Mendedahkan Ciri Terpilih Yang Mempengaruhi Pengecaman Botnet HTTP

Wan Ahmad Ramzi Y.W[1]
[1] Kolej komuniti Masjid Tanah,Melaka
ramzi@kkmt.edu.my

Nur Hidayah M. S [2]
[2] Universiti Teknikal Malaysia Melaka (UTeM)
[2] nurhidayahmohdsaudi@gmail.com

Faizal M. A[2]
[2] Universiti Teknikal Malaysia Melaka (UTeM)
[2] faizalabdollah@utem.edu.my

ABSTRACT. Botnet dikenalpasti sebagai salah satu ancaman yang paling banyak muncul kerana penjenayah Siber berusaha gigih untuk menjadikan sebahagian besar pengguna rangkaian komputer sebagai sasaran mereka. Oleh itu, ramai penyelidik telah menjalankan banyak kajian mengenai botnet dan cara untuk mengesan botnet dalam trafik rangkaian. Kebanyakan mereka hanya menggunakan ciri di dalam sistem tanpa menyebut pengaruh ciri dalam pengesanan botnet. Pemilihan ciri adalah penting dalam pengesanan botnet kerana ia boleh meningkatkan ketepatan pengesanan. Selain itu, penyelidikan sedia ada lebih menumpukan kepada teknik pengecaman daripada mendedahkan tujuan di sebalik pemilihan. Dalam penyelidikan ini, kaedah pembelajaran mesin yang diselia telah digunakan dan fokus utama adalah pada teknik pemilihan ciri yang akan mendedahkan ciri pengaruh dalam pengesanan botnet menggunakan kaedah statistik. Keputusan yang diperoleh menunjukkan ketepatan adalah kira-kira 91% yang boleh diterima untuk menggunakan ciri pengaruh dalam mengesan aktiviti botnet seterusnya mengesahkan pendekatan statistik terbukti membezakan kehadiran botnet HTTP dalam trafik rangkaian.

KATA KUNCI: botnet HTTP; pengesanan botnet; pemilihan fitur; analisis rangkaian; pendekatan statistikal

1 PENGENALAN

Mutakhir ini, botnet telah dikenal pasti sebagai salah satu ancaman yang paling berkembang terhadap keselamatan internet yang juga menjadi perhatian ramai. Ancaman ini mampu memudaratkan organisasi korporat atau kerajaan, menurut Laporan M. C. E. R. T(2022) yang dilaporkan oleh Keselamatan Siber Malaysia, ejen botnet mendahului dengan jumlah kes yang besar terutamanya penipuan dan spam menduduki tempat 2 teratas dalam jumlah kes yang dilaporkan kepada Keselamatan Siber dari Januari hingga Julai tahun 2022. Selain itu, menurut Warwick (2017), pada Januari 2016, laman web perbankan dalam talian dan aplikasi mudah alih HSBC telah dimatikan buat sementara waktu akibat serangan DDoS (Distributed Denial-of-Service). Khattak, S., Ramay, N.R., Khan, K.R., Syed, A.A. and Khayam, S.A (2014) menyatakan bahawa serangan DDoS secara langsung yang menyerang laman sesawang atau sistem yang biasanya disasarkan kepada syarikat, institusi, kerajaan dan syarikat keselamatan. Akibatnya, ia mengedarkan spam melalui tulang belakang rangkaian, memudahkan penafian akses yang sah, perkhidmatan terjejas dan kerugian kewangan. Selain itu, Eric Auchard,(2016) melaporkan pada November 2016, kira-kira 4.5% daripada 20 juta pelanggan talian tetap pelanggan Deutsche Telekom di Jerman mengalami gangguan rangkaian akibat botnet Mirai. Jenis botnet ini bertujuan untuk mengubah peranti rangkaian menjadi "bot" kawalan jauh yang boleh digunakan untuk melancarkan serangan rangkaian berskala besar. Oleh itu, tindakan penting harus dipertimbangkan untuk melindungi organisasi daripada sebarang ancaman jahat keselamatan yang berlaku di dalam rangkaian.

Menurut J.Jabez and Muthukumar (2015) *Intrusion Detection System* (IDS) ialah sistem rangkaian yang dibina untuk menganalisis aktiviti rangkaian dan mengenali corak mencurigakan yang boleh membahayakan rangkaian dan menjana laporan kepada system pengurusan. Dengan memperkenalkan IDS ini, kemungkinan kerosakan yang berlaku di dalam rangkaian mungkin boleh dikurangkan. Pendekatan Berasaskan Anomali dan Berasaskan Tandatangan ialah dua jenis pendekatan untuk mengesan aktiviti pencerobohan. Berdasarkan V. Jyothsna, V. V. Rama Prasad and K.Munivara Prasad(2011) Berasaskan Anomali ialah teknik yang memantau aktiviti trafik rangkaian luar biasa seperti jumlah trafik yang tinggi dan trafik luar biasa pada *port* rangkaian. Manakala, Berasaskan Tandatangan ialah salah satu kaedah pengesanan yang memerhatikan aliran sistem dan mencari contoh botnet sedia ada yang boleh melakukan pengecaman segera dan lebih mudah untuk disediakan.

Menurut Manikandan, Gopi, Abirami, S. (2021) pemilihan ciri memainkan peranan penting dalam kebanyakan masalah ramalan dan banyak aplikasi. Memilih ciri adalah penting sebelum memperkenalkan IDS sebagai alat pertahanan. Ini kerana kejayaan pengesanan bergantung pada set ciri yang terlibat dalam mengesan aktiviti botnet. Selain itu, tiada lagi ciri khusus dalam pengesanan botnet yang mungkin digunakan untuk mengesan serangan botnet dalam rangkaian. Setiap penyelidik menggunakan nama tersendiri untuk subset yang serupa manakala yang lain menggunakan nama yang serupa tetapi jenis yang berbeza. Walaupun pemilihan ciri signifikan ini penting, namun tiada penyelidikan khusus mengenai set ciri yang digunakan dalam mengesan aktiviti botnet *Hypertext Transfer Protocol* (HTTP). Penyelidikan sedia ada lebih memfokuskan kepada teknik pengecaman dan bukannya mendedahkan tujuan di sebalik pemilihan. Selain itu, kebanyakan penyelidik hanya menggunakan ciri di dalam sistem tanpa menyebut ciri dipengaruhi dalam pengesanan atau pengecaman botnet.

Oleh itu, adalah perlu untuk mendedahkan ciri dipengaruhi dalam pengesanan botnet untuk mengatasi kesukaran mengenali aktiviti botnet dalam rangkaian. Sehubungan dengan itu, Kajian ini akan mendedahkan ciri dipengaruhi dalam pengesanan botnet menggunakan pendekatan statistik dan analisis perbandingan daripada penyelidik terdahulu dengan mensyorkan satu set ciri signifikan terpilih untuk digunakan dalam mengesan aktiviti botnet dalam rangkaian.

2 SOROTAN KAJIAN

Penemuan ciri yang signifikan adalah sangat penting untuk mengelakkan masalah seperti set data berlebihan dan bertindihan. Ini kerana ia boleh mengurangkan salah klasifikasi data dan menghasilkan set ciri terbaik daripada dataset kemudian menghasilkan kadar pengesanan yang lebih baik. Sebagai contoh, dalam kajian, Begum et al. (2017) menemui set ciri optimum dalam mengesan kanser payudara daripada set data microarray. Set ciri ini menghasilkan prestasi yang lebih baik dalam ketepatan pengesanan yang mampu mengesan kanser payudara menggunakan pembelajaran aktif dengan ketepatan 94%. Berbeza dengan Abedinia et al. (2017), menggunakan subset minimum ciri daripada set asal untuk mendapatkan beban kecekapan dan ramalan harga kuasa elektrik. Daripada keputusannya, menghapuskan input ciri yang tidak berkesan menunjukkan peningkatan ketepatan ramalan.

Walau bagaimanapun, Chung et al. (2017) memilih ciri signifikan dengan menggunakan rangka kerja *fuzzy rule*. Pengarang mengawal redundansi untuk membuatkan sistem mereka mengelakkan ralat pengukuran dalam ciri tertentu. Hasil percubaan pengarang ini membuktikan bahawa ciri terpilih dengan kawalan berlebihan memberikan prestasi terbaik berbanding set data menggunakan parameter percuma. Pada 2017, Ahsen etal. (2017) memperoleh set ciri setiap sampel untuk meramalkan metastasis dalam kanser endometrium. Ciri ini digunakan untuk menguji latihan dan kohort ujian. Keputusan kohort latihan mencapai ketepatan 100% manakala dalam kohort ujian ia dibahagikan dengan kes positif nod (ketepatan 90%) dan kes nod-negatif (ketepatan 80%). Output ini menyebabkan peningkatan ketara dalam anggaran metastasis limfa dalam pesakit kanser endometrium. Selain itu, Radha et al. (2017) memilih ciri optimum untuk mengurangkan ralat kadar pengelas dan meningkatkan kadar anggaran dengan ketepatan tambahan. Berbeza dengan pengarang Osman et al. (2017), dia menggunakan kaedah ciri tertanam dalam ramalan pepijat. Hasil kajian beliau menunjukkan pengurangan ralat ramalan regressor dan meningkatkan kestabilan. Kesemua penyelidikan sedia ada ini telah dijalankan kajian tentang ciri-ciri signifikan namun pengkaji tidak menumpukan kepada pengesanan botnet. Oleh itu, bahagian ini juga akan diterangkan tentang definisi aliran, pemilihan ciri dan pendekatan statistik.

2.1 DEFINISI ALIRAN

Menurut B. Claise (2008) satu set paket IP yang melepasi titik cerapan semasa selang masa tertentu dalam rangkaian dikenali sebagai aliran. Setiap paket mempunyai maklumat atau ciri tersendiri yang boleh digunakan dalam mengesan kehadiran aktiviti botnet. Kenyataan ini disokong oleh Zhao (2013) yang menyatakan ciri berasaskan aliran adalah penting bagi mengenal pasti kehadiran bot serta aktivitinya dalam rangkaian. Selain itu, Stevanovic, M. and Pedersen, J.M (2014) menyatakan ketepatan pengesanan yang tinggi boleh disediakan dengan memantau aliran trafik. Sementara itu, Saad et al. (2011) menyatakan bahawa pengesanan botnet boleh dikenal pasti dengan memerhatikan ciri trafik berasaskan aliran. Ciri aliran mesti diekstrak daripada pengepala paket untuk memilih ciri yang paling sesuai. Oleh itu, pemilihan ciri adalah penting kerana ia bergantung pada ciri untuk menghasilkan hasil yang lebih baik dalam pengesanan botnet.

2.2 PEMILIHAN CIRI

Pemilihan ciri ialah kaedah memilih subset pembolehubah dalam set latihan dan hanya menggunakan subsetnya sebagai ciri untuk memberikan hasil ramalan yang lebih baik. Menurut Bolon-Canedo (2011), pemilihan ciri adalah teknik untuk menghapuskan ciri-ciri yang bertindih dan tidak diperlukan. Ianya penting kerana sesetengah ciri mungkin mempunyai subset ciri lain. Pemilihan ciri dikategorikan kepada Model Penapis yang menggunakan kaedah anggaran dan Model Pembalut dengan kaedah mendapatkan ciri yang sesuai melalui pengulangan proses aplikasi pengelasan algoritma dengan set ciri yang berbeza. Dalam kajian ini, model Pembalut Pilihan Hadapan akan digunakan. Ini kerana menurut Beigi et al. (2014), model pembalut adalah kaedah terbaik untuk mengesan botnet kerana mampu menentukan subset ciri yang paling berkesan yang menghasilkan ketepatan pengesanan.

Menurut Chen et al. (2010) Kesukaran untuk mengesan botnet HTTP kerana ia bersembunyi di sebalik aliran biasa HTTP adalah faktor untuk mengenal pasti aliran tidak normal dalam trafik web. Penulis telah menggunakan nombor port sumber, nombor port destinasi, alamat IP sumber, alamat IP destinasi dan protokol untuk membezakan trafik web yang tidak normal daripada permintaan web biasa. Berdasarkan kenyataan Cai, T and Zou, F (2012) Botnet HTTP tidak mengekalkan sambungan ke pelayan Perintah & Kawalan (C&C) dan akan menamatkan sesi dan akan mewujudkan semula untuk transaksi baharu. Ini akan membawa kepada bilangan percubaan sambungan TCP keluar menjadi besar. Oleh itu, untuk mengesan keadaan ini berlaku, nisbah paket TCP masuk kepada keluar setiap selang masa dan nisbah paket TCP kepada jumlah bilangan paket setiap selang masa perlu dicari.

Pada 2017, E.popoola, A.Adewumi (2014) memperoleh set ciri daripada dataset pencerobohan NSL-KDD dengan menggunakan evolusi pembezaan diskret dan algoritma pembelajaran mesin C4.5. Keputusan mereka menunjukkan perubahan ketara dalam ketepatan pengesanan yang mampu mengesan serangan baharu dengan ketepatan 88.73%.

Oleh itu, memahami hubungan antara ciri-ciri yang mempengaruhi dalam mengesan aktiviti botnet adalah perlu untuk mengelakkan ciri-ciri yang dipilih berlebihan dalam pengesanan botnet. Penyelidik terdahulu hanya menggunakan ciri di dalam sistem tanpa menyebut ciri dipengaruhi dalam pengesanan botnet. Oleh itu, kajian ini akan mendedahkan ciri yang mempengaruhi pengesanan botnet. daripada trafik rangkaian untuk pengesanan tingkah laku mengganggu yang memudahkan keberkesanan Sistem Pengesanan Pencerobohan (IDS) mereka terhadap pelbagai jenis serangan rangkaian.

2.3 PENDEKATAN STATISTIKAL

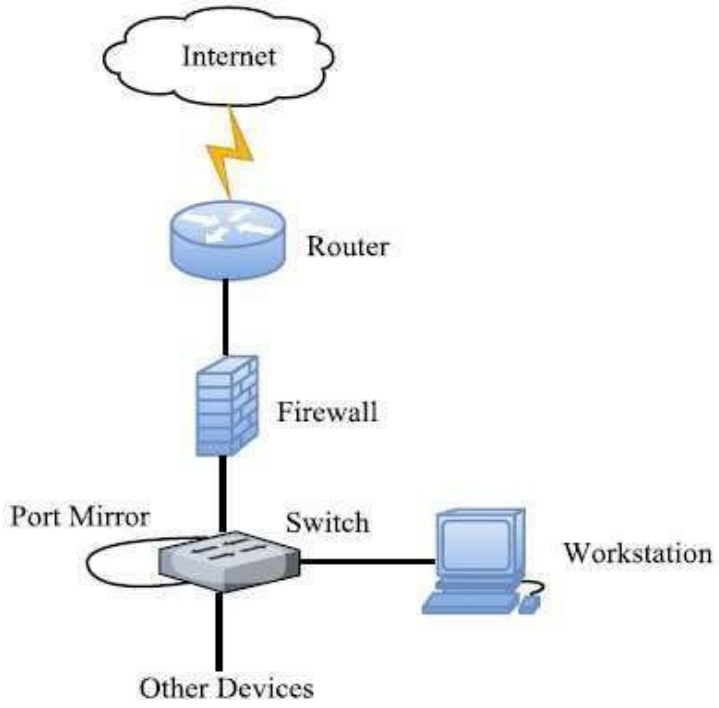
Menurut D.W. Hosmer dan S.Lemeshow,(2000) Kelebihan kaedah logistic regresi adalah mudah digunakan, fleksibiliti dan berkeupayaan digunakan dalam pelbagai bidang. Manakala Ott, R.L. dan Longnecker(2010) menyatakan kaedah statistik ialah satu set prinsip dan prosedur yang melibatkan pengumpulan data, rumusan data dan analisis statistik data berkaitan yang diperhatikan. Penyelidik sebelum ini menggunakan teknik ini dengan menggunakan nilai anggaran parameter untuk mengesan botnet. Kenyataan ini disokong oleh Ghanaei(2015) yang menyerlahkan pengiraan parameter statistik seperti maksimum atau minimum boleh digunakan untuk membezakan aktiviti anomali. M. S. Mok, S. Y. Sohn, and Y. H. Ju (2010) mencadangkan pengesanan rawak model logistic regresi untuk meramalkan pengesanan anomali. Penyelidikan berdasarkan sampel dataset KDD-99 dengan 42 pembolehubah yang mengandungi sambungan Normal dan Anomali. Enam ciri terpilih dengan lima

pemilihan pembolehubah input dilakukan terdiri daripada kategori jenis data diskret, berterusan dan Binari dan menghasilkan ketepatan klasifikasi dan peratusan pengesahan set data yang tinggi.

Tambahan pula, menurut Hughes.K and Qu.Y (2012) menggunakan regresi logistik untuk mengira kebarangkalian bahawa satu paket mengandungi Malware. Pendekatan ini adalah kaedah terbaik berbanding kaedah pengesanan Tandatangan semasa dan kaedah pengesanan Anomali. Ini kerana regresi logistik boleh menggantikan semua tandatangan yang berkaitan dengan satu keluarga Malware dengan ketepatan yang sama seperti pengesanan Tandatangan. Selain itu, regresi logistik adalah lebih tepat dengan kurang positif palsu daripada kaedah pengesanan Anomali. Oleh itu, pendekatan statistik adalah cara terbaik untuk mengesan botnet HTTP.

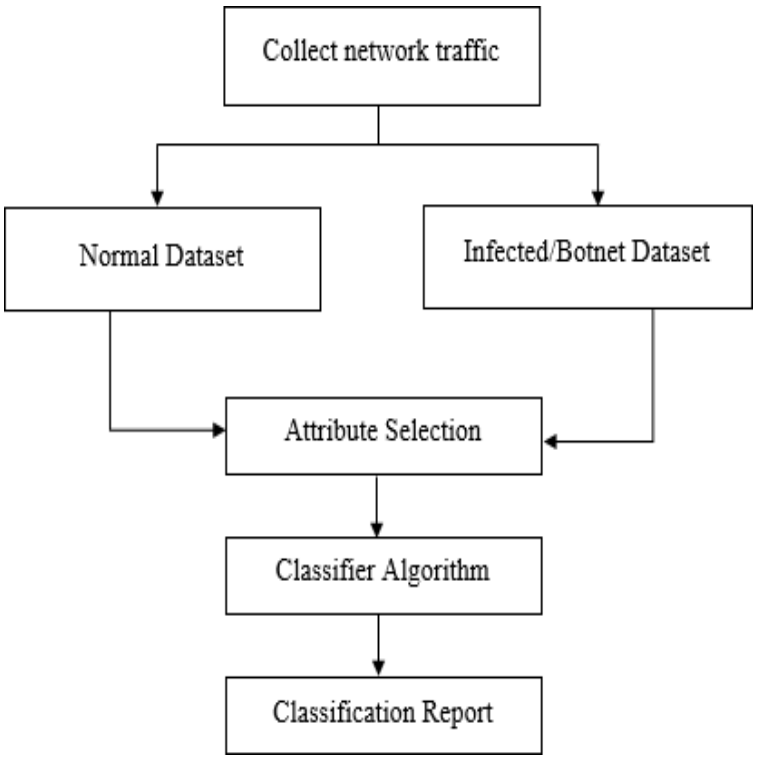
3 METODOLOGI

Untuk memahami cara mengesan botnet dalam rangkaian, testbed dilaksanakan. Trafik rangkaian dari dalaman ke rangkaian luaran akan ditangkap dengan menyambungkan satu hos yang beroperasi pada Linux ke port biasan. Satu port biasan atau port cermin telah dikonfigurasikan pada suis rangkaian untuk memantau trafik rangkaian pada antara muka tembok api. Kad rangkaian di dalam hos perlu dikonfigurasikan juga selepas mengkonfigurasi suis rangkaian. Kad rangkaian perlu ditetapkan sebagai mod rambang. Oleh itu, hos yang dipantau boleh menangkap sebarang aktiviti botnet di dalam trafik seperti yang ditunjukkan di Rajah 1. Sementara itu, proses menangkap botnet HTTP bermula dengan mengkonfigurasi persekitaran botnet HTTP. Kemudian, paket yang ditangkap akan disimpan dalam fail *tcpdump.gz. Fail itu akan diletakkan di mod istimewa root dalam folder dumpit untuk langkah seterusnya.



Rajah 1**.** Rekabentuk Rangkaian

Selain itu, proses pemilihan ciri boleh mendedahkan ciri yang mempengaruhi pengesanan botnet. Selepas pra-pemprosesan data dilakukan, data akan dianalisis dengan menggunakan algoritma pemilihan ciri. Rajah 2, menunjukkan proses pemilihan ciri. Daripada rajah tersebut, paket tangkapan yang melibatkan set data biasa dan botnet telah dikumpul dan melalui prapemprosesan data.



Rajah 2. Proses Pemilihan Ciri

Selepas itu, data (57 daripada ciri) akan dianalisis dengan menggunakan pemilihan ciri untuk memilih ciri yang boleh digunakan dalam pengelas pembelajaran mesin. Kemudian, set data akan dilatih dengan menggunakan tiga algoritma pengelas iaitu Naïve Bayes, Decision Tree dan Random Forest. Jadual 1, menunjukkan penerangan tentang algoritma pengelasan.

Jadual 1. Keputusan Ketepatan Pengelas Feizollah et al. (2014)

Algoritma Pengelasan	Deskripsi
Naïve Bayes	Berdasarkan peraturan Bayes kebarangkalian bersyarat. Ia menggunakan semua ciri yang terkandung dalam data, dan menganalisisnya secara individu seolah-olah ia adalah sama penting dan bebas antara satu sama lain.
Decision Tree	Model pembelajaran mesin ramalan yang memilih anggaran objektif sampel baharu berdasarkan beberapa nilai ciri data yang tersedia.
Random Forest	Teknik gabungan model pembelajaran yang menghasilkan pelbagai pembelajaran individu dan menjumlahkan hasilnya. Parameter terbaik pada setiap nod dalam pepohon pemutus (*decision tree)* dihasilkan menggunakan nombor yang dipilih dari komponen secara sembarangan.

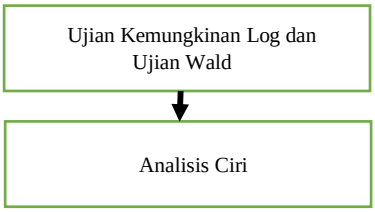
Perbandingan peratusan ketepatan antara ketiga-tiga pengelas ini telah dilakukan bagi mengenal pasti hasil yang terbaik dan sesuai yang boleh digunakan untuk fasa seterusnya. Jadual 2, menunjukkan keputusan pengelas dengan operator yang berbeza. Daripada rajah tersebut, didapati bahawa operator optimize dan operator Forward memberikan prestasi yang lebih baik dalam ketepatan pengesanan. Seperti yang dinyatakan sebelum ini, penyelidik memutuskan untuk memilih pemilihan hadapan kerana ketepatan hasil pengesanannya. Akhir sekali, klasifikasi akan dijana dengan ciri terpilih yang terlibat untuk pengesanan botnet.

Selain itu, ujian Nisbah Kemungkinan dan ujian Wald adalah ujian yang memberikan sumbangan sokongan kepada model dalam mengesan serangan botnet. Ciri yang mempengaruhi pengesanan botnet boleh didedahkan dengan meneroka sumbangan penting daripada ciri. Apabila hasil ciri yang dipilih memberi pengaruh yang baik kepada model maka, ciri tersebut boleh digunakan. Di samping itu, penyelidik lain yang menumpukan pada pengesanan botnet boleh merujuk hasil pengaruh ciri dalam kertas ini untuk kajian akan datang. Rajah 3 menunjukkan proses pengaruh ciri.

Jadual 2. Keputusan ketepatan pengelas

Operator	Pengelasan		
	Naïve Bayes	Random Forest	Decision Tree
Forward	91.08%	91.65%	91.90%
Optimize	91.08%	91.65%	91.90%
Backward	16.35%	91.12%	91.92%

Setiap ciri yang dipilih (7 daripada ciri) akan diuji dengan menggunakan ujian Nisbah Kemungkinan dan model ujian Wald. Selepas itu, semua ciri terpengaruh akan dianalisis bagi membuktikan sama ada model tersebut memberi impak yang baik kepada model atau tidak dalam menjangkakan hasilnya.



Rajah 3**.** Proses Pengaruh Ciri

Merujuk kepada penulis Field .A (2009) Ujian Nisbah Kemungkinan (1) dan Ujian Wald (2) adalah dua teknik yang melibatkan dalam menilai sumbangan ciri tersebut. Kedua-dua ujian ini boleh digunakan untuk menentukan sama ada ciri-ciri mempunyai ramalan yang baik terhadap keputusan atau menentukan hasil pembolehubah.

(i) Ujian Kemungkinan Log

Ujian Nisbah Kemungkinan ialah perbandingan antara model dengan dan tanpa peramal tertentu. Peramal model memberikan kesan yang baik dalam meramalkan keputusan apabila nilai pengurangan model nisbah kemungkinan tanpa peramal jika peramal dimasukkan ke dalam model. Persamaan ujian nisbah kemungkinan log ialah:

x2 = 2 [Kemungkinan Log(Baharu(dengan peramal) – Kemungkinan Log(Garis Asas(tanpa peramal)) (1)

(ii) Ujian Wald

Ujian Wald mempunyai taburan khas yang dikenali sebagai taburan khi kuasa dua yang digunakan untuk menganggar akibat statistik bagi setiap pekali b dalam model. Menurut Field, A(2009) Ujian Wald memberitahu kita sama ada pekali b untuk peramal itu berbeza dengan ketara daripada sifar. Penunjuk membuat komitmen yang besar terhadap jangkaan keputusan apabila pekali sama sekali tidak sama dengan sifar dengan menggunakan syarat (2).

$$Wald = b / SE_b \quad (2)$$

Dengan cara ini, jika anggaran ujian Wald di dalam model adalah lebih besar, maka penunjuk memberikan komitmen yang signifikan kepada model dalam menjangkakan keputusan.

4 DAPATAN KAJIAN

Dalam kertas kerja ini, hanya 7 ciri dipilih daripada 57 ciri yang boleh digunakan untuk mengesan aktiviti botnet dalam rangkaian seperti ditunjukkan dalam Jadual 3 di bawah. Terdapat enam ciri yang dipilih selepas melalui proses pemilihan ciri iaitu: avg_segm_size_b2a, initial_window_bytes_a2b, unique_bytes_sent_b2a, max_win_adv_b2a, max_win_adv_a2b, min_segm_size_a2b dan max_segm_size_a2b. Kesemua ciri ini diuji dan dianalisis menggunakan regresi logistik binomial dalam Statistik SPSS. Tujuan untuk menjalankan ciri tersebut menggunakan pendekatan statistik adalah untuk meneroka ciri pengaruh dan mungkin berguna untuk meningkatkan sistem pengesanan terutamanya serangan botnet HTTP. Kesemua tujuh ciri ini mempunyai ciri tersendiri yang penting dalam mengesan botnet HTTP.

Jadual 3. Menentukan Ciri Terpengaruh

Ciri/Fitur	Deskripsi
avg_segm_size_b2a	Saiz segmen biasa yang dikesan sepanjang hayat hubungan yang dikira sebagai nilai yang dilaporkan dalam medan bait data sebenar dibahagikan dengan paket data sebenar yang dilaporkan.
initial_window_bytes_a2b	Jumlah keseluruhan bait dihantar dalam tetingkap pertama
unique_bytes_sent_b2a	Kuantiti terhad bait dihantar.
max_win_adv_b2a & max_win_adv_a2b	Bilangan iklan tetingkap tinggi yang telah ditemui. Apabila kedua-dua pihak merundingkan penskalaan tetingkap, maka ia adalah tetingkap iklan berskala maksimum yang dapat dilihat.
min_segm_size_a2b	Bilangan terkecil saiz segmen yang dikesan sepanjang hayat sambungan.
max_segm_size_a2b	Bilangan saiz segmen yang lebih tinggi dikesan sepanjang hayat sambungan.

Selain itu, mendedahkan pengaruh ciri dan tujuan di sebalik pengumpulan ciri adalah peluang yang baik kerana kebanyakan penyelidik terdahulu tidak memberi tumpuan mengenainya. Sebab di sebalik mendedahkan ciri yang mempengaruhi boleh membantu menilai sumbangan penting ciri yang digunakan untuk mengesan botnet. Tambahan pula, dengan memahaminya, ia mungkin membantu untuk memberi pengetahuan tentang hubungan ciri dalam menyumbang peranan untuk mengesan aktiviti botnet. Hasilnya dibincangkan berdasarkan nilai statistik daripada ujian nisbah kemungkinan dan ujian Wald.

4.1 Ciri Avg_segm_size_b2a

Ciri ini telah dipilih untuk mengesan aktiviti botnet dalam rangkaian. Dengan menggunakan nilai statistik merujuk kepada persamaan ujian Kemungkinan Log. Pada ketika hanya konsisten dimasukkan, -2LL = 404980.373, walau bagaimanapun avg_segm_size_b2a telah dimasukkan ini telah diturunkan kepada 402052.506. Ciri-ciri mempunyai kesan kritikal dalam menjangkakan hasil (botnet) berdasarkan nilai pengurangan nisbah kemungkinan.

Jadual 4. Ringkasan Avg_segm_size_b2a

-2 Log Likelihood	Wald
402052.506	2173.349

Jadual 4 juga menunjukkan nilai Wald ialah 2173.349. Menurut Ott et al.(2010) jika nilainya berbeza daripada sifar, ia menunjukkan penunjuk memberi impak yang baik kepada model dalam menjangkakan hasilnya. Ciri yang dipilih memberi kesan yang baik kepada model dalam menjangkakan hasil kerana nilai hasil adalah berbeza dengan ketara daripada sifar.

4.2 Ciri Initial_window_bytes_a2b

Initial_window_bytes_a2b juga memberikan pengaruh yang signifikan dalam pengesanan botnet. Jadual 5 menunjukkan nilai statistik nisbah kemungkinan selepas elemen dimasukkan ke dalam model. Sementara itu, apabila hanya konsisten dimasukkan, -2LL = 402052.506 dan ini telah menurun kepada 400666.766. Pengurangan ini menunjukkan bahawa ciri mempunyai kesan kritikal dalam menjangkakan hasil (botnet).

Jadual 5. Ringkasan Initial_window_bytes_a2b

-2 Log Likelihood	Wald
400666.766	7445.696

Nilai Wald untuk model baharu, iaitu 7445.696 seperti yang ditunjukkan di atas. Oleh kerana, keputusan yang berbeza dengan ketara daripada sifar bermakna ciri yang dipilih memberi kesan yang baik kepada model dalam menjangkakan hasilnya.

4.3 Ciri Unique_bytes_sent_b2a

Jadual 6 menunjukkan nilai statistik nisbah kemungkinan selepas elemen dimasukkan ke dalam model. Dengan cara ini, nilai model garis dasar (tanpa peramal) boleh dikira. Pada titik apabila hanya konsisten dimasukkan -2LL = 400666.766 dan nilai ini telah menurun kepada 400651.702. Pengurangan ini menunjukkan bahawa unsur-unsur mempunyai kesan kritikal pada jangkaan keputusan (botnet). Manakala nilai ujian Wald bagi ciri ini ialah 0.322.

Jadual 6. Ringkasan Unique_bytes_sent_b2a

-2 Log Likelihood	Wald
400651.702	0.322

4.4 Ciri Max_win_adv_a2b

Hasil daripada analisis menunjukkan bahawa max_win_adv_a2b memberi kesan yang baik kepada model dalam menjangkakan pengesanan botnet. Jadual 7 menunjukkan nilai statistik nisbah kemungkinan selepas elemen dimasukkan ke dalam model. Apabila hanya konsisten dimasukkan, -2LL = 400651.702 dan nilai ini telah diturunkan kepada 398849.06. Pengurangan ini menunjukkan bahawa unsur-unsur mempunyai kesan kritikal pada jangkaan keputusan (botnet). Jadual 7 juga menunjukkan nilai Wald ialah 0.854.

Jadual 7. Ringkasan Max_win_adv_a2b

-2 Log Likelihood	Wald
398849.06	0.854

a.

4.5 Ciri Min_segm_size_a2b

Jadual 8 menunjukkan anggaran statistik nisbah kemungkinan selepas elemen dimasukkan ke dalam model. Dengan cara ini, anggaran model garis dasar (tanpa penunjuk) boleh dikira. Pada ketika hanya konsisten dimasukkan, -2LL = 398849.06 dan nilai ini telah menurun kepada 378687.144. Pengurangan ini menunjukkan bahawa unsur-unsur mempunyai kesan kritikal pada jangkaan keputusan (botnet).

Jadual 8**.** Ringkasan Min_segm_size_a2b

-2 Log Likelihood	Wald
378687.144	13961.988

Nilai Wald bagi model baharu iaitu 13961.988 seperti yang ditunjukkan dalam Jadual 8. Oleh kerana, keputusannya adalah berbeza secara signifikan daripada sifar yang bermakna ciri yang dipilih memberi kesan yang baik kepada model dalam menjangkakan keputusan.

4.6 Ciri Max_segm_size_a2b

Max_segm_size_a2b juga memberi pengaruh yang signifikan dalam pengesanan botnet dan pengesahan dibuat dengan menggunakan ujian yang sama dengan ciri dipengaruhi yang lain. Jadual 9 menunjukkan nilai statistik nisbah kemungkinan selepas elemen dimasukkan ke dalam model. Apabila hanya konsisten dimasukkan, -2LL = 378687.144 dan nilai ini telah diturunkan kepada 378445.002. Pengurangan ini menunjukkan bahawa ciri mempunyai kesan kritikal dalam menjangkakan hasil (botnet). Manakala nilai ujian Wald ialah 0.124.

Jadual 9. Ringkasan Max_segm_size_a2b

-2 Log Likelihood	Wald
378445.002	0.124

4.7 Ciri Max_win_adv_b2a

Keputusan daripada analisis menunjukkan bahawa max_win_adv_b2a memberi kesan yang baik kepada model dalam menjangkakan pengesanan botnet. Jadual 10 menunjukkan nilai statistik nisbah kemungkinan selepas elemen dimasukkan ke dalam model. Apabila hanya konsisten dimasukkan, -2LL = 378445.002 dan nilai ini telah menurun kepada 377280.474. Pengurangan ini menunjukkan bahawa unsur-unsur mempunyai kesan kritikal pada jangkaan keputusan (botnet). Jadual 10 juga menunjukkan nilai Wald ialah 0.354.

Jadual 10. Ringkasan Max_win_adv_b2a

-2 Log Likelihood	Wald
377280.474	0.354

Daripada perbincangan di atas merumuskan bahawa daripada tujuh ciri yang dipilih, hanya tiga ciri yang memberi kesan yang baik kepada model dalam menjangkakan hasilnya. Ciri dipengaruhi ialah avg_segm_size_b2a, initial_window_bytes_a2b dan min_segm_size_a2b dengan nilai ujian Wald berbeza daripada sifar, iaitu 2173.349, 7445.696 dan 13961.988.Ketiga-tiga ciri ini boleh digunakan untuk mengenal pasti pemilihan ambang.

6. PENUTUP DAN KERJA MASA DEPAN

Memilih ciri penting adalah penting kerana ia memberi sumbangan dari segi ketepatan pengesanan. Kebanyakan pengkaji hanya tertumpu kepada kaedah pengecaman dan bukannya mendedahkan sebab di sebalik pemilihan. Penyelidik terdahulu hanya menggunakan ciri di dalam

sistem tanpa menyebut ciri dipengaruhi dalam pengesanan botnet. Selain itu, untuk mengelakkan ciri bertindan, memahami hubungan antara ciri dipengaruhi boleh mengurangkan potensi memilih ciri yang tidak diperlukan yang mungkin memberi kesan dalam mengesan aktiviti botnet. Hasil analisis menunjukkan ciri yang dijana memberikan sumbangan yang baik dalam pengesanan botnet HTTP dengan kadar pengesanan yang lebih tinggi, sekali gus memenuhi objektif penyelidikan. Berdasarkan ketepatan hasil pengesanan dalam penyelidikan ini, kajian lanjut diperlukan untuk mengenal pasti nilai ambang yang sesuai. Penilaian yang baik pada kadar pengesanan tidak bermakna ia boleh digunakan dalam mana-mana teknik. Ia masih memerlukan penambahbaikan dalam membangunkan teknik yang lebih baik untuk mengenal pasti nilai ambang untuk pengesanan Botnet HTTP itu sendiri untuk meningkatkan kadar pengesanan dan mungkin digunakan dalam teknik yang berbeza. Memandangkan kajian ini hanya memfokuskan pada protokol TCP, niat untuk menggunakan protokol lain seperti UDP juga disyorkan. Selain itu, kajian masa depan juga bertujuan untuk melihat log IDS, untuk menentukan ambang yang sesuai dalam trafik yang boleh menyumbang kepada ketepatan pengesanan IDS untuk membezakan aktiviti normal dan abnormal dalam rangkaian. Memilih ciri penting juga bukanlah tugas yang mudah dicapai. Data yang tidak mencukupi dan tidak relevan ciri akan menjejaskan pemilihan ciri dalam membezakan aktiviti botnet. Pada masa kini, jenis serangan botnet telah berlaku perubahan yang ketara dan sukar untuk dikenal pasti sebagai botnet HTTP menyembunyikan komunikasi mereka melalui trafik HTTP. Ini kekal sebagai cabaran terbuka dalam komuniti penyelidikan. Lebih-lebih lagi, kekangan pada pengesanan botnet iaitu tidak boleh membezakan dan mengenali aktiviti botnet baharu yang perlu ditambahbaik.

7. PENGHARGAAN

Kajian ini telah disokong oleh penyelidik Universiti Teknikal Malaysia Melaka dan Kolej Komuniti Masjid Tanah, Kementerian Pengajian Tinggi Malaysia. Penulis ingin mengucapkan terima kasih kepada Kumpulan Penyelidikan INSFORNET atas sokongan mereka yang luar biasa dengan menyediakan kemudahan untuk menjalankan kajian.

RUJUKAN

Abedinia, O., Amjady, N. and Zareipour, H., “A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems,” IEEE Transactions on Power Systems, Vol. 32, Issues 1, pp.62-74, 2017.

Ahsen, M.E., Boren, T.P., Singh, N.K., Misganaw, B.,Mutch, D.G., Moore, K.N., Backes, F.J., McCourt, C.K., Lea, J.S., Miller, D.S. and White, M.A., “Sparse feature selection for classification and prediction of metastasis in endometrial cancer,” BMC genomics, 18(3), p. 233, 2017.

B. Claise, 2008. “Specification of the IP flow information export (IPFIX) protocol for the exchange of IP traffic flow information”.Retrieved from https://tools.ietf.org/html/rfc5101 [Accessed on March 8, 2017].

Begum, S., Bera, S.P., Chakraborty, D. and Sarkar, R., “Breast cancer detection using feature selection and active learning,” In Computer, Communication and Electrical Technology, pp. 43-48, CRC Press, 2017.

Bolon-Canedo, V., Sanchez-Marono, N. and Alonso- Betanzos, A., “Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset,” Expert Systems with Applications, Vol. 38, No. 5, pp. 5947- 5957, 2011.

Cai, T. and Zou, F., “Detecting HTTP botnet with clustering network traffic,” In Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on Shanghai, China , pp. 1-7, IEEE, September, 2012.

Chen, C.M., Ou, Y.H. and Tsai, Y.C., “Web botnet detection based on flow information,” In Computer Symposium (ICS), 2010 International on Tainan, Taiwan, Taiwan, pp. 381-384, IEEE, December 2010.

Eric Auchard, 2016. “German internet outage was failed botnetattempt:report”. Retrieved from http://www.reuters.com/article/us-deutsche-telekom-outages- dUSKBN13N12K [Accessed on February 13, 2017].

E. Popoola, A. Adewumi, “Efficient feature selection technique for network intrusion detection system using discrete differential evolution and decision tree,” International Journal of Network Security, Vol.19, No.5, pp. 660-669, Sept. 2017.

Feizollah, A., Anuar, N.B., Salleh, R., Amalina, F., Ma’arof,R.U.R. and Shamshirband, S., “A study of machine learning classifier for anomaly-based mobile botnet detection” Malaysian Journal of Computer Science, 26(4), 2014.

Field, A, 2009. Logistic regression. Discovering statistics using SPSS, pp 264-315.Hughes, K. and Qu, Y., “A theoretical model: Using logistic regression for malware signature based detection,” In the 10th International Conference on Dependable, Autonomic, and Secure Computing (DASC-2012), 2012.

J.Jabez and Muthukumar, “Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach,” Procedia Computer Science 48 (2015) 338 – 346

Khattak, S., Ramay, N.R., Khan, K.R., Syed, A.A. and Khayam, S.A., “A taxonomy of botnet behavior, detection, and defense,” IEEE Communications Surveys & Tutorials, Volume 16, No. 2, pp. 898-924, 2014.

Manikandan, Gopi, Abirami, S. “ Feature Selection Is Important: State-of-the-Art Methods and Application Domains of Feature Selection on High-Dimensional Data” 177- 196, 2021

M. C. E. R. T. MYCERT, “Incidents Report of General Incident Classification Statistic 2022,” 2022. [Online].Available:https://www.mycert.org.my/portal/statistics-content?menu=b75e037d-6ee3-4d11-8169-66677d694932&id=574bf33f-7291-4b6e-bb61-9adcf6a6259c . [Accessed: 26-August-2022].

M. S. Mok, S. Y. Sohn, and Y. H. Ju, “Randomeffects logistic regression model for anomalydetection,” Expert Syst. Appl., vol. 37, no. 10, pp.7162–7166, 2010.

Ott, R.L. and Longnecker, M.T., 2010. An introduction to statistical methods and data analysis. Cengage Learning.

Saad, S., Traore, I., Ghorbani, A., Sayed, B., Zhao, D., Lu, W., Felix, J. and Hakimian, P., “Detecting P2P botnets through network behavior analysis and machine learning,” In Privacy, Security and Trust (PST), 2011 Ninth Annual International Conference on Montreal, QC, Canada, pp. 174- 180, IEEE, July, 2011.

Stevanovic, M. and Pedersen, J. M., “An efficient flow-based botnet detection using supervised machine learning,” In Computing, Networking and Communications (ICNC), 2014 International Conference on Honolulu, HI, USA, pp. 797- 801, IEEE, February, 2014.

V. Jyothsna, V. V. Rama Prasad and K.Munivara Prasad “A Review of Anomaly based Intrusion Detection Systems “International Journal of Computer Applications (0975 – 8887), 2011

Warwick Ashford, 2017. Lloyds Bank hit by massive DDoS attack. Retrieved from http://www.computerweekly.com/news/450411443/Lloyds-Bank-hit-by-massive-DDoS-attack [Accessed on March 8, 2017].

Zhao, D., Traore, I., Sayed, B., Lu, W., Saad, S., Ghorbani,A. and Garant, D., “Botnet detection based on traffic behavior analysis and flow intervals,” Computers & Security, 39, pp. 2-16, 2013.