УДК 004.031.2+004.424]:621.391 DOI: 10.31673/2412-4338.2024.048351

Sahaidak V.A.
State university of information and communication technologies, Kyiv
ORCID 0009-0000-9724-958X

OVERVIEW OF TELECOMMUNICATIONS NETWORK DATA COLLECTION METHODS BY PROBE

Annotation: *In this article overview was provided on two systems for traffic collection in real time. These systems are Gigamon fabric solution and Huawei NetProbe. Gigamon fabric solution performs information collection/filtering/enrichment. Following solution supports SS7, IP, 3G, LTE. Information from network elements can be collected by network TAPs (splitters) or traffic port mirroring. Data processing stages with description of each function and used devices description were provided. Huawei NetProbe is used for decoding and creation of call detailed records (CDRs), storing raw signaling data and real time session tracing. This system can recognize more than 1300 protocols and applications by self-developed by Huawei Service Awareness Engine. NetProbe supports traffic collection from NGN, GSM (CS and PS), UMTS (CS and PS), LTE, IMS. DPI (Deep protocol inspection) base, used by Probe to recognize data, is released periodically to support latest protocols. In case if protocol is missing in DPI, it can be configured. System workflow was described. Detailed overview was provided on secondary devices used by both systems and risks were described during usage on telecommunications network. Example of network was provided and analyzed changes that should be done for reviewed data collection method, depending on network elements change, update or new site deployment. Following conclusions were made: Real time data collection provides a lot of pros, but also requires careful investigation before implementation of one or another method. Port mirroring can duplicate information without additional hardware installation, but can have significant impact on network element performance. TAPs can duplicate traffic with cost of signal power on both ends. Active TAP doesn't have such issue, but power outage on such device can cause service loss; Every network change like hardware update, vendor change or new site deployment should be done with thought that data collection should be also updated. Nether less, device configuration should be changed in order to clean duplicates in messages. It increases cost and time for operation and maintenance; Data collection by Probe can be excessive due to realization, required processing resources and amount of collected information, which most probably will never be used. In such case, if only one customer from company requires it, it is better to look for cheaper and easy to use sources of data.*

Key words: *GTP, CS network, PS network, information technology, VNF, LTE, IMS, SS7, VoIP, IP, Big data.*

Сагайдак Віктор Анатолійович
Державний університет інформаційно-комунікаційних технологій, м. Київ
ORCID 0009-0000-9724-958X

ОГЛЯД МЕТОДІВ ЗБОРУ ДАНИХ НА МЕРЕЖІ ТЕЛЕКОМУНІКАЦІЙ ЗА ДОПОМОГОЮ МЕРЕЖЕВОГО ЗОНДУ

Анотація: *В цій статі було наведено опис двох систем для збору інформації у режимі реального часу, а саме Gigamon fabric solution та Huawei NetProbe.Gigamon fabric solution виконує функцію збору, фільтрування, доповнення інформації. Дана система підтримує SS7, IP, 3G, LTE. Інформація з мережевих елементів збирається за допомогою мережевих відгалужувачів (сплітерів) або віддзеркалювання трафіку з порту. Наведено основні етапи обробки даних з описом кожної функції та допоміжні пристрої. Huawei NetProbe*

© Сагайдак В.А. *2024*

використовується для декодування, створення записів сенсу зв'язку та збереження необробленої сигналізації. Система підтримує обробку інформації з мереж NGN, GSM (канальної та пакетної комутації), UMTS (канальної та пакетної комутації), LTE, IMS. За допомогою DPI, що постійно оновлюється та конфігурується, система підтримає більше ніж 1300 протоколів та дозволяє додати свої. Був наведений принцип роботи система. Були детально розглянуті допоміжні пристрої обох систем та наведено ризики при їх використанні на мережі телекомунікацій. Було проаналізовано та наведено приклад мережі для якої дані методи збору інформації потрібно корегувати в залежності від зміни елементів, оновленню або розширенню самої мережі. Були зроблені наступні висновки: збір даних з мережі у режимі реального часу надає багато переваг, але також потребує ретельної підготовки перед наступною інтеграцією одного чи іншого методу. Зеркалювання трафіку з портів може копіювати дані з мережі, але може мати значний вплив на продуктивність мережевих елементів. Мережеві відгалужувачі можуть дублювати трафік за рахунок зменшення рівня сигналу на обох кінцях. Активні мережеві відгалужувачі не мають такої властивості, але для своєї роботи потребують джерело електроенергії для їх роботи, але у випадку перебоїв з живленням можуть спричинити проблеми у роботі мережевого обладнання; При кожній зміні на мережі як оновлення апаратного забезпечення, заміна виробника мережевого обладнання або створення нової площадки для обладнання потрібно мати на увазі, що збір трафіку теж повинен бути оновлений. Тим паче, конфігурація повинна бути змінена, щоб видалити дуплікацію у даних. Це збільшує час та вартість супроводу та оновлення; Збір даних за допомогою мережевого зонду може бути надмірним завдяки реалізації, великого використання ресурсів та обсягу інформації, що скоріш за все ніколи не буде використана. У такому випадку, якщо один замовник потребує в компанії такий метод, то краще пошукати дешевший та простіше у реалізації джерело інформації.

Ключові слова: *GTP, мережа з канальною комутацією каналів, мережа з пакетною комутацією, інформаційні технології, VNF, LTE, IMS, SS7, VoIP, IP, Великі дані.*

1. Introduction

Data, type of data, it's structure, amount of it always matters during analysis. Information can be collected for one research purpose and later on reused for another if it`s purpose changed. In Big data term sometimes companies collect all types of data no matter if it is useful or not and if it could be used in future.

For carrier network one of the most desired way to collect and analyze data from elements is network Probe. This method allows to check quality of provided services and apps that mostly used by users. Probe provides real time information collection, it`s enrichment and divide it in smaller parts and delete duplicates. But like any other data collection it has pros and cons.

2. Analysis of Probe capabilities and collection devices

In this section two probing systems will be reviewed - Gigamon fabric solution and Huawei NetProbe. Gigamon solution can be used as standalone system integrated with other third-party analytics software, while NetProbe can be integrated with Huawei analytics systems.

Gigamon fabric solution (Figure 1) features real-time data collection by network taps installed between VNF equipment like servers, leaf and "spine" switches, routers [1]. This probing solution can also collect data by port mirroring.

Gigamon fabric is controlled by GigaVUE Fabric Management (FM). Following management software provides a centralized management platform for GigaVUE nodes and clusters, Traffic Filtering, reporting dashboard for the whole fabric, Application Intelligence to monitor and identify application usage for filtering [7].

Traffic Filtering is a core intelligence consisting of Flow Mapping for flow extraction policies definition, GigaStream load balancing for distribution of traffic across tools, Terabit-scale configurations, Inline Bypass features for threat prevention tools and enforcement points, visibility across different infrastructure types (physical, virtual, cloud) [8]. Flow Mapping is a feature located in GigaVUE nodes, that support line-rate traffic at 1Gb, 10Gb, 40Gb, or 100Gb from physical or virtual SPAN/port mirroring, network TAP. Following feature processes data from a set of user-defined mapping rules to other tools and applications. Flow Mapping can filter traffic based on Layer 2-4 parameters like IP address, application port number, VLAN ID, MAC, etc. It also allows user to define custom rules and apply them to specialized application, tunneled traffic, high layer protocols. Flow Mapping can be used with GigaSMART for additional actions like packet modification, load balancing, flow masking de-duplication, packet slicing to reduce size of information for processing and storing, header operations like stripping and addition, de-tunneling of ERSPAN, IP, GRE, L2GRE, GMIP, VXLAN to provide data, that analysis tools couldn't handle before [9]. Following GigaSMART operation features belong to Traffic Intelligence [10].

GigaSMART provide other operations on solutions like Subscriber Intelligence and Application Intelligence [11].

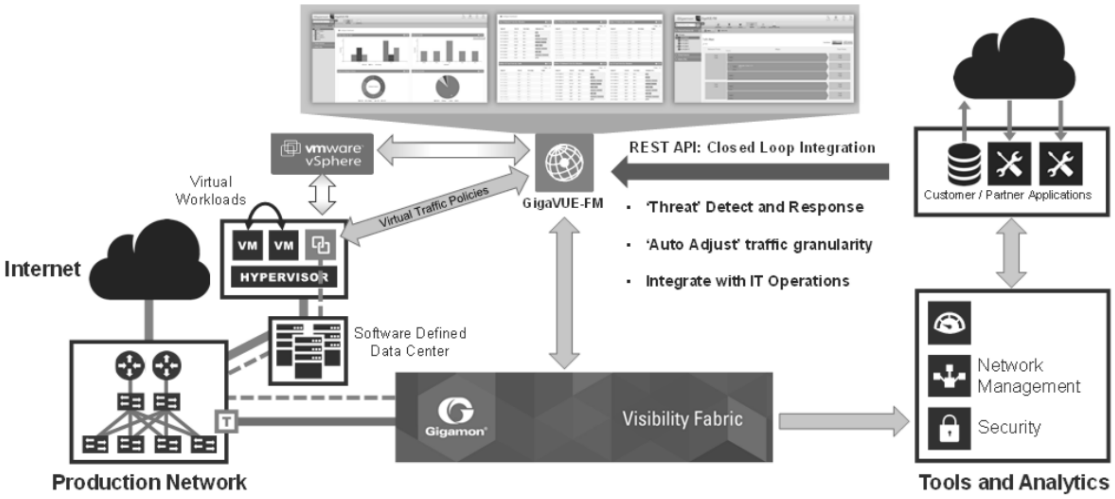

Figure 1 Gigamon fabric scheme

GigaSMART Subscriber Intelligence application provides correlation of GTP, SIP/RTP, Diameter S6a and FlowVUE to perform different operations on GPRS Tunneling Protocol (GTP) [11,12]. This subset of features correlates and filters data based on subscriber IDs (IMSI, IMEI, and MSISDN), GTP version or EPC interface, GTP-c (control plane) with GTP-u (user plane) traffic, relationship between subscriber ID and tunnel ID [13]. Subscriber Intelligence allows flow sampling and mapping with GTP whitelisting to send selected data to multiple destinations [14]. FlowVUE supports forwarding of filtered sourced based on subscriber device IPs, subscriber IPs, IP ranges or/and at specified sampling rates with configurable timeouts for detection and replacement of inactive devices, sampling of IP-based flows and traffic encapsulated in GTP-u tunnels [12].

GigaSMART Application Intelligence identifies applications in network traffic, isolates flow for specific application to redirect it to analysis tool, exports application metadata to perform actions on its attributes for analysis like relevant usage context, that enable indicators of compromise for forensics and security. It reports total application and bandwidth consumption over a selected period of time in bytes, packets and flows, enables application layer filtering to check high-volume or low-risk traffic [15].

Gigamon fabric solution supports SS7, IP, 3G, LTE, which allows next services processing:

–Mobile Calls—ISUP, TD.35, MAP, SMS, BSSAP, GTP, Diameter, xCDR, etc

–Fixed Calls —ISUP, Diameter, xCDR, etc

–VoIP/VoLTE Calls—SIP, H.323, Diameter, xCDR, etc

Huawei NetProbe is a part of Huawei SmartCare and GENEX Discovery solutions. NetProbe is used for decoding and creation of call detailed records (CDRs), storing raw signaling data and real time session tracing. This system can recognize more than 1300 protocols and applications by Service Awareness Engine self-developed by Huawei. Probe supports traffic collection from NGN, GSM (CS and PS), UMTS (CS and PS), LTE, IMS. DPI (Deep protocol inspection) base is used by Probe to recognize data and it is released periodically to support latest protocols. In case if protocol is missing in DPI, it can be configured [3].

Table 1

Example of services/protocols supported by NetProbe

Service name	**Example of services/protocols supported by NetProbe**
Basic service	WAP1.X/2.0, HTTP/ HTTPS, Facebook, Twitter, Radius, Gaming, Win_Update, etc
Email	SMTP (SSL), POP3 (SSL), IMAP4 (SSL), Webmail, MS_Exchange, LotusNotes, Blackberry, etc
P2P	eDonkey, Bittorrent, FlashGet, Thunder, HotLine, GNUTELLA, DirectConnect, etc
VoIP	Skype Out/In, SIP, Diameter, H323, MGCP, Net2Phone, GoogleTalk, Shutter, UUCall, etc
Streaming	RTP/RTSP, RealPlayer, MS_Media, Flash_Yahoo, PPLIve, YouTube, AOL_Video, etc
IM	MSN, GoogleTalk, YahooMsg, Skype IM, ICQ, Viber, Whatsapp, etc

For PS network Probe works in next way (Figure 2):

1. At first stage, system identifies L3/L4 protocols from data stream with help of Flow Table Match. After that Probe performs L3/L4 Parse.
2. At second stage, NetProbe performs L7/L7 Protocol Identification of information identified at previous stage. At next step system performs parse of L7/L7 protocols and application.
3. At third stage, Probe sends parsed data mediation servers (MES) and pre-processing servers (DPS) of SEQ Analyst or GENEX Discovery.

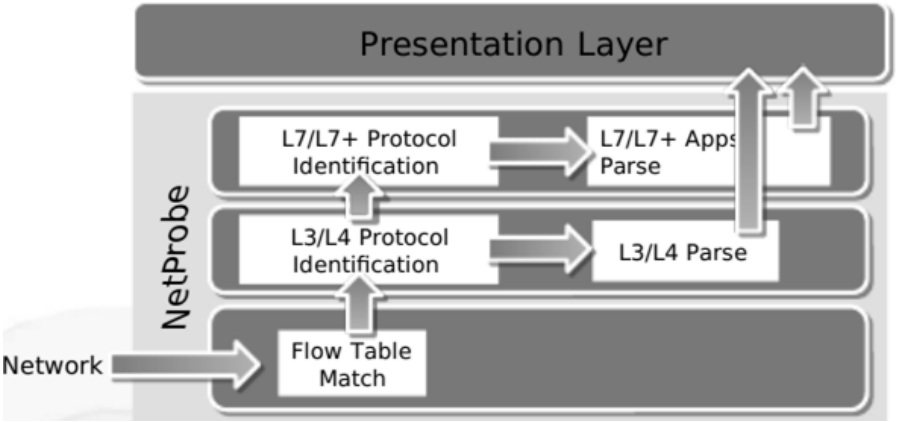

Figure 2 – Probe PS network data processing flow

NetProbe uses splitters or network TAPs for information collection as Gigamon fabric solution. System supports line-traffic at 1Gb, 10Gb, 40Gb or 100Gb (figure 3).

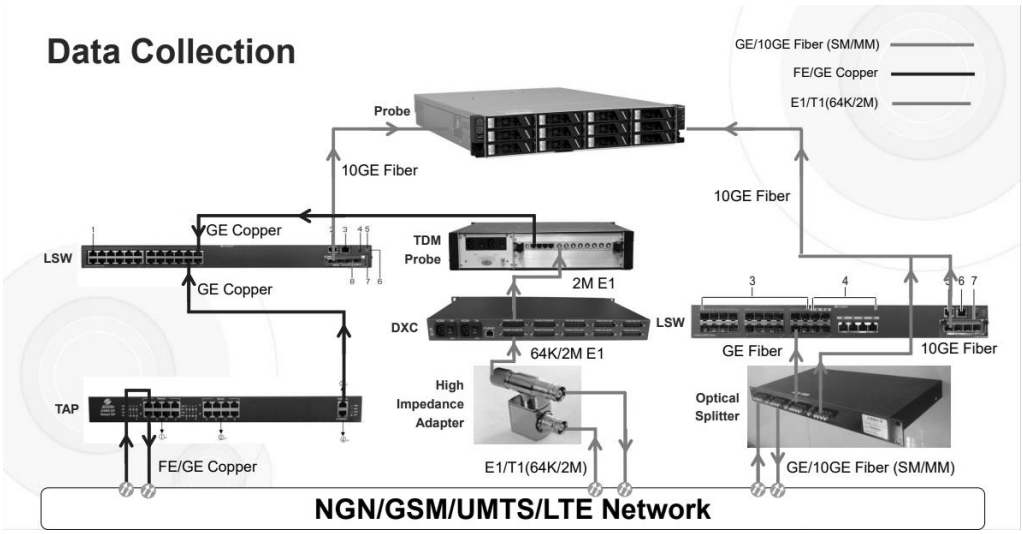

Figure 3 Huawei NetProbe data collection methods

NetProbe network TAPs and splitters are installed between network elements of mobile network (figure 4). As shown in figure 3 these devices are connected to Probe by network switches (LSW). Looks like those following switches are performing network data filtering functions to eliminate duplication, because NetProbe don't have such functions.

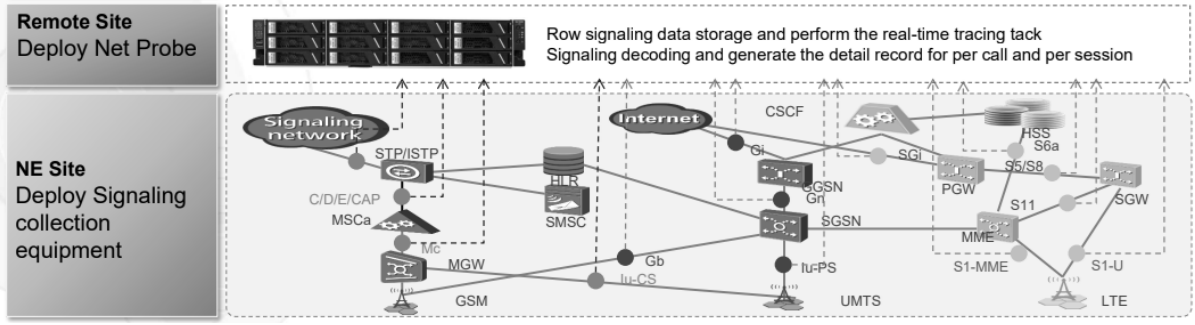

Figure 4 - Example of devices installation for network data collection

3. Purpose of the study

The purpose of following study is to define possible pros and cons for real time data collection and required scope of predefined tasks, that should be performed.

To achieve defined purpose next tasks resolved:

1. Analyze network collection device (TAP and splitter) works approach, what should be checked before implementation.
2. Investigate SPAN/mirror port, it's purpose and possible impact on device utilization during usage.
3. Check in VNF scenario with multiple network entities deployed on different sites what is missing in proposed deployment and check possible impact with suggestions how to eliminate it.

4. Network collection device and mirroring methods review

As can be noticed, for information collection both solutions use traffic mirroring from device ports or network TAP (Test Access Point) also known as splitter.

Splitter is connected directly to the cabling infrastructure in order to split or copy packets. Captured packets are used in analysis, security or general network management [4]. Figure 5 show device working principal.

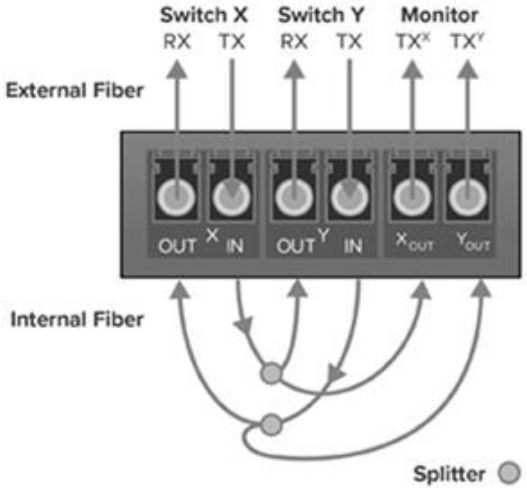

Figure 5 Passive TAP working approach

In first pair of ports installed optical cable RX/TX (receive/transfer) from Switch X and in second pair of ports connected RX/TX Switch Y. Third pair of ports is splitted TX from both devices. Such data collection becomes available due to physical divert from portion of light from its original source, so before device installation, it is recommended to calculate signal power loss.

In table 2 maximum signal level loss are provided for network and monitoring ports, based on split ration.

Table 2

Example of services/protocols supported by NetProbe

Multimode Passive TAPs			
Split ratio	50/50	60/40	70/30
Max network loss	3.9dB	3.15dB	2.2dB
Max monitor loss	3.9dB	5.15dB	6.2dB
Singlemode Passive TAPs			
Split ratio	50/50	60/40	70/30
Max network loss	3.7dB	3.05dB	2.0dB
Max monitor loss	3.7dB	4.95dB	6.1dB

There are 2 type of TAP devices – active and passive. The main difference between active and passive device is that active device can regenerate signal power and requires power source for it's work. It is making such type less reliable compared to passive TAP.

SPAN/mirror port – it is a function, that duplicates from one switch port to another (figure 6).

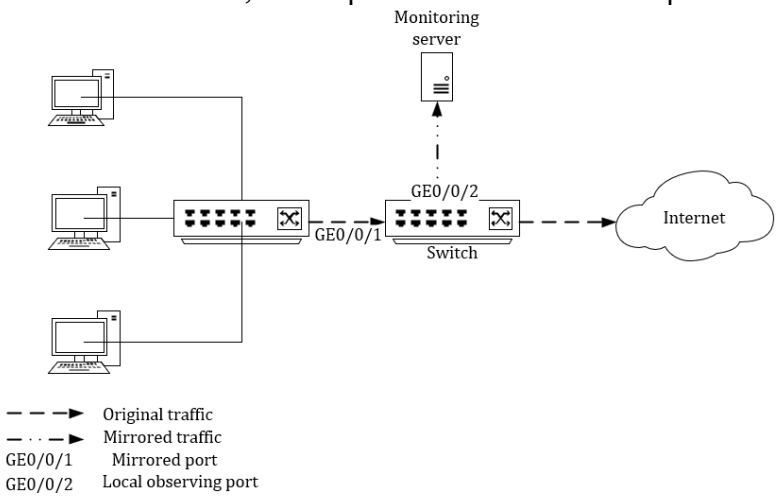

Figure 6 port mirroring scheme

Usually following functionality is used to determine point of failure and troubleshooting. After issue was located, device manufactures recommend to disable it. During long time usage device

resource utilization can drastically increase and can cause impact on other services configured and provided by device. In other words, before configuration available resources should be checked for long term usage [5,6].

5. Possible impact on VNF network with multiple sites

Except reviewed impacts on telecommunication network during implementation of one or another real time collection methods, in reviewed schemes of both solutions all network elements are located within one site. Usually, carriers balance network workload of provided services and perform reservation and disaster recovery between several sites with help of virtualized network functions (VNF). In case, when LTE core network, located for instance in city A, fails to provide service for some reason it's workload and services should be maintained by LTE (or vEPC – virtual Evolved Packet Core) core network located in city B (figure 7).

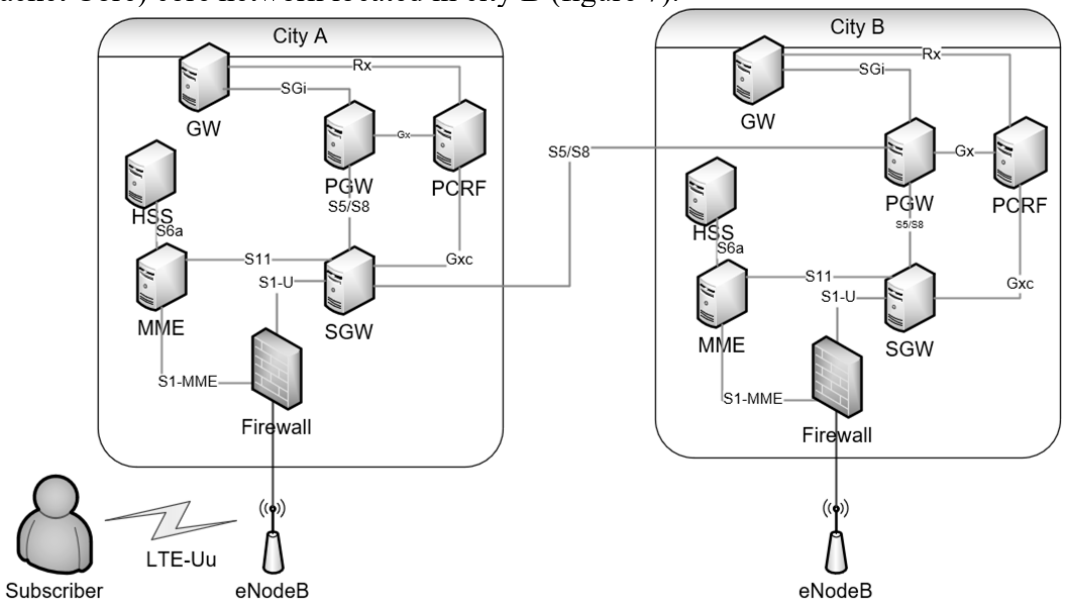

Figure 7 Example of VNF LTE network

In following scenario some points can be discovered:

–Additional network TAPs should be installed in city B, in order to cover services provided by network elements locate there. In case of hardware update, installation of new nodes or even replacement one core network vendor for another, new splitters also should be installed, reconnected or change to mirroring depending of capabilities and specification of new NE;

–Despite new additional collection points, data duplication could appear. For instance, in figure 7 SGW, located in city A, attached to subscriber, but PGW, that provides service to him, could be located in city B. In such case additional measures should be added to prevent duplicate message collection which can 33% of all received records;

–Probe should be installed in city A and city B. If only one network Probe will be collecting data from both cities, it will increase workload on transport network in 2 or even more times, depending on Probe location and how much traffic is served by both sites. Probe license also should be updated, because number of traffic, that Core Network can process increases for 50%;

For calculation of maintenance cost, Total Cost of Operation formula is used [7]:

$$Cost_{operation} = R\left(M_{total} * S_{avg} + IT_{dep} + \sigma_l\right)$$

where R – number of racks with equipment, M_{total} – number of IT support personnel, S_{avg}- average salary of specialist, IT_{dep} – hardware depreciation costs, σ_l – software and licensing costs.

$$\mathrm{M}_{total} = \frac{1}{N_R}$$

where N_R – number of full racks assigned to one specialist.

$$\mathrm{IT}_{dep} = \frac{P_{rprice}}{T_{rlife}}$$

where P_{rprice} – overall hardware purchase cost, T_{rlife} – lifetime of rack with equipment.

$$\sigma_l = \frac{P_{ltotal}}{R}$$

where P_{ltotal} – total software and license cost.

In order to include discovered Probe data collection in VNF scenario, formula was modified:

$$Cost_{operation} = \sum_{i=1}^{R} \left(\mathrm{M}_{total} * \mathrm{S}_{avg} + \mathrm{IT}_{depi} + \sigma_{li} + \sigma_{swli} + \mathrm{IT}_{spli}\right)$$

where R - number of vEPC racks with equipment, M_{total} - number of support personnel with Probe support specialist, σ_{swli} – software licensing cost based on number of network traffic, IT_{spli} – splitter expenses usage.

For Probe support except of IT support personnel, specialist is required, that have knowledge and experience of Core Network, data communication and network probes. In other words, 3 more specialists are required.

$$\mathrm{M}_{total} = \frac{1}{N_R} + \frac{3}{N_{project}}$$

where $N_{project}$ – number of systems assigned to one specialist.

$$\sigma_{swl} = \frac{P_{swtotal}}{T_{monlife}}$$

where $P_{swtotal}$ – Probe license expense, $T_{monlife}$ - lifetime of rack with vEPC equipment.

$$\mathrm{IT}_{spl} = \frac{N_{monp} * P_{sprice}}{N_{splp} * T_{monlife}}$$

where N_{monp} – number of physical ports for monitoring, N_{splp} – number of monitoring ports on splitter device, P_{sprice} – price of one splitter device.

6. Conclusions

1) Real time data collection provides a lot of pros, but also requires careful investigation before implementation of one or another method. Port mirroring can duplicate information without additional hardware installation, but can have significant impact on network element performance. TAPs can duplicate traffic with cost of signal power on both ends. Active TAP doesn't have such issue, but power outage on such device can cause service loss;

2) Every network change like hardware update, vendor change or new site deployment should be done with thought that data collection should be also updated. Nether less, device configuration should be changed in order to clean duplicates in messages. It increases cost and time for operation and maintenance;

3) Data collection by Probe can be excessive due to realization, required processing resources and amount of collected information, which most probably will never be used. In such case, if only one customer from company requires it, it is better to look for cheaper and easy to use sources of data;

Список використаних джерел

1. GigaVUE-FM Overview. GigaVUE 5.8 Online Documentation. URL: http://surl.li/zrdava
2. wangshupeng. SC1002 HUAWEI SmartCare SEQ Analyst & NetProbe Technical Slides V2.4 | PDF | Service Level Agreement | Websites. Scribd. URL: http://surl.li/vvbtbq
3. Understanding international telecoms fraud. Network-Level Intelligence for Observability Tools | Gigamon. URL: http://surl.li/joztlh
4. Example for Configuring Local Port Mirroring (1:1 Mirroring) - S600-E Series Switches Typical Configuration Examples. Huawei. URL: http://surl.li/qfdhia
5. Configuring Local Port Mirroring - CloudEngine S8700 V600R022C00 Configuration Guide - System Monitoring. Huawei. URL: http://surl.li/xfheiu
6. Patel C. D., Shah A. J. Cost Model for Planning, Development and Operation of a Data Center. ResearchGate. URL: http://surl.li/nhuupr
7. GigaVUE Fabric Management. GigaVUE Online Documentation. URL: http://surl.li/opqtwp
8. Traffic Filtering. GigaVUE Online Documentation. URL: http://surl.li/rbmaci
9. Flow Mapping Overview. GigaVUE Online Documentation. URL: http://surl.li/fxnlfd
10. Traffic Intelligence Solutions. GigaVUE Online Documentation. URL: http://surl.li/esfmds
11. GigaSMART Operations. GigaVUE Online Documentation. URL: http://surl.li/jehbmy
12. GigaSMART FlowVUE. GigaVUE Online Documentation. URL: http://surl.li/buemlz
13. GigaSMART GTP Correlation. GigaVUE Online Documentation. URL: http://surl.li/dqzxug
14. GTP Overlap Flow Sampling Maps. GigaVUE Online Documentation. URL: http://surl.li/ieprqs
15. About Application Intelligence. GigaVUE Online Documentation. URL: http://surl.li/wlroki

References

1. GigaVUE-FM Overview. GigaVUE 5.8 Online Documentation. URL: http://surl.li/zrdava
2. wangshupeng. SC1002 HUAWEI SmartCare SEQ Analyst & NetProbe Technical Slides V2.4 | PDF | Service Level Agreement | Websites. Scribd. URL: http://surl.li/vvbtbq
3. Understanding international telecoms fraud. Network-Level Intelligence for Observability Tools | Gigamon. URL: http://surl.li/joztlh
4. Example for Configuring Local Port Mirroring (1:1 Mirroring) - S600-E Series Switches Typical Configuration Examples. Huawei. URL: http://surl.li/qfdhia
5. Configuring Local Port Mirroring - CloudEngine S8700 V600R022C00 Configuration Guide - System Monitoring. Huawei. URL: http://surl.li/xfheiu
6. Patel C. D., Shah A. J. Cost Model for Planning, Development and Operation of a Data Center. ResearchGate. URL: http://surl.li/nhuupr
7. GigaVUE Fabric Management. GigaVUE Online Documentation. URL:

http://surl.li/opqtwp

8. Traffic Filtering. GigaVUE Online Documentation. URL: http://surl.li/rbmaci

9. Flow Mapping Overview. GigaVUE Online Documentation. URL: http://surl.li/fxnlfd

10. Traffic Intelligence Solutions. GigaVUE Online Documentation. URL: http://surl.li/esfmds

11. GigaSMART Operations. GigaVUE Online Documentation. URL: http://surl.li/jehbmy

12. GigaSMART FlowVUE. GigaVUE Online Documentation. URL: http://surl.li/buemlz

13. GigaSMART GTP Correlation. GigaVUE Online Documentation. URL: http://surl.li/dqzxug

14. GTP Overlap Flow Sampling Maps. GigaVUE Online Documentation. URL: http://surl.li/ieprqs

15. About Application Intelligence. GigaVUE Online Documentation. URL: http://surl.li/wlroki