Халдлага Илрүүлэх Системийн Хосолмол Загварын Судалгаа

Н.Угтахбаяр[1], Б.Өсөхбаяр[1], С.Байгалтөгс[2], Ж.Нямжав[1]
[1] ХШУИСургууль, МУИС, Монгол [2] БУХСургууль, ШУТИС, Монгол
И-мэйл: 44911.n@gmail.com

Хураангуй — Гажигт суурилсан системийг мэдээллийн технологийн салбарт сүүлийн жилүүдэд ихээр ашиглаж байгаа бөгөөд үүнийг халдлага илрүүлэх систем (ХИС)-д ашигласнаар сүлжээний орчинд халдлага танилтыг сайжруулах, сургалтын өгөгдлийг боловсруулах зэрэг судалгааны ажлууд хийгдэж байна. Өнөөг хүртэл ХИС-ийн хосолмол загвар болон гажигт суурилсан систем ашигласнаар үүсэх асуудлуудыг шийдэхээр олон судлаачид судалгааны ажлаа танилцуулсаар байгаа боловч үйлдвэрлэлд нэвтэрсэн загвар хараахан гараагүй. Энэхүү судалгааны ажлын хүрээнд гажигт суурилсан систем болон сигнатурт суурилсан системийг хамтад нь ашигласнаар сүлжээний халдлагыг илрүүлэх танилтыг сайжруулах боломжтой болохыг өөрсдийн боловсруулсан загварын үр дүнгээр харуулах юм. Уг ажилд өөрсдийн боловсруулсан хосолмол загварын үр дүнг туршихдаа KDD 99 сан болон өмнөх ажлын хүрээнд боловсруулсан NUM 15 халдлагын санг ашиглах юм. Бидний санал болгож буй загвар нь Naive Bayes ангиллын аргачлал болон Snort ХИС-ийг параллелаар ашигласнаараа давуу талтай бөгөөд сургалтын машиныг дахин сургах нь илүү хялбар болсон. Уг загварыг ашиглан халдлагыг 99,4%-ийн үр дүнтэйгээр таньж байна.

Түлхүүр үгс — ХИС-ийн хосолмол загвар, халдлага илрүүлэх систем, weka, Naive Bayes

I. ОРШИЛ

Сүүлийн жилүүдэд мэдээллийн технологийн хэрэглээ үйлдвэрийн болон бусад бүх салбарт хүч түрэн орж байна. Үүний улмаас сүлжээний халдлагын төрөл болон хэлбэр өөрчлөгдөн улам бүр аюултай болсоноос шалтгаалж интернэт ашиглах үед мэдээллийн аюулгүй байдлыг хангах зорилгоор халдлага илрүүлэх болон эсэргүүцэх систем ашиглах нь улам бүр чухал болсоор байна. Мэдээллийн аюулгүй байдлыг хангах шифрлэх, баталгаажуулалт хийх, эрхийн хязгаарлалт хийх, галт хана, сигнатурт суурилсан халдлага илрүүлэх болон эсэргүүцэх систем зэрэг олон аргачлал, технологийг өнөөгийн түвшинд ашиглаж байгаа боловч шинээр гарч буй халдлагыг илрүүлэх, таних тал дээр дутагдалтай хэвээр байгаа. Үүний улмаас олон тооны 0 өдрийн халдлага (zero day) өдөр бүр үйлдэгдэж байна.

ХИС-ийн тусламжтайгаар сүлжээний халдлага болон гажиг илрүүлэх боломжтой [12]. Халдлага илрүүлэх аргачлалыг гажигт суурилсан болон сигнатурт суурилсан гэж 2 ангилж үздэг [12].

Сигнатурт суурилсан аргачлал нь 0 өдрийн халдлагыг судлан тухайн халдлагын онцлогыг агуулсан буюу тухайн халдлагыг таних боломжтой сигнатурыг үүсгэх замаар халдлагын мэдээлэл бүхий сигнатурын сан үүсгэж уг сантай урсгалын мэдээллийг тулгаснаар халдлагыг таньж илрүүлдэг [1]. Энэхүү аргачлалын хувьд халдлагын төрөл болон тоо нэмэгдэх тусам ашиглаж буй сангийн хэмжээ нэмэгдэж үүнээс үүдэн урсгалын мэдээллийг шалгах хугацаа нэмэгдэх дутагдалтай талтайгаас гадна 0 өдрийн халдлагыг илрүүлж чаддаггүй. Хэдий дээрх дутагдалтай тал бий боловч судлагдсан халдлагыг алдаагүй таньдагаараа давуу талтай.

Гажигт суурилсан системийн хувьд тухайн илрүүлэх урсгалын шинж чанар, өгөгдлийн мэдээлэл, давтагдах байдал зэрэг хэд хэдэн мэдээллийг ашиглан боловсруулалт хийсний үр дүнд халдлага мөн эсэхийг тодорхойлдог [2]. Энэхүү аргачлалыг ашиглах үед өгөгдлийн боловсруулалт хийх, ангилал хийх нь хамгийн чухал ажил юм. Бидний энэхүү судалгааны үр дүнд гажигт суурилсан системийг сүлжээний ХИС-д ашигласнаар дараах асуудлыг шийдвэрлэх шаардлага гарна гэж үзсэн:

- Сүлжээний урсгалын мэдээлэл нь их өгөгдөлд хамаарах тул боловсруулалтын хугацаа богино буюу бодит хугацааны агшинд байх шаардлагатай тул өгөгдлийн урьдчилсан боловсруулалт чухал.
- Халдлага таних танилтыг нэмэгдүүлэхийн тулд өгөгдлийн боловсруулалтыг зөв гүйцэтгэх, хуурмаг өгөгдөл зэрэг өгөгдлийг цэвэрлэх.
- Сүлжээний урсгалд шинэ төрлийн халдлагын мэдээлэл агуулагдаж байх үед танилт хийх боломжтой.

Өнөөдрийн байдлаар мэдээллийн аюулгүй байдлыг хангах шийдлийг анализд суурилсан (analyst-driven), машин сургалтын аргад суурилсан (machine learning-driven) гэсэн хоёр ангилалд хуваан үзэж байна [25]. Энэхүү судалгааны ажлын хүрээнд 0 өдрийн халдлагыг таних танилтын хувийг нэмэгдүүлсэн зэрэгцээ ажиллагаа бүхий хосолмол загварыг танилцуулна. Бидний танилцуулах аргачлалд сигнатурт суурилсан ХИС болох Snort, гажигт суурилсан ХИС-д Naive Bayes-ийн машин ашиглах юм. Энэхүү хосолмол систем нь өгөгдөл олборлох аргачлалын өгөгдлийн боловсруулалт болон дахин сургах аргачлалуудыг ашиглан халдлага илрүүлэх танилтын хувийг ихэсгэхээс гадна сигнатурт суурилсан систем, гажигт суурилсан систем мөн аюулгүй байдлын аналистын хамтын ажиллагааг ашиглаж байгаагаараа давуу талтай. Уг ажилд энэ төрлийн судалгааны ажилд ихээр ашиглагддаг KDD 99 сан болон бидний өмнөх ажилд [12] цуглуулсан NUM15 санг хослуулан ашиглах юм. [12] ажлын хүрээнд Backtrack үйлдлийн системийг ашиглан туршилтын сүлжээний орчинд 2010 оноос

хойш гарсан халдлага хийж тухайн санг үүсгэсэн. KDD 99-ийн сургалтын сан нь халдлагын төрлөөр тэмдэглэгдсэн 5 сая орчим холболтын мэдээллийг агуулсан бөгөөд уг судалгаанд ашиглахаас өмнө энэхүү санд урьдчилсан боловсруулалт хийсэн. Харин тестийн сан нь сургалтын санд агуулагдаж буй болон агуулагдаагүй холболтын мэдээллээс бүрддэг [3].

II. СУДЛАГДСАН БАЙДАЛ

Сигнатурт суурилсан [4] болон гажигт суурилсан [6] системийг сүлжээний ХИС-д ашиглах судалгаа 1980 оноос эхэлсэн гэж үздэг. Уг чиглэлээр олон тооны эрдэмтдийн бүтээл бий бөгөөд тэдгээрийн гол зорилго нь төрөл бүрийн алгоритмуудаас халдлага зөв таних танилтыг нэмэгдүүлэх аргачлал бий болгоход чиглэж байна. Андерсон болон бусад эрдэмтэд [4] хэрэглэгчийн шинж чанарт тулгуурлан статистикийн аргачлалд суурилсан ХИС-ийн санааг боловсруулж байв. 1987 онд ХИС-ийн анхны загвар бий болсон гэж үздэг [5] бөгөөд энэхүү санаа нь асар хурдацтайгаар нийтэд түгсэн юм. Хэд хэдэн судалгааны ажил сүлжээний орчны халдлага болон гажигыг илрүүлэхэд өгөгдөл олборлолтын аргыг ашиглах талаар хийгджээ [6], [7]. Пакетд суурилсан болон урсгалд суурилсан өгөгдлийн илгээгчийн мэдээлэлд тулгуурлан анализ хийдэг. Пакетд суурилсан аргачлалыг сигнатурт суурилсан ХИС-д түлхүү ашиглагддаг бол урсгалд суурилсан аргачлалыг гажигт суурилсан ХИС-д илүү ашигладаг [8]. [10] судалгааны ажилд машин сургалтын ангилагчдыг давхарга хэлбэртэйгээр ашигласнаар халдлага илрүүлэх танилтын хувийг нэмэгдүүлсэн байна. Уг ажилд Naive Bayes, fuzzy K-NN, back-propagation NN зэргийг ашиглаж KDD 99 сангийн өгөгдлийн 30 онцлогоор туршиж үзжээ. [11] ажилд Naive Bayes ангилагч ашиглан халдлагын үндсэн 4 төрлийг илрүүлэх судалгааг хийж гүйцэтгэжээ. Эдгээр болон бусад ажлуудад DARPA, KDD 99 сангуудыг өргөнөөр ашиглаж байна. Эдгээрээс гадна хосолмол ХИС-ийн судалгаа сүүлийн жилүүдэд олон судлаачдын судалгааны объект болжээ. [16] ажилд Snort ХИС-ийн препроцессийг сайжруулах замаар гажигт суурилсан хосолмол ХИС-ийг танилцуулсан бол Özge Cepheli болон бусад эрдэмтдийн [17] ажилд H-IDS нэртэй DDoS халдлагыг илрүүлэх хосолмол загварыг мөн танилцуулжээ. Уг ажлын үр дүнд H-IDS нь банкны өгөгдөлд анализ хийхэд танилтын хувийг 27.4 хувиар нэмэгдүүлж чаджээ.

Hussein .S. F болон бусад эрдэмтэд [23] ажилд Snort болон Naive Bayes-ийн аргачлалуудыг шатласан хэлбэртэйгээр хослуулан ашиглажээ. Уг аргачлалыг шалгахдаа KDD 99 сан болон Weka програмыг ашигласан байна. Энэхүү судалгааны ажилд Bayes-ийн сүлжээ, J48 graft, Naive Bayes зэрэг аргачлалуудыг харьцуулсан бөгөөд эдгээр дундаас Naive Bayes 92%-ийн танилт үзүүлсэн бөгөөд загварыг дахин сургахад ойролцоогоор 10 минут зарцуулсан нь бусад аргуудаас хурдан хугацаанд суралцаж байжээ.

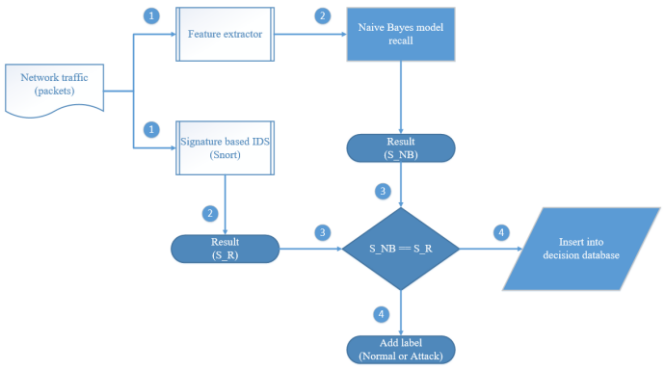
Mradul Dhakar болон бусад эрдэмтэд [24] Tree Augmented Naive Bayes (TAN) болон Reduced Error Pruning (REP) аргачлалуудыг хослуулсан ХИС-ийн хосолмол загварыг ашиглажээ. TAN ангилагчыг үндсэн ангилагчаар REP ангилагчыг TAN ангилагчын мэдээллээс суралцах байдлаар цуваа хэлбэртэйгээр ашиглаж судалгааны үр дүнг KDD 99 сангаар шалгахад 99.96% хувийн танилт үзүүлсэн нь одоогийн байдлаар хамгийн өндөр хувийн танилт болоод байна.

2016 оны 4 сард MIT AI2 [25] нэртэй загвараа танилцуулсан бөгөөд энэхүү загвар нь удирдамжтай болон удирдамжгүй аргачлалуудыг аюулгүй байдлын аналисттай хослуулан ашигласан аргачлал юм. Уг аргачлалын давуу тал нь их өгөгдөлийн орчинд шинж чанарт суурилан ажиллах боломжтой, хуурамч мэдээллийг илрүүлэх, аюулгүй байдлын аналистаас санал авах зэрэг давуу талуудтай гэж үзжээ. Энэхүү аргачлалаа 3,6 тэрбум логын өгөгдөл дээр туршиж 85%-ийн зөв танилт үзүүлжээ.

III. АРГАЧЛАЛ

Бид сургалтын машинаа үүсгэхээс өмнө онцлог сонгох, өгөгдлийг боловсруулах ажлуудыг хийж гүйцэтгэсэн. Үүний үр дүнд хэсэг онцлог нь танилтын хувийг өндөр үзүүлж буйг илрүүлсэн. Уг онцлогыг сангаас ялгасны үндсэн дээр Naive Bayes-ийн аргачлал бүхий машиныг суралцуулсан.

Бидний санал болгож буй загвар нь сигнатурт суурилсан болон гажигт суурилсан аргачлалуудыг Зураг 1-д үзүүлсний дагуу зэрэгцээ ашигласан. Snort ХИС [15] нь өөрийн сигнатурын тусламжтайгаар халдлага илрүүлэх бөгөөд уг үр дүнг гажигт суурилсан ХИС-ийн үр дүнтэй харьцуулах юм. Энэхүү системийн зорилго нь 0 өдрийн халдлагын танилтыг сайжруулахад оршино. Үүнийг аюулгүй байдлын аналисттай хамтран сургалтын машиныг дахин сургах замаар хийж байгаа юм. Хэрвээ уг хоёр системийн үр дүн ижил гарвал тухайн холболтыг хадгалахгүй ба ялгаатай тохиолдолд шинжилгээ хийх санд (decision database) нэмэх юм. Энэхүү шинжилгээ хийх санг администратор эсвэл аюулгүй байдлын аналистууд шалгах ба халдлага мөн эсэхийг тодорхойлж сургалтын машиныг зөв өгөгдлөөр дахин сургах замаар уг загвар ажиллана.



Зураг. 1. Санал болгож буй загвар

Бид уг судалгааны ажилд 2 дахь үеийн Intel i5 2.4GHz CPU, 8GB DDR3 RAM, 1TB SATA диск бүхий компьютер дээр MySql 5.7 бүхий Ubuntu 16 үйлдлийн систем суулгаж ашигласан.

Сүлжээний урсгалыг Wireshark-ийн интерфэйсийн тохиргоог promiscuous горимд ашигласан бөгөөд урсгалын мэдээллийг ".pcap" хэлбэртэйгээр хадгалсан.

Үүний дараа цуглуулсан өгөгдлийг боловсруулж онцлог ялгах ажлыг хийж гүйцэтгэсэн.

Бидний санал болгож буй загвар нь дараах хэсгүүдээс бүрдэнэ:

- Санг шинжлэх болон урьдчилан боловсруулах
- Сигнатурт суурилсан систем
- Naive Bayes ангилагч бүхий гажигт суурилсан систем
- Үр дүн харьцуулагч

A. *Санг шинжлэх болон урьдчилан боловсруулах*

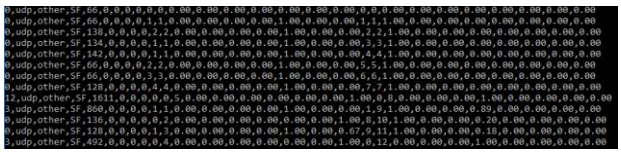
KDD 99 сан нь DoS, Probe, U2R, R2L гэсэн 4 төрлийн халдлагын сантай. Уг санд хадгалагдаж буй өгөгдөл бүр 41 онцлогоос бүрдэх бөгөөд энэхүү сан нь туршилтын болон тестийн гэсэн хоёр хэсгээс гадна нийт сангийн 10% бүхий сангуудаас бүрдэнэ. Энэхүү 41 онцлог нь үндсэн онцлог, контетын онцлог, хугацааны онцлог болон хостын онцлог гэж 4 ангилагддаг [20]. Бид энэхүү судалгаандаа өмнөх судалгаанд ашигласан NUM15 болон KDD 99 сангийн 10 хувийн өгөгдлүүдийг ашигласан. NUM15 нь 300,000 мөр бичиглэл бүхий сан юм.

KDD 99 сан нь дискрет, үргэлжилсэн зэрэг олон төрлийн утгууд бүхий сан бөгөөд эдгээр утгыг сургалтын машинд [0;N] утгаар нормчилж сургасан бөгөөд энгийн болон 4 төрлийн халдлага бүхий 5 ангилал бүхий урсгалын мэдээлэл агуулагдаж байгаа.

KDD 99 сан нь протоколын төрөл, үйлчилгээний төрөл, флаг, ланд, холболтын мэдээлэл, рүүт шэлл, зочин болон хэрэглэгчийн хандалтын тоо зэрэг 9 дискрет утгыг агуулдаг [12, 14]. Энэхүү утгуудыг нормчлоход эвклидийн дистансийн аргачлалыг ашигласан.

Сүлжээний бодит урсгалын мэдээллээс 41 онцлогыг ялгаж авсан жишээг зураг 2-д харуулж байна. Өмнө цуглуулсан урсгалыг tcpreply хэрэгсэлийн тусламжтайгаар интерфэйсд дахин дамжуулж онцлогуудыг ялгаж авсан.

Үүний дараа Markov Blanket болон Pearson корреляцын тусламжтайгаар боломжит онцлогийг



ялган авсан.

Зураг 2. Онцлог ялгах

B. *Сигнатурт суурилсан систем*

Сигнатурт суурилсан ХИС нь өөрийн санд буй буюу өмнө нь судлагдаж тухайн халдлагын онцлогыг бүрэн тодорхойлсон халдлагыг өөрийн сангийн сигнатурын тусламжтайгаар илрүүлдэг. Энэ төрлийн системийн хувьд шинэ халдлагыг илрүүлэх боломжгүй байдаг бөгөөд энэхүү судалгаандаа Snort ХИС-ийг ашиглах ба энэ төрлийн хамгийн түгээмэл ашиглагддаг ХИС юм. [21]. Сигнатурт суурилсан ХИС сонгосон шалтгаан нь [12] [21]:

- Ашиглахад хялбар
- Өөрийн санд буй халдлагыг илрүүлэх буруу танилтын хувь ~0

Энэхүү судалгаанд Snort ХИС-ийг ХИС болон пакет логийн горимд ажиллуулах бөгөөд энэ нь дараах хэсгүүдээс бүрдэнэ [22]:

- Пакет задлагч болон боловсруулагч
- Халдлага илрүүлэх хэсэг
- Логийн болон алертийн хэсэг

C. *Naive Bayes-ийн загвар*

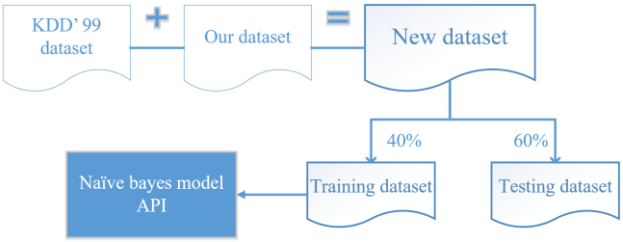
Компьютерийн сүлжээ нь зангилаануудаас тогтох бөгөөд дамжих мэдээлэл нь олон төрлийн хувьсагчууд байх ба тэдгээр нь хоорондоо харилцан хамаарал бүхий байна. Bayes-ийн сүлжээ нь тэдгээр хувьсагчуудын хамаарлыг боловсруулахаас гадна тэдгээрт агуулагдаж буй онцлогууд нь халдлага болон энгийн урсгалтай хэрхэн хамааралтай байгааг илрүүлэх боломжтой юм. Bayes-ийн сүлжээ нь тухайн онцлогуудыг ашиглан профайл үүсгэх бөгөөд тухайн профайлын тусламжтайгаар таних процессыг гүйцэтгэдэг. Өөрөөр хэлбэл тухайн профайл нь хувьсагчуудын мэдээллийг агуулсан цэц гэж ойлгож болох юм. Хэрвээ тухайн профайлын мэдээллээс оролтын өгөгдөл зөрсөн буюу хязгаарыг давсан тохиолдолд алерт өгөх юм. Профайлуудыг агуулсан системийг Bayes-ийн машин эсвэл загвар гэж нэрлэж болох юм.

Bayes-ийн сүлжээний нэгэн төрөл болох Naive Bayes-ийн ангилагчийг уг судалгаанд ашигласан. Тухайн аргачлал нь оролтын өгөгдлийн тусламжтайгаар профайлууд үүсгэх боломжтой бөгөөд үүнийг өөрөөр ангилал гэж хэлж болно. Bayes-ийн ангилагчаар дурын к-ийн хувьд y = C_k гэсэн ангилал үүсгэх боломжтой бөгөөд томъёог доор үзүүлэв:

$$\hat{y} = \underset{k \in \{1,\dots,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i|C_k) \qquad (1)$$

Энэхүү ажилд хэвийн (C_1) болон халдлагатай (C_2) гэсэн хоёр төрлийн ангилал үүсгэн ашигласан.

Зураг 1 болон 2-д үзүүлсэний дагуу бид сүлжээний урсгалыг ‘arff’ форматад хөрвүүлж онцлог сонгох, өгөгдлийг цэвэрлэх, нормчлох зэрэг ажлуудыг гүйцэтгэсэн. Учир нь бидний туршилтанд ашиглах Weka [13] програм нь ‘arff’ форматын өгөгдлийн боловсруулалтыг гүйцэтгэх боломжтой юм. Үүний дараа Зураг 3-д үзүүлсний дагуу өгөгдлийг сургалтын (нийт өгөгдлийн 40%) болон туршилтын (нийт өгөгдлийн 60%) гэсэн 2 хэсэгт хуваасан.



Зураг 3. Naive Bayes model API creating

D. Үр дүн харьцуулагч

Бидний санал болгож буй загварын нэг чухал хэсэг нь сигнатурт суурилсан болон гажигт суурилсан системүүдийн үр дүнг харьцуулах юм. Бид тухайн системүүдийн үр дүнг харьцуулахдаа хугацааны мэдээлэл, холболтын мэдээлэл (session), илгээгч болон хүлээн авагчын логик хаяг зэрэг 4 өгөгдлийг ашигласан. Хэрвээ тухайн хоёр системийн хоорондын өгөгдөл ялгаатай үед тухайн өгөгдлийг аюулгүй байдлын аналист эсвэл системийн администратор шинжлэн зөв төрөлд оруулан Naive Bayes-ийн машиныг дахин сургах ажлыг гүйцэтгэнэ.

```
Start program
If S_NB.source equal S_R.source and not equal to 0 Then
  If S_NB.destination equal S_R.destination Then
    If S_NB.packet_session equal S_R.packet_session Then
      Insert Into decision table
    EndIf
  EndIf
EndIf
End program
```

E. Тооцооллын хэмжигдэхүүн

True positive (TP) – Халдлагыг халдлага гэдгээр нь зөв ангилсан тоо хэмжээ.

True negative (TN) – Хэвийн урсгалыг хэвийн урсгал гэдгээр нь зөв ангилсан тоо хэмжээ.

False positive (FP) – Хэвийн урсгалыг халдлага гэж ангилсан тоо хэмжээ.

False negative (FN) – Халдлагыг хэвийн урсгал гэж ангилсан тоо хэмжээ.

ХИС-ийн accuracy-г дараах томъёогоор бодож олно.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \qquad (2)$$

IV. Туршилтын үр дүн

Энэ хэсэгт бидний санал болгож буй загварын үр дүн болон бусад ижил төстэй аргуудын үр дүнтэй харьцуулсан үр дүнг үзүүлнэ. Уг ажлыг гүйцэтгэхэд бид 41 онцлог болон уг судалгаанд ашиглагдсан Markov blanket болон Pearson-ий корреляцийн үр дүнд сонгогдсон 20 онцлогуудыг ашиглан үр дүнг харьцуулан харуулах юм. Сонгогдсон 20 онцлогын мэдээллийг Хүснэгт 1-д харууллаа.

Хүснэгт 1. Сонгогдсон онцлогууд

№	*Онцлогын нэр*	*Төрөл*	*Ангилал*
1	duration	Cont.	Basic
2	protocol_type	Disc.	Basic
3	Service	Disc.	Basic
4	src_bytes	Cont.	Basic
5	Land	Disc.	Basic
6	wrong_fragment	Cont.	Basic
7	Urgent	Cont.	Basic
8	Flag	Disc.	Basic
9	num_failed_logins	Cont.	Content
10	logged_in	Disc.	Content
11	num_file_creations	Cont.	Content
12	is_guest_login	Disc.	Content
13	Hot	Cont.	Content
14	srv_serror_rate	Cont.	Traffic
15	diff_srv_rate	Cont.	Traffic
16	num_compromised	Cont.	Traffic
17	serror_rate	Cont.	Traffic
18	dst_host_same_src_port_rate	Cont.	Traffic
19	dst_host_srv_serror_rate	Cont.	Traffic
20	dst_host_serror_rate	Cont.	Traffic

Бид туршилтын санг нийт сангийн 20% байхаар буюу 3 тэнцүү хэсэгт хувааж Naive Bayes-ийн загварыг гурван удаа туршиж үзсэн. Хүснэгт 2-д эхний 20%-ийн өгөгдлөөр сургасан үр дүнг харуулж байна. Уг хүснэгтэд нийт 41 онцлог болон сонгогдсон 20 онцлогуудын хоорондын ялгааг харьцуулан харууллаа.

Хүснэгт 2. Сонгогдсон 20 болон нийт 41 онцлогуудын accuracy

	Naive Bayes	
	Сонгогдсон онцлог болон өгөгдлийн боловсруулалт хийгдсэн	Нийт онцлог болон өгөгдлийн боловсруулалт хийгдсэн
Хэвийн	96.2%	99.2%
Халдлага	89.8%	96.8%
Дундаж accuracy	***93%***	***98%***

Үүний дараа үлдсэн 2 хэсэг туршилтын санг ашиглан туршилт хийж гүйцэтгэсэн үр дүнг График 1-ийн 2 (98.40%) болон 3 (99.30%)-р баганад харууллаа. Энэхүү туршилтыг илүү нарийвчлан гаргахын тулд Backtrack үйлдлийн системийн аюулгүй байдлын шинжилгээ хийх хэрэгсэлүүдийг ашиглан Apache вэб сервер, vsFTP серверүүдийг Ubuntu сервер хувилбар дээр сулган, Виндоус сервер 2012-ийг үндсэн байдлаар нь суулгаж халдлага хийж 3531 мөр бүхий тестийн шинэ сан үүсгэж тухайн санг 2 хэсэг хуваан 3 дахь туршилтын дараа сургасан машинаа ахин 2 удаа туршиж гарсан үр дүнг График 1-ийн 4 (99.33%) болон 5 (99.44%)-р баганад харуулав.

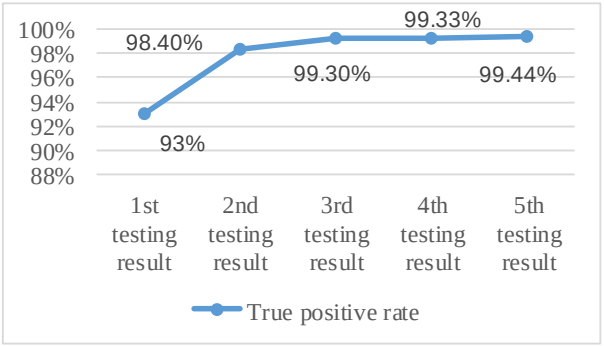


График 1. Туршилтын үр дүн

График 1-ээс дүгнэвэл тодорхой тооны сургалтын дараа тухайн загварын халдлага болон энгийн урсгал таних танилтын хувь тогтвортой буюу бага хувиар өсөж байгаа нь харагдаж байна. Бидний туршилтын үр дүнгээр 3 удаагийн сургалтын дараа зөв танилтын өсөлтийн хувь тогтворжсон үр дүнг үзүүллээ.

Бидний дараагийн туршилтын үр дүнд сигнатурт суурилсан болон гажигт суурилсан системүүдийн танилтын хугацааг харьцуулан харууллаа. Үр дүнг 20 онцлогтой үед тооцож График 2-д харуулав.

Энэхүү графикаас харвал сигнатурт суурилсан систем нь гажигт суурилсан системээс зарим үед илүү хурдан ажиллаж байгаа нь харагдаж байна. Иймд уг хоёр системийг хослуулан ашиглахад халдлагыг илүү хурдан таних боломжтой гэж үзэж байна.

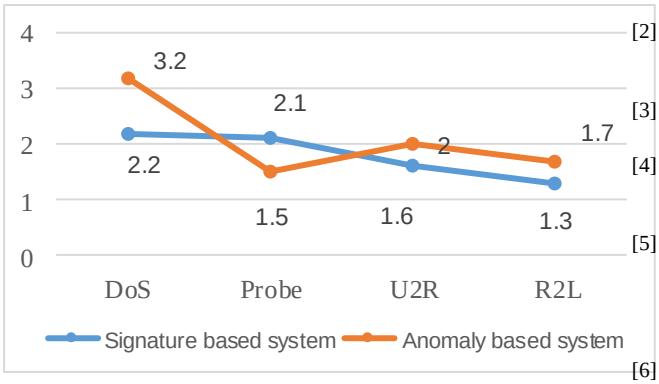
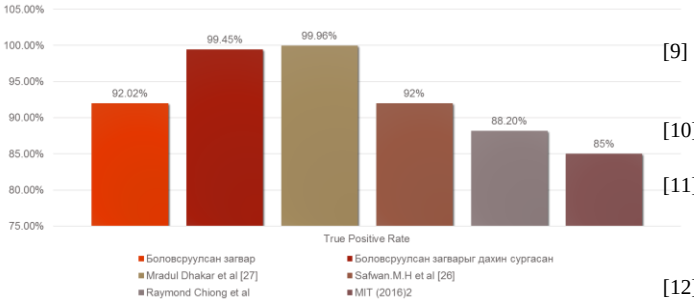


График 2. Халдлага таних хугацаа

Бид судалгааны ажлын үр дүнгээ бусад ижил төстэй судлаачдын ажилтай харьцуулан дараах хүснэгтэд харууллаа.



Хүснэгт 3. Ижил төстэй ажлуудтай харьцуулсан судалгаа

Хүснэгт 3 болон судлаачдын ажлаас дүгнэвэл энэ төрлийн ажил нь сургалтын болон туршилтын сангийн өгөгдлөөс мөн тухайн өгөгдөлд урьдчилсан боловсруулалт хэрхэн хийснээс хамаарч ялгаатай танилт үзүүлж байгаа нь ажиглагдаж байна.

Дүгнэлт

Бидний санал болгож буй систем нь сигнатурт суурилсан болон гажигт суурилсан системүүдийг зэрэгцээ байдлаар ашигласнаараа хосолмол загварын хувьд нэгэн шинэ санааг бий болгож байгаа юм. Уг системийн нэгэн давуу тал нь уг системүүдийг хослуулан ашигласнаар Naive Bayes загварыг Snort ХИС-ээр туршиж үзэх боломжийг олгож байгаад оршино. Учир нь Snort ХИС нь сигнатурт суурилсан ХИС-ийн хувьд хамгийн өргөн ашиглагддаг систем юм.

Уг судалгааны үр дүнд ХИС-ийн хосолмол загвар нь халдлага таних танилтыг сайжруулахаас гадна халдлагыг илүү богино хугацаанд таних боломжийг олгож байна гэж дүгнэж байна. Цаашид уг системийг өөрөө сурдаг систем болгон сайжруулснаар халдлагыг цаг алдалгүй хурдан таних боломжтой болно гэж үзэж байна.

Ном зүй

[1] A. K. Pathan, “The state of the Art in Intrusion Prevention and Detection,” CRC press, 2014.

[2] M.Naga Surya Lakshmi, Dr. Y. Radhika “A complete study on intrusion detection using data mining techniques” Volume IX, IJCEA Issue VI, June 2015.

[3] Miroslav Stampar “Artificial Intelligence in Network Intrusion Detection”.

[4] J.P.Anderson, “Computer security threat monitoring and surveillance” technical report, James P. Anderson Co., Fort Washington, Pennsylvania, 1980.

[5] L.Zenghui, L.Yingxu, “A data mining framework for building Intrusion detection models based on IPv6” Proceedings of the 3rd International conference and Workshops on Advances in Information Security and Assurance. Seoul, Korea, Springer-Verlag, 2009.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] G.Androulidakis and S.Papavassiliou, “Improving network anomaly detection via selective flow-based sampling”, Communications IET, pp. 399-409, 2008.

[9] Machine learning, <http://en.wikipedia.org/wiki/Machine_learning/>, 2014 (accessed 01.25.16)

[10] C.Te-Shun, J.Fan and M.Kia, “Ensemble of machine learning algorithms for intrusion detection.” pp. 3976-3980

[11] S.Neelam and M.Saurabh, “Layered approach for intrusion detection using Naive Bayes classifier” Proceedings of the international conference on Advances in computing, Communications and Informatics, India, 2012.

[12] Ugtakhbayar.N, Usukhbayar.B and Nyamjav.J "Improving accuracy for anomaly based IDS using signature based system" International Journal of Computer Science and Information Security, Vol.14, No.5, pp. 358-361, 2016

[13] Weka. <http://weka.sourceforge.net/>, 2012 (accessed 05.02.15)

[14] Y.Wang, K.Yang, X.Jing, H.L.Jin, "Problems of KDD Cup 99 Dataset Existed and Data Preprocessing", Applied Mechanics and Materials, Vol. 667, pp. 218-225, 2014

[15] Snort. <http://www.snort.org/>, 2001 (accessed 21.01.15)

[16] J.Gómez, C.Gil, N.Padilla, R.Baños and C.Jiménez, "Design of Snort-Based Hybrid Intrusion Detection System", IWANN 2009, pp. 515-522, 2009

[17] Özge Cepheli, Saliha Büyükçorak and Güneş Karabulut Kurt, "Hybrid Intrusion Detection System for DDoS attacks", Journal of Electrical and Computer Engineering, vol. 2016, Article ID 1075648, 2016

[18] R.A.Maxion and R.R.Roberts. "Proper use of ROC curves in intrusion/anomaly detection", Technical report CS-TR-871, 2004