Learning Techniques", *Proc. PAM* (Antibes Juan-les-Pins, France), April 2004.

[13] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamatian, "Traffic Classification On The Fly", *SIGCOMM Comput. Commun*. Rev. 36 (2006), no. 2, 23–26.

[14] F. Tan, "Improving feature selection techniques for machine learning", *Ph.D. thesis*, Atlanta, GA, USA, 2007, Adviser-Bourgeois, Anu G.

[15] L. Bernaille, et al., "Traffic Classification On The Fly", *ACM SIGCOMM Computer Communication Review*, Volume 36, Number 2, April 2006

[16] J. Cao et al. "Online Identification of Applications Using Statistical Behavior Analysis", *IEEE "GLOBECOM"* 2008

[17] Kyoungwon Suh, Daniel R. Figueiredo, Jim Kurose, Don Towsley, "Characterizing and Detecting Skype-Relayed Traffic", *Infocom2006*, Barcelona, Spain, April 2006.

[18] Saikat Guha, Neil Daswani, Ravi Jain, "An Experimental Study of the Skype Peer-to-Peer VoIP System*", IPTPS2006*, June 2006.

[19] Gergely Kún: „IP traffic anaysis", 30th Kando Conference, 2014.

[20] Li, Fangfan, et al. „Classifiers Unclassified: An Efficient Approach to Revealing IP Traffic Classification Rules" *Proceedings of the 2016 ACM on Internet Measurement Conference*. ACM, 2016.

[21] Ertam, Fatih, and Engin Avci. „Classification with Intelligent Systems for Internet Traffic in Enterprise Networks." *Int. J. Comput. Commun. Instrum. Eng.(IJCCIE)* 3.1 (2016).