УДК 621.931

В.І. Романчук, С.В. Алексєєв,
В.В. Червенець, Р.С. Колодій
Національний університет "Львівська політехніка"

АНАЛІЗ СТРУКТУРИ МЕРЕЖЕВОГО ТРАФІКУ ТА МЕРЕЖЕВИХ АНОМАЛІЙ НА ПРИКЛАДІ СЕГМЕНТА ЛОКАЛЬНОЇ МЕРЕЖІ КАМПУСУ НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ "ЛЬВІВСЬКА ПОЛІТЕХНІКА"

Новітні мережі підтримують тенденцію щодо підвищення інтеграції різних типів мережевої інфраструктури і технологій, формуючи так звану концепцію "Інтернету речей". Це вимагає створення нових підходів до забезпечення безпеки і цілісності мереж, боротьби з мережевими аномаліями. Використання методів "точкового аналізу" трафіку і формування базових принципів, необхідних для створення так званих "розумних мереж", допомагає наблизитися до відповіді на ці виклики.

Ключові слова: аналіз трафіку, мережеві аномалії, безпека мереж.

V.I. Romanchuk, S.V. Aleksieiev,
V.V. Chervenets, R.S. Kolodiy
Lviv Polytechnic National University

ANALYSIS OF NETWORK TRAFFIC STRUCTURE AND NETWORK ANOMALY DETECTION IN THE LOCAL SEGMENT OF LVIV POLYTECHNIC NATIONAL UNIVERSITY CAMPUS NETWORK

The main objectives of this article are to analyze and propose new solution to increase network efficiency and security. New approaches to the network designs are emerging with changing of network services and network availability requirements. To create network environment that can fully support ability to freely interconnect between various devices and inside different network types (BYOD (Bring Your Own Device), Internet of Things, etc.), the new mechanism of traffic analysis and control must be developed. When discussing basic principles of this new approach, we need to take in consideration all the major challenges that this "freely interconnected" concepts are facing. The main problem will be network security and integrity. Also significantly will increase percentage of "parasite" network traffic and network anomalies occurrence. But with high traffic rates transmitted over network segments and mobility of network users that are using wireless network infrastructure, current approaches Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), and other network and security monitoring approaches just not flexible enough and inefficient. The future of network security and traffic control solutions lies within "smart network" concept. New types of "smart networks" need to have an ability to react autonomously on any "threat" or anomaly in network flow. This requires an algorithm that can in real time make a correct suggestion regarding current network problems. Basically the whole process will include three stages: monitoring (finding and locating certain network problem and defining its "pattern"), classification (using "pattern" that was established on previous stage classification algorithm can now find suggested source of the problem and mechanism of solving it), resolving problem (finally we can apply solution that algorithm choose to solve the problem). All this stages must be combined in one procedure that is executed by autonomous controlling unit. For decreasing

an amount of false positives and information that needs to be proceeded, this control mechanism need to include problem detection solution that will use "critical point" method of information gathering (only collect traces and traffic patterns of certain minimum that needed to identify problem), and machine learning (to dynamically improve quality of selected solutions).

Key words: traffic analysis, network anomalies, network security.

Вступ. В процесі розвитку сучасних інформаційних мереж все гостріше постає проблема забезпечення їх надійності, збіжності та ефективності. Одним з найсерйозніших викликів нині стає поширення та поглиблення таких концепцій, як BYOD, та поступовий перехід від концепції Internet of Devices до нової концепції Internet of Things, враховуючи прагнення розробників цих концепцій до підвищення рівня їх доступності для користувача, а також зростання вимог до динамічної маштабованості мереж, що відрізняються за масштабами, типом чи функціональністю (WAN, MAN, LAN, PAN, BAN, etc.). Саме наявність різноманітних типів трафіку та динамічне середовище, максимально орієнтоване на надання послуг доступу до мережі, формують ряд інженерно-технічних викликів. При цьому мережа також має містити елементи корпоративного середовища і вимоги щодо забезпечення певного рівня надійності й конфіденційності даних. Звичайно, в реальних умовах важко дослідити такі інноваційні підходи до побудови мереж, але мережу кампусу Національного університету "Львівська політехніка" певною мірою можна представити як приклад квазікорпоративної мережі. Саме тому вивчення структури й особливостей трафіку в одному з її віртуальних локальних сегментів (VLAN) може стати основою для розроблення підходів і алгоритмів автоматизації збіжності, безпеки й ефективного функціонування новітніх прототипів програмованих мультисервісних мереж з високим рівнем доступу.

Аналіз та оцінка структури трафіку сегмента локальної мережі кампусу. В досліджуваному локальному сегменті є близько 250 кінцевих пристроїв користувачів, частина з них підключена як стаціонарні ПК, частина через портативні (мобільні) пристрої (нетбуки, лептопи тощо). Певний відсоток також підключений до мережі опосередковано, за допомогою самостійно встановлених комутаторів чи маршрутизаторів та з використанням різноманітних типів технологій передачі та обміну даними. Ключовим фактором є наявність певної хаотичності у підключенні користувачів і пов'язані з цим труднощі з ефективним контролем і забезпеченням якості роботи мережі. Тому, зібравши дані щодо потоку трафіку і динаміки його структури у певні моменти часу (підвищеної активності користувачів чи відсутності активного навантаження), можна класифікувати певні патерни, які можуть слугувати індикаторами рівня "здоров'я" мережі, що, своєю чергою, дозволяє розробити механізми зворотної дії для компенсації негативних тенденцій. На відміну від активних систем моніторингу, які вже реалізовані, потрібні гнучкіші засоби для контролю потоку трафіку, засоби, які спільно із системами моніторингу можуть застосовувати точковий аналіз мережі, своєрідну реакцію на виникнення тих чи інших "подразників" та порушень штатного режиму роботи мережі. Для початку розглянемо схему збереження дампів трафіку в досліджуваному сегменті мережі (рис. 1).

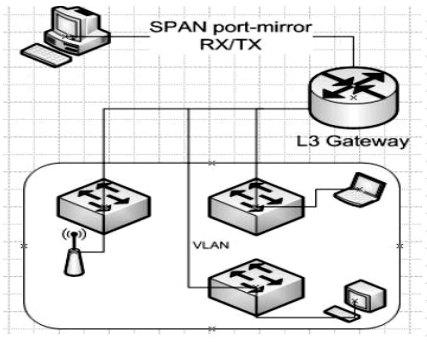


Рис. 1. Схема підключення SPAN порту

Забір трафіку на аналіз відбувається на магістральній лінії, що підімкнена до інтерфейсу комутатора 3 рівня (див. рис. 1), що слугує шлюзом за замовчуванням. Згідно з оцінкою активності мережевих інтерфейсів, зібраною за допомогою систем моніторингу, фази високої активності мережі кампусу припадають на періоди з 10^{00} до 13^{00} та з 14^{00} до 16^{00}. Для проведення дослідження трафіку створено чотири вибіркові дампи трафіку: в фази високої мережевої активності, а також для порівняння контрольний у фазу пониженої мережевої активності (див. табл. 1).

Примітка: трафік зібрано на порту, налаштованому у SPAN RX/TX режимі, під час запису трафік розділено на окремі файли розміром 52 MB, загальний об'єм однієї сесії запису даних 1 GB. Нижче наведена статистика для кожної сесії запису відповідно до окремих файлів.

Аналіз трафіку під час підвищеної завантаженості мережі

Таблиця 1

Загальна статистика потоку трафіку в період більшої активності мережі (*відсотки трафіку вказані від загального об'єму трафіку, зібраного в конкретному файлі, також надана кількість пакетів)

Файл №:	1	2	3	4	5
Загальна статистика:					
TCP	75 651 (78,4 %)	65 903 (70,6 %)	71 590 (78,4 %)	65 390 (88,4 %)	64 638 (87,3 %)
UDP	18 869 (19,6 %)	26 310 (28,2 %)	18 611 (20,4 %)	8 095 (10,9 %)	7 834 (10,6 %)
L 2-3	1 981 (2 %)	1131 (1,2 %)	1125 (1,2 %)	505 (0,7 %)	1569 (2,1 %)
Файл №:	11	12	13	14	15
Загальна статистика:					
TCP	42 232 (68,2 %)	40 533 (64,4%)	46 212 (58,7%)	40 533 (64,4%)	59 500 (88,3%)
UDP	19 592 (31,4 %)	25 155 (35,1 %)	22 053 (39,7 %)	31 214 (39,7 %)	6 859 (10,2%)
L 2-3	280 (0,5 %)	310 (0,5 %)	633 (1,0 %)	1325 (1,7 %)	1054 (1,6 %)
Файл №:	6	7	8	9	10
Загальна статистика:					
TCP	62 641 (88,5 %)	57 370 (90,7 %)	61 027 (84,5 %)	44 747 (63,4 %)	32 585 (45,9 %)
UDP	7 279 (10,3 %)	5 045 (8,0 %)	10 159 (14,1 %)	21 697 (30,7%)	37 905 (53,4 %)
L 2-3	826 (1,2 %)	833 (1,3 %)	1076 (1,5 %)	4159 (5,9 %)	493 (0,7 %)
Файл №:	16	17	18	19	20
Загальна статистика:					
TCP	65 678 (83,0%)	67 198 (84,4%)	64 514 (81,7%)	63 276 (85,6%)	51 595 (64,3%)
UDP	12 088 (15,3 %)	11 781 (14,8 %)	13 542 (17,1 %)	9 776 (13,2 %)	27 793 (34,4 %)
L 2-3	1406 (1,8 %)	607 (0,8 %)	918 (1,2 %)	861 (1,2 %)	1047 (1,3 %)

Як можна побачити з табл. 1, загальний розподіл співвідношення трафіку між рівнем 4 стека протоколів TCP/UDP та протоколами 3 та 2 рівнів залишається доволі пропорційним. Та щоб краще зрозуміти особливості розподілу трафіку по специфічних протоколах, розглянемо окремі випадки для максимальних значень активності наведених вище видів трафіку, а також у випадку суттєвої зміни пропорції наявних протоколів у потоці. Як видно з табл. 1, найбільший відсоток активності за наведеними типами мережевих протоколів зафіксовано у 5, 7, і 10 файлах дампу трафіку. У цей момент зафіксовано пікову активність протоколів 2 і 3 рівня, TCP і UDP відповідно, а отже, саме вони найбільше підходять для своєрідного "зрізу" потоків трафіку і вивчення активності конкретних протоколів, що і формують цю картину.

Аналіз трафіку під час зниженої завантаженості мережі. Наявність трафіку, що використовує протоколи IPv6, є негативним фактором, оскільки офіційно в мережі немає жодних сервісів чи пристроїв, що працюють за цим протоколом. Якщо розглянути протокол TCP, то найбільший відсоток з чітко розпізнаних типів протоколів займає протокол HTTP. Якщо ж розглядати трафік за протоколом UDP, то найбільше зафіксовано саме передавання даних, хоча насторожує постійна активність протоколу Toredo IPv6 over UDP. Саме вивчення такого "зрізу" мережі у різні періоди її активності чи в період певних збоїв може допомогти сформувати певні залежності. Як бачимо, під час активного використання мережі найбільшу активність виявляє протокол TCP, що в принципі логічно, але наявність таких факторів, як нестандартні протоколи, деформовані пакети, високий рівень повторного пересилання пакетів чи велика кількість їх "дублікатів" вказує на проблеми з ефективністю, надійністю та безпекою мережі. Для порівняння розглянемо трафік мережі під час пасивнішого її використання та зміну процентного співвідношення наявних мережевих протоколів. Як можна побачити з табл. 3, у випадку з пасивним використанням мережі дещо змінилося співвідношення між мережевими протоколами.

Таблиця 2

Загальна статистика потоку трафіку в період меншої активності мережі

Файл №:	1	2	3	4	5
Загальна статистика:					
TCP	21 649 (31,7 %)	32 269 (47,2 %)	26 767 (44,2 %)	30 505 (45,3 %)	24 622(38,5 %)
UDP	45 915 (67,3 %)	35 616 (52,1 %)	24 393 (40,3 %)	28 956 (43,0 %)	7 970 (13,3 %)
L 2-3	692 (0,1 %)	518 (0,8 %)	9 386 (15,5 %)	7 819 (11,6 %)	15 229 (23,8 %)
Файл №:	11	12	13	14	15
Загальна статистика:					
TCP	45 843 (61,2 %)	32 291 (42,7 %)	46 628 (62,9 %)	37 178 (50,2 %)	36 873 (48,6 %)
UDP	28 026 (37,4 %)	24 961 (33,0 %)	24 687 (33,3 %)	28 243 (38,2 %)	32 900 (43,4 %)
L 2-3	1081 (1,4 %)	18 399 (24,3 %)	2 782 (3,8 %)	8 606 (11,6 %)	6 052 (8,0 %)
Файл №:	6	7	8	9	10
Загальна статистика:					
TCP	29 638 (45,9 %)	27 784 (43,6 %)	28 858 (44,2 %)	37 546 (59,7 %)	48 986 (45,9 %)
UDP	29 201 (45,2 %)	34 442 (54,0 %)	35 906 (55,0 %)	24 862 (39,6 %)	20 277 (29,0 %)
L 2-3	5 751 (8,9 %)	1551 (2,4 %)	534 (0,8 %)	452 (0,7 %)	639 (0,9 %)
Файл №:	16	17	18	19	20
Загальна статистика:					
TCP	35 580 (49,9 %)	49 300 (80,4%)	20 665 (32,0 %)	17 131 (26,5 %)	29 130 (45,7 %)
UDP	28 261 (39,6 %)	11 346 (18,5 %)	43 168 (66,9 %)	47 145 (72,9 %)	34 022 (53,4 %)
L 2-3	7 467 (10,5 %)	679 (1,1 %)	685 (1,1 %)	425 (0,7 %)	553 (0,8 %)

Як і в попередньому випадку, виберемо три файли дампу з найбільшою активністю за відповідними типами протоколів: 12 (L 2-3 24,3 %), 17 (TCP 80,4 %), та 19 (UDP 72,9 %). Як бачимо, пропорційне співвідношення протоколів стало різкішим, а також на фоні зменшення відсотка корисного завантаження мережі можна чіткіше прослідкувати активність наявного паразитного трафіку та мережевих аномалій.

Виявлення мережевих аномалій на основі отриманих даних. З наведених даних видно, що певна частина трафіку в цьому локальному сегменті мережі є паразитною, а отже, становить певні загрози як для мережі загалом, так і для безпеки і комфорту роботи у ній користувачів, тому що так званий паразитний трафік ідеально підходить для маскування різноманітних атак та вірусної активності в мережі. Це, своєю чергою, призводить до дискредитації певних користувачів мережі чи навіть певних мережевих сервісів загалом. Якщо спробувати виконати певну класифікацію мережевих аномалій, то їх можна поділити на дві основні категорії: явні та неявні.

Явні – це наявність таких активних протоколів, як IPv6, Quake 3 Arena Protocol, Teredo, Sinec H1 Protocol, GSM over IP, etc.. Неявні, своєю чергою, пов’язані з деформованими пакетами, надмірним повторним передаванням пакетів у межах однієї сесії, а також наявністю активних сесій до загальновідомих недовірених або ж і відверто скомпрометованих IP адрес. Глибше вивчивши трафік, можна побачити наявність запитів до неавторизованих DNS серверів та іншу схожу негативну активність. Для пошуку і аналізу мережевих аномалій можна застосувати техніку аналізу, що базується на вивченні залежностей і ентропії трафіку. Для цього дослідження використаємо методику, запропоновану в дослідженні [3]. Враховуючи вищенаведену статистику, виберемо три файли з періоду пасивного використання мережі (12,17,19) і проведемо оцінку трафіку кінцевого користувача з найбільшою кількістю активних сесій відносно зміни показника ентропії трафіку. Для початку проведемо аналіз і визначимо найактивнішого кінцевого користувача для IPv4 трафіку. Один з користувачів займає в 12 файлі друге місце, а в 17 і 19 третє за об’ємом обміну пакетами. На основі співвідношення кількості пакетів, що відправляються до конкретного сокета, та загальної кількості пакетів за досліджуваний період ми отримаємо значення ймовірності для кожної унікальної сесії протягом досліджуваного періоду:

$$p(x[i]) = \frac{U(x[i])}{Z_{sum}}, \tag{1}$$

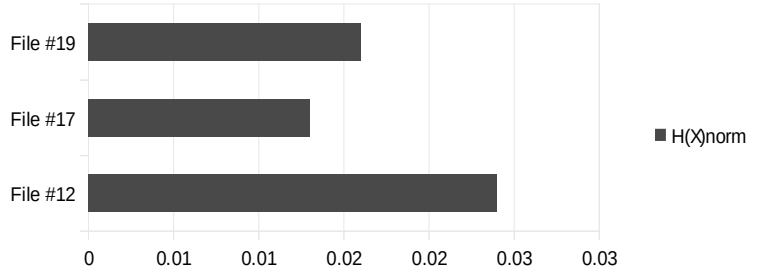
де $U(x[i])$ – кількість пакетів $x[i]$ з певною адресою призначення, Z_{sum} – загальна кількість пакетів.

Маючи значення ймовірності, виберемо фактор нормалізації $\log(N[0])$, де $N[0]$ – це кількість унікальних адрес призначення. Тепер проведемо розрахунок ентропії:

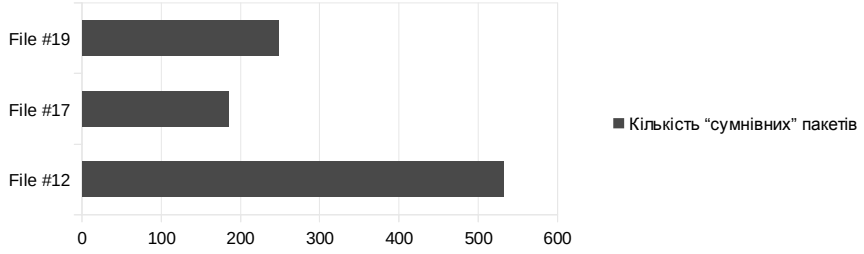
$$H(x) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i), \tag{2}$$

де $H(x)$ – ентропія випадкової величини x, n – кількість унікальних адрес призначення, $p(x_i)$ – ймовірність, що x набуде значення x_i, значення b для логарифма дорівнює 2. Розрахунок нормалізованої ентропії (значення в межах від 0 до 1) здійснимо за співвідношенням $H / \log(N[0])$.

Маючи значення ентропії, можемо порівняти його з кількістю “сумнівних” пакетів, наявних для цих сесій у кожному з трьох випадків.



а



б

Рис. 2. Порівняння показників нормалізованої ентропії (а) і кількості “сумнівних” пакетів (повторної передачі, деформованих тощо) (б)

Висновки. У разі високого рівня свободи підключення до мережі й використання різноманітних технологій побудови мереж потрібен розвиток нових технологій і підходів “точкового аналізу”, які могли б в автоматизованому режимі в сукупності з системами моніторингу і конфігурації мережі активно вливати на роботу і функціонування мережі.

Аналіз аномального трафіку, а також застосування методу визначення ентропії для аналізу трафіку може виявитися одним з ефективних інструментів побудови “патернів”.

Таблиця 3

Детальний аналіз структури трафіку 12,17,19 файлів (пасивний період роботи мережі)

Файл №:	12	17	19
Загальна статистика:			
TCP	32 291 (42,7 %)	49 300 (80,4 %)	17 131 (26,5 %)
UDP	24 961 (33,0 %)	11 346 (18,5 %)	47 145 (72,9 %)
L 2-3	18 399 (24,3 %)	679 (1,1 %)	425 (0,7 %)
	=	=	=
IPv4	75 221 (99,49 %)	61 208 (99,81 %)	64 520 (99,72 %)
TCP	32 291 (42,71 %)	49 300 (80,39 %)	17 131 (26,48 %)
HTTP	3 844 (5,08 %)	5 375 (9,35 %)	5 336 (8,25 %)
SSL	6 174 (8,17 %)	3 987 (6,50 %)	732 (1,13 %)
NetBIOS	49 (0,06 %	=	=
BitTorrent	126 (0,17 %)	71 (0,12 %)	70 (0,11 %)
Data	1 395 (1,85 %)	336 (0,55 %)	1 082 (1,72 %)
Sinec H1 Protocol	=	=	1 (0,00%)
PPPTP	38 (0,05 %)	4 (0,01 %)	14 (0,02 %)
Gnutella	=	=	=
POP	=	=	=
FTP	=	=	=
FTP Data	=	=	=
DCE RPC	16 (0,02 %)	=	=
Web Socket	=	=	=
XMPP Protocol	4 (0,01 %)	=	=
UDP	24 741 (28,71 %)	11 304 (18,43 %)	47 075 (72,76 %)
Data	21 923 (29,00 %)	10 334 (16,85 %)	45 834 (70,84 %)
Domain Name Service	290 (0,38 %)	44 (0,07 %)	85 (0,13 %)
Teredo IPv6 over UDP	1 702 (2,25 %)	670 (1,09 %)	906 (1,40 %)
NetBIOS NS	666 (0,88 %)	189 (0,31 %)	172 (0,27 %)
Dropbox LAN DP	16 (0,02 %)	=	=
GPRS NS	=	=	1 (0,00 %)
NetBIOS DS	10 (0,01 %)	6 (0,01 %)	=
HTTP (SSCM)	10 (0,01 %)	=	=
Bootstrap Protocol	6 (0,01 %)	9 (0,01 %)	9 (0,01 %)
ICMP	39 (0,05 %)	12 (0,02 %)	9 (0,01 %)
Quake 3 Arena Protocol	2 (0,00 %)	2 (0,00 %)	=
NAT Port Mapping	=	=	=
IPv6	158 (0,21 %)	36 (0,06 %)	74 (0,11 %)
UDP	138 (0,18 %)	36 (0,06 %)	62 (0,10 %)
HTTP (SSCM)	46 (0,06 %)	18 (0,03 %)	22 (0,03 %)
DHCPv6	36 (0,05 %)	14 (0,02 %	28 (0,04 %)
DNS	56 (0,07 %)	4 (0,01 %)	12 (0,02 %)
ICMPv6	20 (0,03 %)	=	12 (0,02 %)
TCP	=	=	=
LLC	128 (0,17 %)	56 (0,09 %)	72 (0,11 %)
ARP	99 (0,13 %)	31 (0,05 %)	35 (0,05 %)
GRE (PPP)	18 149 (27,45 %)	591 (0,96 %)	305 (0,47 %)
Data	1 (0,01)	1 (0,00 %)	=

Необхідно розробити нову систему класифікації надійності та довіреності мережевої активності та протоколів, яка могла б, використовуючи відповідне маркування активних сесій, потоків і протоколів, регулювати і систематизувати їх активність.

Запропоновані підходи чудово вписуються в концепцію так званих автоматизованих (програмно контрольованих мереж), які мають набути здатності до самоорганізації, відновлення і контролю незалежно від технологій, що використовуються для передавання інформації у мережі.

1. Wireshark, www.wireshark.org 2. Wireshark network analysis (Second edition), Laura Chappell, Founder of Wireshark University 3. An empirical evaluation of entropy-based traffic anomaly detection, George Nychis, Vyas Sekar, David G. Andersen, Hyong Kim, Hui Zhang; Carnegie Mellon University. 4. Karamcheti V., Geiger D., Kedem Z., and Muthukrishnan S. Detecting malicious network traffic using inverse distributions of packet contents. In Proc. of ACM SIGCOMM Workshop on Mining Network Data (MineNet), 2005 5. The TCP/IP Guide, Charles M. Kozierok, 2005.