Хиймэл оюуны хэрэгсэл ашиглан сүлжээний урсгалыг шинжлэх

Н.Угтахбаяр, Ш.Содбилэг
МУИС, ХШУИС
ugtakhbayar@num.edu.mn, sdblg@num.edu.mn

Хураангуй **– Халдлагатай болон халдлагагүй сүлжээний урсгалыг цуглуулан халдлагыг таних систем хийх талаар судалгаа сүүлийн жилүүдэд олноор хийгдэж байна. Тэр дундаа KDDCUP өгөгдлийн санг ашиглан халдлага илрүүлэх системийн загварчлалыг хийсэн судалгаа илүү хийгдэж байгаа.**

Ихэнх халдлага илрүүлэх системийн хувьд хэвийн урсгалыг халдлагын урсгал гэж таних буюу хуурамч анхааруулга өгөх явдал их гардаг. Бид энэхүү судалгааны ажлаараа халдлагад өртсөн урсгалаар KDDCUP'99 санг ашиглах бөгөөд энгийн сүлжээний урсгалыг МУИС-ийн сүлжээг ашиглан цуглуулж хиймэл оюуны програмаар тэдгээрийг харьцуулан судлана. Энгийн урсгалыг Watchguard XTM, Cisco IOS ACL ашиглан цуглуулна. Цуглуулсан урсгалын OSI-ийн 2, 3, 4-р түвшний мэдээллүүдийг задлан өгөгдлийн санд цуглуулна. Өөрсдийн цуглуулсан урсгалыг ашиглан хиймэл оюуны хэрэгсэлийг суралцуулах бөгөөд KDDCUP санг ашиглан танилтын хувийг тодорхойлно.

Түлхүүр үг **— сүлжээний урсгал цуглуулах, Хиймэл оюун ухаан, Weka, J-48, KDDCUP'99.**

I. ОРШИЛ

Хиймэл оюуны аргачлал ашигласан халдлага илрүүлэх болон эсэргүүцэх системийн хөгжүүлэлтийг хийх талаарх судалгаа, туршилтын ажлууд сүүлийн жилүүдэд ихээр гарах болсон. Мөн байгууллага, хувь хүний интернэт хэрэглээ асар хурдтай өсөн нэмэгдэхийн хэрээр тухайн хэрэглэгчийн аюулгүй байдлыг хангах асуудал улам бүр түвэгтэй болж байна.

Сүлжээ болон системийн администрууд аюулгүй байдлыг сайжруулах зорилгоор янз бүрийн арга ашигладаг бөгөөд энэ нь эргээд сүлжээний ачааллыг нэмэгдүүлэх, хурдыг багасгах хортойгоос гадна зарим аюулгүй байдлын хэрэгсэлээр дамжиж сүлжээнд аюул учрах боломжтой болдог. [1]

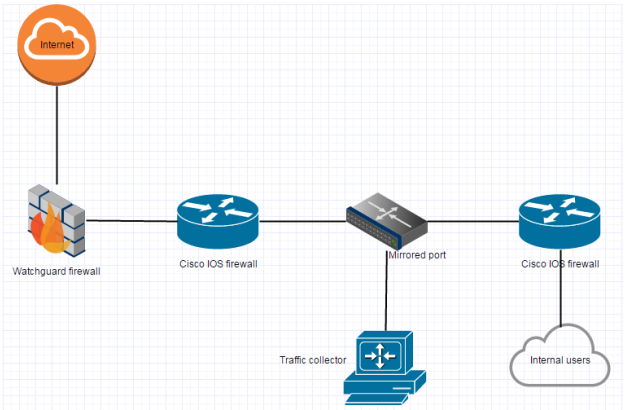
Иймд сүлжээний аюулгүй байдлын төхөөрөмжүүд нь сүлжээний хурдад нөлөөлөх нөлөөлөл багатай байх ёстой гэсэн үндсэн шаардлага тавигдаж байна. Энэхүү ажлыг хийж гүйцэтгэснээр цаашид холимог халдлага илрүүлэх систем хөгжүүлэх суурийг бүрдүүлж өгөх юм. Уг аргачлал нь сүлжээнийн халдлагын хувьд өмнө танигдсан халдлагыг тухайн агшинд илрүүлэх бөгөөд урьд өмнө нь тухайн системд тохиолдож байгаагүй халдлагыг системд ачаалал өгөхгүйгээр илрүүлэх аргачлал юм. Халдлагууд дундаас DoS, probe төрлийн халдлагууд хамгийн их тохиолддог мөн хамгийн үр дүнтэй халдлагын төрөл юм. Жишээ нь arbornetworks.com сайтад бичсэнээр 2015 оны 1 сарын 15-д Франц улсад DoS төрлийн халдлага хийгдэж ойролцоогоор 19 мянган сайт хэдхэн өдрийн дотор халдлагын бай болсон байна.

Мөн энгийн DoS халдлага нь SYN-flooding халдлага байдлаар TCP2ийн үйлчилгээг бусниулах үйл ажиллагаатай байдаг [10]. Нийт TCP халдлагын 90 хувь нь DoS-ийн төрлийн халдлага байна [11].

Иймд энэхүү судалгааны ажлын хүрээнд компьютерийн сүлжээний халдлага дотор илүү тохиолддог DoS, probe төрлийн халдлагуудад анализ хийж тэрхүү халдлагын төрлийг таних хиймэл оюунт алгоритмыг турших ажлыг хийж гүйцэтгэнэ. Энэхүү судалгааны ажлыг хийж гүйцэтгэснээр цаашид олон алгоритм дээр туршиж дээрх 2 төрлийн халдлагыг хурдан, найдвартай илрүүлэх боломжтой алгоритмыг олоход чухал нөлөө үзүүлнэ гэдэгт итгэлтэй байна. Судалгааны ажил нь үндсэн хоёр хэсгээс бүрдэж байгаа бөгөөд эхний хэсэгт МУИС-ийн гарцны төхөөрөмжийг ашиглан сүлжээний урсгалыг цуглуулж хиймэл оюуны аргачлалаар суралцуулах. Дараагийн алхамд өөрсдийн сангаар сургасан өгөгдлөө KDDCUP'99 сангаар турших, ингэхдээ үр дүнг сайжруулах зорилгоор шугаман корреляцын арга ашигласан. Дээрх туршилтыг хийж халдлагыг илрүүлэхэд OSI загварын 2, 3, 4-р түвшний толгойн мэдээллийн аль параметрууд илүү нөлөөтэй болохыг тодорхойлох юм.

II. СУДАЛГААНЫ АРГА ЗҮЙ

Уг ажлын хувьд хэвийн урсгал цуглуулах системийг CPU: Core2Duo 2.0GHz, RAM: 8GB, 1Gbps сүлжээний карт бүхий машин дээр суулгаж МУИС-ийн гарцны switch-ийн портыг бүх дотоод болон гадаад урсгалыг хуулбарлан дамжуулах байдлаар тохируулан сүлжээний эрүүл урсгалыг tcpdump хэрэгсэлийг ашиглан барьж авсан /Зураг 1-д сүлжээний загварыг харуулав/.



Зураг 1. Сүлжээний урсгалыг цуглуулсан топологи

Цуглуулсан сүлжээний урсгалаас OSI загварын 2, 3, 4-р түвшний мэдээллийг KDDCUP өгөгдлийн сангийн хэлбэрт оруулан csv файл болгон бэлдсэн. Туршилтад ашиглах компьютерийн хүчин чадлаас шалтгаалан 1GB хэмжээтэй урсгалыг боловсруулалтад ашигласан бөгөөд хугацааны хувьд өглөөний 9 цагаас 11 цагийн хоорондох урсгалаас түүвэрлэн авсан. Уг урсгалаа TCP, UDP протоколоор тусд нь файл бэлдсэн.

III. ТУРШИЛТ, ҮР ДҮН

Туршилтын хэсгийн хамгийн эхний ажил туршилтанд ашиглах онцлог шинжүүдийг сонгох явдал байсан бөгөөд өмнөх судалгааны ажлууд дээр тулгуурлан дараах байдлаар сонгосон. Үүнд:

- Холболтын хугацаа
- Протокол
- Порт / Төрөл
- Флагууд
- Илгээгч хүлээн авагчийн OSI-ийн 2, 3-р түвшний мэдээлэл зэрэг нийт 15 төрлийн мэдээллийг ашигласан.

Урсгалд ашиглан онцлог шинжийг сонгохоос өмнөх мэдээллийн зарим хэсгийг зураг 2-т харууллаа.

```
@relation Classification
@attribute duration real
@attribute protocol_type {tcp
@attribute service string
@attribute flag string
@attribute src_bytes real
@attribute dst_bytes real
@attribute land real
@attribute wrong_fragment real
@attribute urgent real
@attribute hot real
@attribute num_failed_logins real
@attribute logged_in real real
@attribute num_compromised real
@attribute root_shell real
@attribute su_attempted real
@attribute num_root real
@attribute num_file_creations real
@attribute num_shells real
@attribute num_access_files real
@attribute num_outbound_cmds real
@attribute is_host_login real
@attribute is_guest_login real
@attribute count real
@attribute srv_count real
@attribute serror_rate real
@attribute srv_serror_rate real
@attribute rerror_rate real
@attribute srv_rerror_rate real
@attribute same_srv_rate real
```

Зураг 2 Онцлог сонгохоос өмнөх мэдээллийн зарим хэсэг

Онцлогийг сонгосны дараа Weka хэрэгсэл ашиглан J-48 алгоритмаар цуглуулсан өгөгдлөө сургаж KDDCUP’99 санг ашиглан танилтын хувийг тодорхойлох ажлыг хийж гүйцэтгэсэн. Суралцуулахад нийт 30000 мөр өгөгдөл ашигласан.

Аттрибут	TCP урсгалыг танилтын хувь	UDP урсгалыг танилтын хувь
Танилтын хувь	87.3	85.1

Хүснэгт 1 Танилтын хувь

Сонгосон нийт онцлогуудыг ашиглахад таних болон суралцах процесс харьцангуй удаан буюу суралцах хугацаа 5 минут таних хугацаа 3,5 минут орчим байсан тул зарим онцлогыг хасах шаардлагатай болсон. Мөн танилтын хугацааг богиносгохын тулд Spearman-ы корреляцын аргыг ашиглан танилтын хувийг дахин бодсон. Үр дүнг дараах хүснэгтүүдэд харууллаа. Хүснэгт 2-д KDDCUP’99 өгөгдлийн утгын зарим хэсгийг хүснэгт 3-д өөрсдийн цуглуулсан урсгалын өгөгдлийн утгын зарим хэсгийг тус бүр харуулав.

ip_dst	time	tcp_sport	tcp_dport	tcp_seq	tcp_ack
0	0	0	0	0	0
1	0	0	0	0	0
0	1	0.97	0	0.0074	0
0	0.97	1	0	0.0086	0
0	0	0	1	0	0
0	0.0074	0.0086	0	1	0

Хүснэгт 2 халдлага бүхий урсгал

ip_dst	time	tcp_sport	tcp_dport	tcp_seq	tcp_ack
-0.07121	0.549692	0.223295	-0.1057	0.0022	0.165121
1	0.54553	-0.18875	0.3111	0.217	-0.016
0.54553	1	0.034116	0.1884	0.2611	0.1413
-0.18875	0.034116	1	-0.9692	-0.463	0.4427
0.311146	0.18847	-0.96924	1	0.511442	-0.4119
0.217049	0.261126	-0.463	0.5114	1	-0.26352

Хүснэгт 3 Энгийн урсгал

Дээрх аргын дагуу бүх урсгалыг боловсруулсан бөгөөд үүний дараа танилтын хувийг DoS, Remote to Local (R2L), Probe халдлагын хувьд боловсруулж үр дүнг хүснэгт 4, 5, 6-д тус тус харууллаа.

Аттрибут	Сонгож авсан 10 онцлог
Танилтын хувь	91.6

Хүснэгт 4 Spearman-ий корреляц ашигласны дараа DoS төрлийн халдлага илрүүлэх танилтын хувь

Аттрибут	Сонгож авсан 10 онцлог
Танилтын хувь	89.8

Хүснэгт 5 Spearman-ий корреляц ашигласны дараа R2L халдлага илрүүлэх танилтын хувь

Аттрибут	Сонгож авсан 10 онцлог
Танилтын хувь	89.5

Хүснэгт 6 Spearman-ий корреляц ашигласны дараа Probe төрлийн халдлага илрүүлэх танилтын хувь

Энэхүү туршилтын дараагийн алхам болгон танилтын хувийг илүү нарийн тодорхойлохын тулд халдлагыг таних туршилтанд халдлагатай урсгал 50 хувь + энгийн урсгал 50 хувь байхаар KDDCUP болон өөрсдийн цуглуулсан өгөгдлийг хольж үр дүнг боловсруулахад гарсан үр дүнг хүснэгт 7-д харууллаа. Энэхүү туршилтыг хийхэд аттрибут тус бүрээр хийсэн. Үүний үр дүнг хүснэгт 7-д үзүүллээ.

	DoS	R2L	Probe
Танилтын хувь	95.3	92	92.2

Хүснэгт 7 Холимог урсгал ашиглахад халдлагыг илрүүлэх танилтыг хувиар илэрхийлэв

IV. ДҮГНЭЛТ

Энэхүү судалгааны ажлын хүрээнд МУИС-ийн сүлжээг ашиглан энгийн өгөгдөл цуглуусан бөгөөд өөрийн гэсэн энгийн урсгалын сантай болсон. Мөн халдлага дундаас DoS, R2L, probe төрлийн халдлагуудыг таних танилтын хувийг J-48 алгоритмаар тодорхойллоо.

Мөн туршилтын дараагийн хэсгээр холимог өгөгдөл боловсруулж халдлага илрүүлэх танилтын хувийг тодорхойллоо.

Энэхүү ажлын хувьд танилтын хувь бага гарч байгаа нь халдлагагүй өгөгдлийг өөрсдийн систем ашиглан цуглуулсантай холбоотой гэж үзсэн. Хиймэл оюуны хэрэгсэл ашиглан халдлага илрүүлэх ажилд онцлог сонгохоос гадна халдлагатай болон энгийн урсгал цуглуулах тэдгээрийг нормчлох ажил чухал болохыг дээрх судалгаа харууллаа.

Уг ажлыг цааш үргэлжлүүлэн халдлагагүй урсгал цуглуулах системийг илүү нарийн зохион байгуулах шаардлагатай гэж үзсэн.

АШИГЛАСАН МАТЕРИАЛЫН ЖАГСААЛТ

[1] Daniel J. Arndt, A. Nur Zincir-Heywood, “A Comparison of Three Machine Learning Techniques for Encrypted Network Traffic Analysis”

[2] Ciza Thomas, Vishwas Sharma, N.Balakrishnan “Usefulness of DARPA Dataset for Intrusion Detection System Evaluation”

[3] M. V. Mahoney, P. K. Chan, An analysis of the 1999 DARPA /Lincoln Laboratory evaluation data for network anomaly detection, Technical Report CS-2003-02

[4] Supreeth Burji, Kathy J.Liszka, C.C.Chan “Malware analysis using reverse engineering and data mining tool” International conference on system science and engineering. 2010.\

[5] S.Ganapathy, P.Yogesh and A.Kannan, “Intelligent Agent-based Intrusion Detection System Using Enhanced Multiclass SVM”, Computational Intelligence and Neuroscience 2012

[6] Ketki Arora, Krishan Kumar, Monika Sachdeva “Impact analysis of recent DDoS attacks” IJCSE.

[7] Pars Mutaf “Defending against a Denial-of-service attack on TCP”.

[8] Richard Lippmann, David Fried, Keith Piwowarski, William Streilein, “Passive Operating System Identification From TCP/IP Packet Headers*”

[9] Yasemin Gokcen, Vahid Aghaei Foroushani, A. Nur Zincir-Heywood “Can we identify NAT behavior by analyzing Traffic Flows?”, IEEE Security and Privacy Workshops, 2014

[10] Computer Emergency Response Team “TCP SYN flooding and IP spoofing attacks” CERT advisory: CA 96-21, 1996.

[11] G.Gemona, I.Duncan, A.Miller “NEMESI: Using a TCP finite state machin against TCP SYN flooding attacks”