

**PENGENALAN POLA TRAFIK DATA TOR BROWSER
MENGGUNAKAN CLUSTERING K-MEANS**

TUGAS AKHIR



Oleh :

**DONI SAPUTRA
09011181520120**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020**

**PENGENALAN POLA TRAFIK DATA TOR BROWSER
MENGGUNAKAN CLUSTERING K-MEANS**

TUGAS AKHIR

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**



Oleh :

**DONI SAPUTRA
09011181520120**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN

PENGENALAN POLA TRAFIK DATA TOR BROWSER MENGGUNAKAN CLUSTERING K-MEANS

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

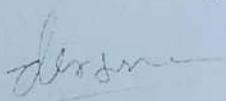
OLEH :

**Doni Saputra
09011181520120**

Indralaya, 31 Desember 2020

Pembimbing I

Pembimbing II


Deris Stiawan, M.T., Ph.D.
NIP 197806172006041002


Ahmad Heryanto, S.Kom., M.T.
NIP 198701222015041002

Mengetahui,
Ketua Jurusan Sistem Komputer


Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERSETUJUAN

Pada hari Kamis 31 Desember 2020 telah dilaksanakan ujian sidang tugas akhir oleh Sarjana Ilmu Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Doni Saputra

Nim : 09011181520120

Judul : Pengenalan Pola Trafik Data *TOR* Browser Menggunakan
Clustering K-Means

Tim Pengaji :

1. Ketua

Rendyansyah, S.Kom., M.T.



(.....)



(.....)



2. Pengaji

Huda Ubaya, M.T.

Mengetahui,
Ketua Jurusan Sistem Komputer



Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Doni Saputra
NIM : 09011181520120
Jurusan : Sistem Komputer
Judul : Pengenalan Pola Trafik Data *TOR* Browser Menggunakan
Clustering K-Means

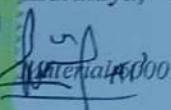
Hasil Pengecekan Software iTThenticate/Turnitin: 9 %

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil *penjiplakan/plagiat*. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, Desember 2020


Indra 14/12/2020

Doni Saputra

NIM. 09011181520120

HALAMAN PERSEMBAHAN

"Menaklukkan ribuan manusia mungkin disebut pemenang, tapi bisa menaklukkan diri sendiri disebut penakluk yang brilian" – Soekarno

Tugas Akhir ini saya persembahkan untuk :

- *Kedua Orang tua dan Adik saya*
- *Dosen Pembimbing dan Penguji*
- *Calon Istri saya*
- *Sahabat – sahabat saya*
- *Teman Seperjuangan Sistem Komputer 2015*
- *Almamaterku*

KATA PENGANTAR



Alhamdulillahirabbil'alamin Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir ini dengan judul “Pengenalan Pola *Trafik Data TOR Browser Menggunakan Clustering K-Means*” di susun untuk memenuhi sebagian persyaratan kelulusan untuk memperoleh gelar Sarjana Komputer pada Jurusan Sistem Komputer Universitas Sriwijaya.

Pada kesempatan ini penulis menyadari keterbatasan dan kelemahan yang ada dalam menyelesaikan tesis ini sehingga penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini, kepada:

1. Allah SWT, yang telah memberikan rahmat dan karunia-Nya sehingga tugas akhir penulisan ini dapat berjalan dengan lancar.
2. Kedua orang tua beserta keluarga yang selalu mendoakan serta memberikan motivasi dan semanga.
3. Bapak Jaidan Jauhari, S.Pd, M.T. selaku dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. Ir. H. Sukemi, M.T. sebagai Ketua Jurusan Sistem Komputer Universitas Sriwijaya.
5. Bapak Deris Stiawan, M.T., Ph.D. selaku pembimbing I yang telah meluangkan waktu, bantuan serta saran dan kritiknya dalam penyusunan tugas akhir ini.

6. Bapak Ahmad Heryanto, M.T. selaku pembimbing II yang telah meluangkan waktu, bantuan serta saran dan kritiknya dalam penyusunan tugas akhir ini.
7. Dosen-dosen pengajar yang telah memberikan ilmu bermanfaat kepada penulis selama menuntut ilmu di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Mba Winda Kurnia Sari dan Mba Renny Virgasari selaku Administrator Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberi kemudahan dalam pengurusan administrasi.
9. Seluruh teman-teman Jurusan Sistem Komputer Angkatan 2015 yang telah membantu dan memberikan semangat pada masa-masa perkuliahan.
10. Semua pihak yang telah memberi dukungan kepada penulis dan tidak bisa disebutkan satu-persatu.

Akhir kata, penulis menyadari bahwa tugas akhir ini masih banyak kekurangan baik dari isi maupun susunan. Semoga tugas akhir ini dapat bermanfaat untuk kita semua.

Indralaya, Desember 2020

Penulis

PENGENALAN POLA TRAFIK DATA TOR BROWSER MENGGUNAKAN CLUSTERING K-MEANS

Doni Saputra (09011181520120)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : d0ni270497@gmail.com

Abstrak

Saat ini anonimitas dan privasi, ada beberapa jenis dan implementasi layanan anonim yang tersedia di Internet. Tor adalah salah satu layanan di antara layanan tersebut. Aktivis, jurnalis dan penulis menggunakan alat ini untuk kebebasan berbicara, tetapi juga digunakan secara salah. misalnya ilegal peretasan dan serangan, menyebarkan malware atau scam, pornografi situs, narkotika dan transaksi ilegal lainnya, dan masih banyak lagi. Tujuan dari penelitian ini adalah melakukan ekstraksi paket trafik pada *Dataset ISCX*, training data untuk menemukan atribut paket data yang digunakan dalam pengenalan pola, mengenali pola secara manual serta menerapkan metode *Clustering K-Means* untuk mengenali pola *Dataset ISCX*. Penelitian ini menggunakan dataset dari *ISCX*, tools yang digunakan dalam pengambilan dataset *ISCX* adalah *Whonix* (<https://www.whonix.org>), OS Linux yang siap pakai dan dikonfigurasi untuk merutekan semua lalu lintas melalui jaringan Tor. Proses analisa dan akuisisi data menggunakan *tools Wireshark* bertujuan untuk keperluan proses pengenalan pola berupa atribut data atau informasi mengenai *Log Activity* dan *Protokol* dari trafik *Dataset TOR*. Pada Penelitian ini berhasil mengenali pola dari *Dataset TOR*, baik secara manual maupun dengan penerapan Clustering K-Means yaitu pada atribut protokol trafik tor dan normal dengan perbandingan jumlah protokol *TLS* trafik tor lebih besar dari trafik normal.

Kata Kunci: *TOR (The Onion Router) Network, TLS (Transport Layer Security), Clustering K-Means.*

INTRODUCTION TO TOR BROWSER DATA TRAFFIC PATTERNS USING K-MEANS CLUSTERING

Doni Saputra (09011181520120)

Computer Engineering Department, Computer Science Faculty,

Sriwijaya University

Email : d0ni270497@gmail.com

Abstract

Nowadays anonymity and privacy, there are several types and implementations of anonymous services available on the Internet. Tor is one of those services. Activists, journalists and writers use this tool for free speech, but it is also used incorrectly. for example illegal hacks and attacks, spreading malware or scams, pornographic sites, narcotics and other illegal transactions, and many more. The purpose of this research is to extract traffic packets on the *ISCX Dataset*, train data to find the attributes of the data packets used in pattern recognition, recognize patterns manually and apply the *K-Means Clustering* method to recognize *ISCX Dataset*. This study uses a dataset from ISCX, the tools used in retrieving the *ISCX Dataset* are *Whonix* (<https://www.whonix.org>), a Linux OS that is ready to use and configured to route all traffic through the Tor network. The process of analyzing and data acquisition using the *Wireshark* tool aims for the purposes of the pattern recognition process in the form of data attributes or information about the *Log Activity* and *Protocols* of the *TOR Dataset* traffic. In this study, it was successful in recognizing the pattern of the *TOR Dataset*, both manually and with the application of *K-Means Clustering*, namely the attributes of the tor and normal traffic protocols with the ratio of the number of TLS protocol traffic tor greater than normal traffic.

Keywords: *TOR (The Onion Router) Network, TLS (Transport Layer Security), Clustering K-Means.*

DAFTAR ISI

HALAMAN JUDULi
HALAMAN PENGESAHANii
HALAMAN PERSETUJUANiii
HALAMAN PERNYATAANiv
HALAMAN PERSEMBAHANv
KATA PENGANTARvi
ABSTRAKvii
ABSTRACTviii
DAFTAR ISIix
DAFTAR GAMBARx
DAFTAR TABELxi

BAB I. PENDAHULUAN

1.1. Latar Belakang	1
1.2. Tujuan	2
1.3. Manfaat	2
1.4. Rumusan Masalah	2
1.5. Batasan Masalah	2
1.6. Metodologi Penelitian	3
1.7. Sistematika Penelitian	5

BAB II. TINJAUAN PUSTAKA

2.1. Diagram Konsep Penelitian	6
2.2. TOR	7
2.3. Intrusion Detection System (IDS)	9
2.4. Clustering K-Means	10
2.5. Dataset ISCX	11

BAB III. METODOLOGI PENELITIAN

3.1. Pendahuluan	12
------------------------	----

3.2. Kerangka Kerja Penelitian	12
3.3. Perancangan Sistem	14
3.4. Skenario Dataset ISCX	14
3.5. Ekstraksi Data	16
3.6. Training Data	18

BAB IV. HASIL DAN PEMBAHASAN

4.1. Pendahuluan	19
4.2. Analisa Dataset ISCX	19
4.3. Training Data	22
4.4. Pengenalan Pola	23
4.5. Pengujian Clustering K-Means	25
4.6. Pengujian dengan Tools Rapid Miner.....	29
4.7. Visual/ Plot Hasil Clustering K-Means dan Rapid Miner.....	30

BAB V. KESIMPULAN DAN SARAN	32
--	-----------

DAFTAR PUSTAKA	34
-----------------------------	-----------

LAMPIRAN

DAFTAR GAMBAR

	Halaman
Gambar 1.1 Diagram Air Metodologi Penelitian	4
Gambar 2.1 Diagram Konsep Penelitian.....	6
Gambar 2.2 Sirkuit TOR	7
Gambar 2.3 Proses TLS Handshake	8
Gambar 3.1 Flowchart Kerangka Kerja Tugas Akhir	13
Gambar 3.2 Skenario Dataset TOR.....	15
Gambar 4.1 Korelasi hasil ekstraksi.....	22
Gambar 4.2 Sampel Percobaan Browsing	24
Gambar 4.3 Hasil Normalisasi	25
Gambar 4.4 Hasil Inisialisasi.....	26
Gambar 4.5 Hasil Program Clustering K-Means	29
Gambar 4.6 Rangkaian Blok Rapid Miner	29
Gambar 4.7 Hasil Dengan Tools Rapid Miner	29
Gambar 4.8 Visual/ Plot Hasil Program Clustering K-Means.....	30
Gambar 4.9 Visual/ Plot Hasil Tools Rapid Miner.....	31

DAFTAR TABEL

	Halaman
Tabel 2.1	Kategori Dataset
Tabel 3.1	Kebutuhan Perangkat Lunak
Tabel 3.2	Atribut Ekstraksi Data.....
Tabel 4.1	Jumlah Paket <i>Trafik Normal</i> pada Dataset.....
Tabel 4.2	Jumlah Paket <i>Trafik TOR</i> pada Dataset.....
Tabel 4.3	Pola <i>Trafik TOR</i> dan <i>Trafik NORMAL</i>

BAB I. PENDAHULUAN

1.1. Latar Belakang

Saat ini anonimitas dan privasi adalah perhatian utama pengguna Internet. Ada beberapa jenis dan implementasi layanan anonim yang tersedia di Internet. Tor adalah salah satu layanan di antara layanan tersebut. Jaringan Tor didasarkan pada konsep perutean onion routing, dan sekarang sangat populer. Anehnya, sangat sedikit penelitian yang telah dilakukan pada jaringan anonimisasi tersebut. Aktivis, jurnalis dan penulis menggunakan alat ini untuk kebebasan berbicara, tetapi juga digunakan secara salah oleh malware, mendistribusikan serangan penolakan layanan, layanan tersembunyi yang menjual barang ilegal, spam, dan banyak lagi [1].

Namun di sisi lain, internet juga bisa disalahgunakan untuk melakukan aktivitas negatif atau ilegal, misalnya ilegal peretasan dan serangan, menyebarkan malware atau scam, pornografi situs, narkotika dan transaksi ilegal lainnya, dan masih banyak lagi [8]. Ada sejumlah besar literatur yang menunjukkan bahwa penyerang menggunakan jaringan Tor untuk melakukan aktivitas ilegal [9].

Dalam penelitian [1], menjelaskan penggunaan Tor deteksi dengan menganalisis koneksi TLS yang digunakan untuk membuat koneksi aman dan menggunakan karakteristik yang ditemukan selama analisis untuk mendeteksi dan memblokir lalu lintas Tor yang berasal dari browser Tor.

Pada penelitian [2], Clustering k-means dapat digunakan untuk mendeteksi paket serangan dan paket normal. Clustering merupakan teknik untuk mengklasifikasi data yang tidak diketahui ke dalam satu grup untuk eksplorasi dan analisis data. Tujuan utama dari clustering meliputi antara lain memperoleh informasi dari data (deteksi anomali, identifikasi feature), classification data and compressing data.

Berdasarkan uraian diatas penulis bermaksud melakukan penelitian untuk mengenali pola trafik data *TOR* menggunakan metode *K-Means Clustering*.

1.2. Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Melakukan ekstraksi paket trafik pada dataset ISCX.
2. Melakukan training data untuk menemukan atribut paket data yang digunakan dalam pengenalan pola.
3. Mengenali pola trafik dataset ISCX..
4. Menerapkan metode Clustering K-Means untuk mengenali pola dataset ISCX.

1.3. Manfaat

Adapun manfaat yang dapat diambil dari penelitian ini adalah:

1. Memberikan kemudahan dalam mengenali pola trafik data *TOR* dan trafik data normal.
2. Dapat membedakan pola trafik data *TOR* dan trafik data normal.

1.4. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan sebelumnya, maka didapatkan perumusan masalah yaitu :

1. Bagaimana menentukan atribut trafik data *TOR* dan trafik data normal dalam dataset *ISCX* ?
2. Bagaimana mengenali pola trafik data *TOR* dan trafik data normal menggunakan Clustering K-Means?

1.5. Batasan Masalah

Selain perumusan masalah diatas, juga terdapat batasan masalah pada tugas akhir ini, antara lain :

1. Menggunakan dataset ISCX.
2. Tidak membahas bagaimana cara pencegahan trafik data *TOR* tersebut.
3. Pengenalan pola trafik *TOR* dan trafik normal tidak diujikan pada lalu lintas jaringan real-time.
4. Mekanisme yang digunakan untuk mengenali pola adalah dengan algoritma

clustering k-means.

1.6. Metodologi Penelitian

Metodologi yang digunakan dalam tugas akhir ini akan melewati beberapa tahapan sebagai berikut :

1. Pengumpulan Data

Metode pengumpulan data yang dilakukan adalah studi pustaka atau literatur.

Studi *literature* dilakukan dengan cara mempelajari dan mengumpulkan informasi mengenai penelitian yang akan dilakukan. *Literature* tersebut diperoleh dari jurnal, buku dan *mailing list* agar dapat menunjang metedologi dan pendekatan yang akan diterapkan pada penelitian.

2. Pengolahan Data

Tahap ini ialah membahas mengenai proses yang telah dilakukan dalam penelitian kedalam bentuk tulisan. Pengolahan data dimaksudkan untuk melihat kesesuaian hasil penelitian serta mengevaluasi jalannya sistem berdasarkan batasan masalah dari penelitian

3. Pengujian

Tahap ini merupakan tahap pengujian metodologi penelitian dan penelitian sebelumnya sehingga didapatkan data hasil uji yang sesuai dan tepat dengan algoritma.

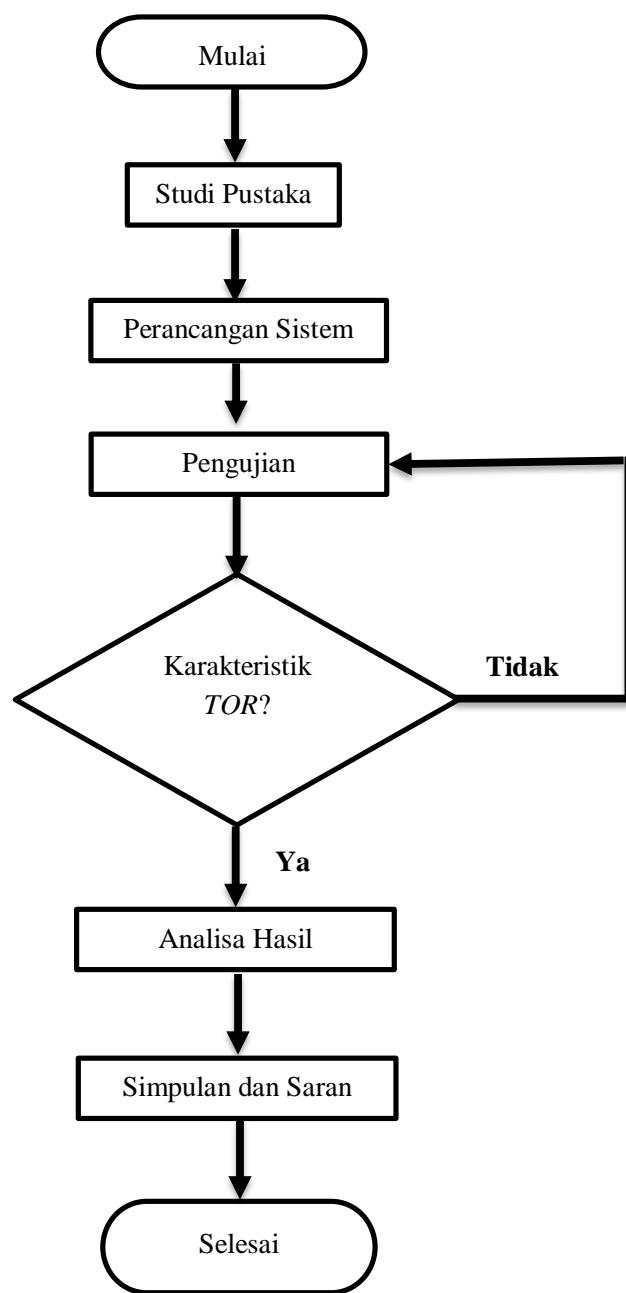
4. Analisa

Hasil dari pengolahan data pada tahap sebelumnya, akan dianalisa sesuai identifikasi permasalahan. Tahapan ini bertujuan untuk mendapatkan data objektif dari analisa hasil pengolahan data serta dapat dilakukannya pengembangan pada penelitian sebelumnya.

5. Kesimpulan dan Saran

Pada tahap ini dilakukan penarikan kesimpulan dalam penulisan tugas akhir. Tahapan ini juga terdapat beberapa poin saran dari penulis untuk penelitian selanjutnya.

Pada Gambar 1.1 berikut ditampilkan metodologi penelitian dalam bentuk diagram air yang merepresentasikan proses pelaksanaan penelitian :



1.7. Sistematika Penulisan

Penyusunan laporan tugas akhir ini, penulis membuat sistematika penulisan agar mempermudah mengetahui isi dari setiap bab yang dibuat pada laporan tugas akhir ini. Adapun seitematika penulisan laporan tugas akhir sebagai berikut :

BAB I. PENDAHULUAN

Bab ini akan menjelaskan tentang latar belakang masalah, tujuan dan manfaat, perumusan masalah, batasan masalah, metodologi penelitian, serta sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Bab ini berisi dasar teori dari penelitian terkait dengan *TOR*, *Intrusion Detection System*, *Clustering K-Means*, dan yang berkaitan langsung dengan penelitian

BAB III. METODELOGI

Bab ini akan menjelaskan tentang langkah-langkah (metodologi) perancangan sistem pada tugas akhir ini..

BAB IV. PENGUJIAN DAN ANALISA

Bab ini akan menjelaskan tentang hasil dari pengujian yang telah dilakukan, dari hasil tersebut akan dilakukan analisa agar mendapatkan data yang akurat.

BAB V. KESIMPULAN

Bab ini akan menjelaskan tentang kesimpulan yang didapat dari data penelitian yang telah dilakukan. dan saran yang diharapkan dapat membuat penelitian ini dikembangkan lebih baik.

DAFTAR PUSTAKA

- [1] P. Mayank, “Tor Traffic Identification,” pp. 85–91, 2017.
- [2] E. A. Winanto, A. Heryanto, and D. Stiawan, “Visualisasi Serangan Remote to Local (R2L) Dengan Clustering K-Means,” *Annu. Res. Semin. 2016*, vol. 2, no. 1, pp. 359–362, 2016.
- [3] F. A. Saputra, I. U. Nadhori, and B. F. Barry, “Detecting and blocking onion router traffic using deep packet inspection,” *Proc. - 2016 Int. Electron. Symp. IES 2016*, no. November, pp. 283–288, 2017.
- [4] L. P. Su and J. A. Zhang, “The Improvement of Cluster Analysis in Intrusion Detection System,” *Proc. - 10th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2017*, vol. 2017-Octob, pp. 274–277, 2017.
- [5] J. Jabez and B. Muthukumar, “Intrusion detection system (ids): Anomaly detection using outlier detection approach,” *Procedia Comput. Sci.*, vol. 48, no. C, pp. 338–346, 2015.
- [6] K. Rajeswari, O. Acharya, M. Sharma, M. Kopnar, and K. Karandikar, “Improvement in k-means clustering algorithm using data clustering,” *Proc. - 1st Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2015*, pp. 367–369, 2015.
- [7] A. H. Lashkari, A. A. Ghorbani, G. D. Gil, M. Saiful, and I. Mamun, “Characterization of Tor Traffic using Time based Features Dark Web detection View project Detecting Malicious URLs View project Characterization of Tor Traffic using Time based Features,” no. Cic, pp. 253–262, 2017.
- [8] F. A. Saputra, I. U. Nadhori, and B. F. Barry, “Detecting and blocking onion router traffic using deep packet inspection,” *Proc. - 2016 Int. Electron. Symp. IES 2016*, no. November, pp. 283–288, 2017.

IES 2016, no. November, pp. 283–288, 2017.

- [9] A. Bermudez-villalva, “A Measurement Study on the Advertisements Displayed to Web Users Coming from the Regular Web and from Tor,” pp. 494–499, 2020.
- [10] Y. Wang, T. T. Gamage, and C. H. Hauser, “Security Implications of Transport Layer Protocols in Power Grid Synchrophasor Data Communication,” *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 807–816, 2016.
- [11] D. S. Dolliver and J. L. Kenney, “Characteristics of Drug Vendors on the Tor Network: A Cryptomarket Comparison,” *Vict. Offenders*, vol. 11, no. 4, pp. 600–620, 2016.
- [12] A. K. Jadoon, W. Iqbal, M. F. Amjad, H. Afzal, and Y. A. Bangash, “Forensic Analysis of Tor Browser: A Case Study for Privacy and Anonymity on the Web,” *Forensic Sci. Int.*, vol. 299, pp. 59–73, 2019.
- [13] S. V. Gajbhiye and G. B. Malode, “Enhancing pattern recognition in social networking dataset by using bisecting KMean,” *Proc. 2017 Int. Conf. Intell. Comput. Control. I2C2 2017*, vol. 2018-January, pp. 1–5, 2018.
- [14] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and L. Fernández-Robles, “ToRank: Identifying the most influential suspicious domains in the Tor network,” *Expert Syst. Appl.*, vol. 123, pp. 212–226, 2019.
- [15] M. Landvoigt and J. Cardoso, “Analysis about publications on Facebook pages: finding of important characteristics,” p. 8, 2017.
- [16] Z. Du, L. Ma, H. Li, Q. Li, G. Sun, and Z. Liu, “Network Traffic Anomaly Detection Based on Wavelet Analysis,” *Proc. - 2018 IEEE/ACIS 16th Int. Conf. Softw. Eng. Res. Manag. Appl. SERA 2018*, pp. 94–101, 2018.