Парето эффектээр халдлага илрүүлэх

П.Минж*, П.Лхавужал†, Ц. Энхтөр‡, Я.Дашдорж§
*Мэдээллийн сүлжээ аюулгүй байдал/Магистр
† Мэдээллийн сүлжээ аюулгүй байдал/Магистрант
‡ Мэдээллийн сүлжээ аюулгүй байдал/Магистр
§ Мэдээллийн сүлжээ аюулгүй байдал/Доктор(Ph.D)
*minj@must.edu.mn †lkhavuu.n@gmail.com ‡enkhtur@sict.edu.mn §dashdorj@must.edu.mn

***Хураангуй*— Энэ судалгааны ажлаар паретогийн эффект ашиглан их хэмжээний пакетын урсгал дээр анализ хийж халдлага илрүүлэхэд оршино. Бид энэ судалгааны ажлаар бодит сүлжээн дээгүүр хэд хэдэн төрлийн халдлага хийж түүний нэгж хугацаан дахь хурд болон пакетын тооны хамаарал, Shannon-ний энтропид суурилсан траффикийн шинжилгээ, тухайн нэгж хугацаан дахь пакетийн хурднаас хоёр суурьтай логарифм авах, паретогийн эффектээр тодорхойлох зэрэг аргуудаар анализ хийж үзсэн. Паретогийн эффект нь маш нарийн тодорхойлох боломжтой болсон.**

Түлхүүр үг*— *траффик, анализ, энтропи, пакет, TCP SYN Flood

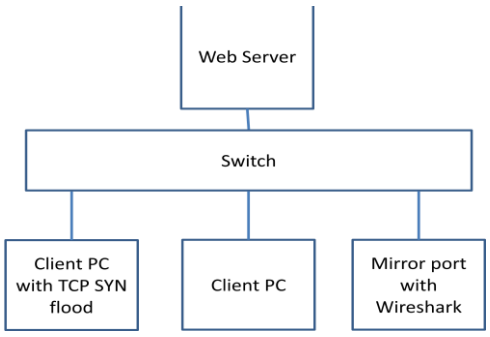
I. УДИРТГАЛ

Сүлжээн дээрх халдлага нь тухайн сүлжээний төхөөрөмжүүд болон серверийн үйл ажиллагааг тасалдуулах, унагах, удаан болгох зэрэг үйлдэл болно. Сүүлийн жилүүдэд сүлжээн дээрх траффикийн хэмжээ эрс нэмэгдэж сүлжээний чанарын үзүүлэлт маш муу болж байгаа. Иймд улс орон бүр өөр өөрийн мэдээллийн аюулгүй байдалд зориулж асар их хөрөнгө оруулалтыг хийж байна. Vantagepoint компанийн мэдээлснээр 2016 оны 10-р сарын 21-ны баасан гаригт тухайн сүлжээний траффик дээр анализ хийж үзэхэд TCP болон UDP протоколын 53-р порт дээр халдлага ажиглагдсан ба DDOS төрлийн халдлага ба түүний хурдны хэмжээ нь 1.2Тбит/с байсан [1]. Касперский компани мэдээлснээр 73 сая вэб хөтөч халдлага хийсэн нь лог файл дотор бичигдсэн. Мөн 3.2 тэрбум халдлага янз бүрийн хэрэглэгч болон мэдрэгчүүдээс орж ирж байгаа нь бүртгэгдсэн байна. IDS (intrusion detection system) нь компьютер болон сүлжээн дээрх пакетын урсгалыг хянах ба аюулгүй байдлыг нь илрүүлдэг[2]. Өдөр бүр шинээр шинэ төрлийн халдлага гарч байгаа үед тухайн байгууллагын сүлжээн дээр аюулгүй байдлын төв байх ёстой ба тэр нь шинэ төрлийн халдлагыг шинээр бүртгэх ба анализ хийж тухайн IDS дээр шинэ дүрмийг нэмж болно. Асар их мэдээлэл дамжиж байх үед түүнийг халдлага эсвэл хуурамч халдлага гэдгийг ялгах нь хамгийн том асуудал болж байна. SNORT нь 1999 DARPA IDS үнэлгээний програм ашиглан туршилт хийхэд дөнгөж 69% нь илрүүлж байсан. Энэ төсөл MIT-ийн LINCOLIN лаборатори дээр хэрэгжсэн. Энэ үр дүнгээс харахад илрүүлэх системийг зайлшгүй сайжруулах хэрэгтэй гэдэг нь харагдаж байна. Сүүлийн үед маш олон төрлийн халдлага илрүүлэх арга хөгжиж байгаа ба үүнд энгийн экспоненциалаар жигнэн дундажлах арга [2], хуурамч халдлагыг мэдэхийн тулд ангилах аргыг мөн хэрэглэж байна. Ангилахдаа пакет бүр дээр шошго тэмдэглэгээ тавих ба энэ нь хэдэн тэрбум пакетын урсгалтай ажиллахад тийм амар биш, мөн дамжуулах чадамжид муугаар нөлөөлж болно. Mine etc нь динамик бүрэлдэхүүн дээр анализ хийх аргаар халдлага илрүүлэх арга судалсан[3]. Бусад судлаачид орж ирсэн пакет дээр статик анализ хийх замаар халдлага илрүүлэх аргыг мөн ашигласан.Ting Liu etc судалгаагаар халдлага илрүүлэхдээ Shannon энтропигийн аргаар анализ хийсэн, мөн Renyi Cross энтропигийн аргыг мөн хэрэглэж анализ хийх улмаар Snort дээр ийм дүрмийг оруулж өгсөн [4]. Бид энэ судалгаагаар бодит сүлжээн дээр TCP SYN FLOOD, UDP FLOOD, ICMP FLOOD зэрэг халдлага хийж түүний траффикийг WireShark хэрэгсэлээр барьж авч, түүний пакетын хурдыг тооцох, анализ хийх, тухайн нэгж хугацаанд дах хурдны хэмжээг хоёр суурьтай логарифмээр илэрхийлэх, Shannon энтропигийн аргаар болон, Паретогийн эффектээр анализ хийж үзсэн.

II. ТРАФФИКИЙН ФАЙЛЫГ ЦУГЛУУЛСАН ТУРШИЛТ

A. Туршилтын зураглал

Туршилтыг 2017 оны 3 сард ШУТИС-ийн МХТС дээр хийсэн ба вэб сервер, клиентүүдээс бүрдсэн дотоод сүлжээ байгуулан нэг клиентээс TCP SYN Flood халдлагыг 2 минутын хугацааны турш хийн, тэр хугацаан дахь траффикийг Wireshark программаар бичиж авсан.



Зураг2. Туршилтын бүдүүвч зураг.

Туршилтын явцад нэг клиент TCP SYN Flood халдлагыг хийсэн ба нөгөө клиент нь ping багцыг мөн явуулж байсан.

Туршилтын хугацаанд баригдсан траффикийн мэдээллийг Хүснэгт 1-т харуулсан. Тус хүснэгтээс харахад 155 секундын хугацаанд 1,151,827 өөр IP хаягаас дамжуулсан 1,152,154 пакет баригдсан ба дундаж пакетийн хэмжээ нь 54.5 байт байгаа нь энгийн үеийнхээс хэд дахин бага байна.

ХҮСНЭГТ 1
ТРАФФИКИЙН СТАТИСТИК ҮЗҮҮЛЭЛТ

Пакет	1152154	Дундаж (B/sec)	399K
Хугацаа (сек)	155.848	IPv4	1151679
Дундаж пакетын тоо	7392.8	Port	1151827
Дундаж пакетын урт (B)	54.5	Нийт байт	62227189

Траффик бичих хугацаанд нөгөө клиент ping хийх явцад эхний ICMP багцууд амжилттай дамжиж байснаа халдлага эхэлснээс хойш хариу ирэх хугацаа огцом нэмэгдэн, цаашлаад амжилтгүй болж байсан. Энэ халдлагад өртсөн тус вэб сервер унан ажиллагаагүй болсныг илэрхийлнэ. Зураг 2-т ICMP ping багцын явцыг харуулсан болно.

```
Reply from 192.168.1.101: bytes=32 time=4ms TTL=64
Reply from 192.168.1.101: bytes=32 time=6ms TTL=64
Reply from 192.168.1.101: bytes=32 time=6ms TTL=64
Reply from 192.168.1.101: bytes=32 time=2ms TTL=64
Reply from 192.168.1.101: bytes=32 time=1794ms TTL=64
Reply from 192.168.1.101: bytes=32 time=15ms TTL=64
Reply from 192.168.1.101: bytes=32 time=28ms TTL=64
Reply from 192.168.1.101: bytes=32 time=158ms TTL=64
Reply from 192.168.1.101: bytes=32 time=1621ms TTL=64
Reply from 192.168.1.101: bytes=32 time=270ms TTL=64
Reply from 192.168.1.101: bytes=32 time=97ms TTL=64
Reply from 192.168.1.101: bytes=32 time=204ms TTL=64
Request timed out.
```

Зураг2. ICMP ping багц.

III. ХАЛДЛАГА ИЛРҮҮЛЭХ АРГУУД

A. Shannon энтропигийн арга

Shannon энтропигийн аргыг мэдээлэл болон тодорхойгүй байдлын хэмжигдэхүүн болгон ашигладаг. X=[x_1, x_2, x_3, … ,x_n] олонлогийн x элемент бүр нь x$\in C_x$ мужид хамаардаг. C_x мужтай X-ийн entropy нь дараах томъёогоор тодорхойлогддог [4] :

$$H(X) = -\sum_{i=1}^{n} p_i log_2 p_i \quad (1)$$

Shannon энтропигийн аргыг ашиглан нийт таван үзүүлэлтээр шинжилгээ хийгддэг. Дамжуулагчийн IP (SNUM) болон хүлээн авагчийн IP (DNUM) хаягийн t хугацааны завсарт entropy нь:

$$H(Sip_t) = -\sum_{i=1}^{m}(snum_i/n)\log(snum_i/n) \quad (2)$$

$$H(Dip_t) = -\sum_{i=1}^{k}(dnum_i/n)\log(dnum_i/n) \quad (3)$$

Дамжуулагч (STHREAT) болон хүлээн авагчийн (DTHREAT) халдлагын зэрэг:

$$H(S\ threat_t) = -\sum_{i=1}^{m}\frac{threat\_of\_sip(i)}{sum\_threat} log_2(\frac{threat\_of\_sip(i)}{sum\_threat}) \quad (4)$$

$$H(D\ threat_t) = -\sum_{i=1}^{k}\frac{threat\_of\_dip(i)}{sum\_threat} log_2(\frac{threat\_of\_dip(i)}{sum\_threat}) \quad (5)$$

Нийт халдлагын дохиог тооцохдоо:

$$sum\_threat = \sum_{i=1}^{n} threat_i \quad (6)$$

Датаграмын уртын энтропи нь:

$$(Dgmlen_t) = -\sum_{i=1}^{x}\left(\frac{dgmNum_i}{n}\right) * \log_2\left(\frac{dgmNum_i}{n}\right) \quad (7)$$

Бид нар өөрсдийн туршилтандаа датаграмын энтропийг тооцон үзсэн.

B. Парето эффектийн арга

Парето тархалтын хувьд x-ийн магадлалыг тооцохдоо:

$$\Pr\{X > x\} = 1 - F(x) \approx \left(\frac{\delta}{x}\right)^{\alpha} \quad (8)$$

Үүнд δ нь 1секунд дэх хамгийн бага утга бөгөөд X нь x→∞ байж болно. α нь shap параметр ба авах утгын муж нь 1→2 байна.

$$F(x) = 1 - \left(\frac{\delta}{x}\right)^{\alpha} \quad (9)$$

Парето тархалтын CDF нь:

$$f(x) = \frac{\alpha}{\delta} \cdot \left(\frac{\delta}{x}\right)^{\alpha+1} \quad (10)$$

Нэг секундэд ирэх дундаж пакетийн тоо нь E(x) бол α параметрийг дараах тэгшитгэлээр тодорхойлно:

$$E(x) = \delta \cdot \frac{\alpha}{\alpha-1} \rightarrow \alpha = \frac{E(x)}{E(x)-\delta} \quad (11)$$

α параметрийн тодорхойлсны дараа Парето эффектийн функцийг тэгшитгэлээр тооцно:

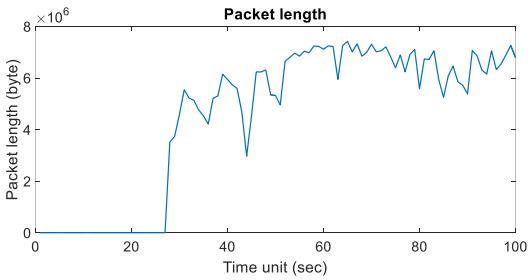
$$batch\_pareto(q,k,\alpha,X) = \begin{cases} 1-q & if\ k=0 \\ \frac{\left[1-\left(\frac{1}{1.5}\right)^{\alpha}\right]}{1+\left(\frac{1}{X+0.5}\right)^{\alpha}} & if\ k=1 \\ \frac{\left(\frac{1}{k-0.5}\right)^{\alpha}-\left(\frac{1}{k+0.5}\right)^{\alpha}}{1-\left(\frac{1}{X+0.5}\right)^{\alpha}} & if\ 1<k<X \\ 0 & if\ k>X \end{cases} \quad (12)$$

Үүнд X нь хамгийн их багцын урт, q – утасгүй төхөөрөмжийн дамжуулах багцын дундаж магадлал, k – нэгж хугацаанд орж ирсэн битийн тоо/пакетийн тоо болно.

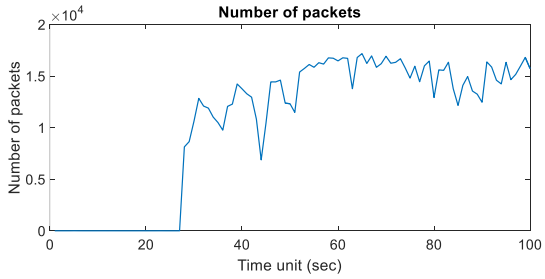
IV. ХАЛДЛАГЫГ SHANNON ЭНТРОПИ БОЛОН ПАРЕТО ЭФФЕКТИЙН АРГААР ҮНЭЛЭХ НЬ

A. Shannon энтропигийн үнэлгээний үр дүн

Wireshark программаар барьж авсан траффикийн мэдээллийг Matlab програм дээр боловсруулсан ба Зураг 3-т нэг секундэд дамжсан урсгалын уртын (байтаар) графикийг харуулсан бол Зураг 4-т нэг секундэд дамжсан пакетын тоог графикаар үзүүлсэн болно.

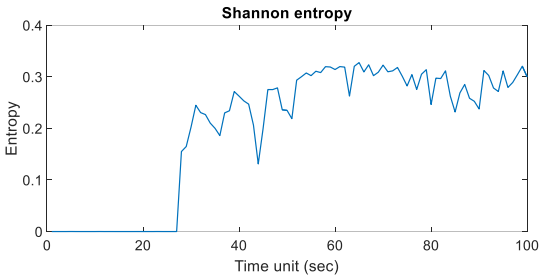


Зураг3. Нэг секундэд дамжсан мэдээллийн урт.



Зураг4 Нэг секундэд дамжсан пакетийн тоо.

Дээрх зурагнуудаас харахад хоорондоо тун төстэй бөгөөд халдлага эхэлснээс хойш огцом нэмэгдсэнийг харж болно.

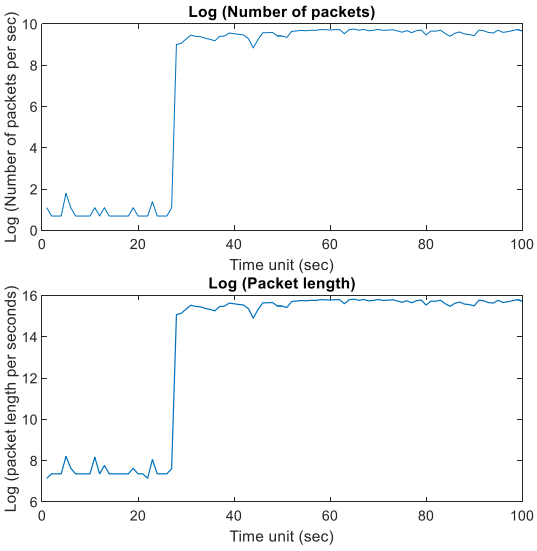


Зураг5. Shannon энтропигийн график.

Үүний дараа нэг секундэд дамжсан пакетийн тоог Shannon энтропигийн аргаар үнэлсэн ба үр дүнг нь Зураг 5-т үзүүлсэн. Графикийг харахад өмнөх хоёр графиктай тун төстэй бөгөөд сүлжээний ачаалал ихэдсэн, эсвэл халдлагад өртсөн эсэхийг шууд дүгнэн хэлэхэд хүндрэлтэй байна.

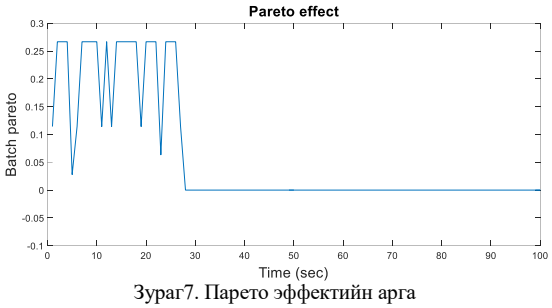
V. ПАРЕТО ЭФФЕКТИЙН АРГЫН ҮР ДҮН

Парето эффектийн аргаар үнэлэхээсээ өмнө бид нэг секундэд дамжсан пакетийн тоо болоод мэдээллийн уртаас логарифм авч график байгуулсныг Зураг 6-аар харуулав.

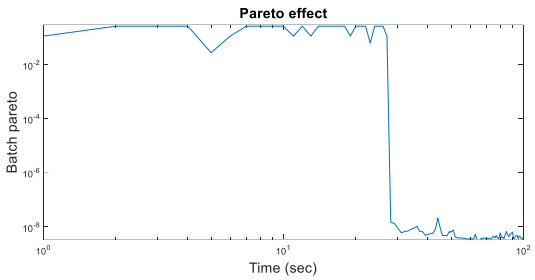


Зураг6. Log(Number of packet) and Log(Packet length)

Дээрх зургаас харахад халдлага илэрсэн эсэхийг Shannon энтропигийн аргаас илүү хялбараар тодорхойлж болохоор байна.



Зураг7. Парето эффектийн арга



Зураг8. Парето эффектийн аргаас логарифм авсан нь.

Зураг 7-д 1 секундэд дамжсан пакетийн тоог Парето эффектийн аргаар үнэлсэн үр дүнг болоод Зураг 8-д түүнээс логарифм авсан графикийг харуулсан болно. Эдгээр графикийг харахад тодорхой хугацааны дараа огцом тэг рүү дөхөж унаснаас халдлагад өртсөн байх магадлалтай гэдэг дүгнэлт гаргаж болно.

ДҮГНЭЛТ

Сүүлийн жилүүдэд интернетийн хэрэглээ хурдтайгаар нэмэгдэхийн зэрэгцээгээр халдлагын тоо болон төрлүүд огцом нэмэгдсээг байна. Иймд бид энэ судалгааны ажлаар халдлага илрүүлэх аргуудыг судлан, сүлжээгээр TCP SYN Flood халдлага хийн, траффик дээр Shannon энтропигийн, логарифмийн аргаас гадна Парето эффектийн аргаар анализ хийж үзсэн болно. Эндээс харахад Парето эффектийн аргыг халдлага илрүүлэхэд боломжтой гэж үзсэн цаашид энэ аргыг өөр олон төрлийн халдлага илрүүлэхэд ашиглан үр дүнг нь гаргана.

НОМ ЗҮЙ

[1] Report news of Vantagepoint http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack

[2] P.Cisar, S. Bosnjak and S.Maravic Cisar “EWMA Algorithm in Network Practice” Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844 Vol. V (2010), No. 2, pp. 160-170

[3] J.Mina and etc “Fault detection for large scale systems using Dynamic Principal Components Analysis with adaptation,” International Journal of Computers, Communications & Control, vol.2, pp. 185-194, 2007.

[4] Ting Liu, Zhiwen Wang, etc “An Entropy based method for Attack Detection in large Scale Network” International J Computer Communication , ISSN 1841-9836 Vol.7 (2012), No.3 (September), pp. 509-517G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.