JISKa, Vol. 4, No. 1, Mei, 2019, Pp. 1 – 10
ISSN : 2527 – 5836 (print) | 2528 – 0074 (online)

Analisis Serangan *Web Defacement* pada Situs Web Pemerintah Menggunakan *ELK Stack*

Fazlurrahman [(1)], Dedy Hariyadi [(2)],
Komunitas NgeSec Yogyakarta [(1)], Universitas Jenderal Achmad Yani Yogyakarta [(2)],
e-mail : fazlurbima@gmail.com [(1)], milisdad@gmail.com [(2)] ,

Abstract

Web defacement is an attack that is categorized as a potential cyber attack. The impact of web defacement attacks is that the process of returning to the original condition requires special mitigation. Monitoring web defacement attacks can be done from two sides, namely internal and external sides. In this study external monitoring and analysis of web defacement attacks. Therefore a special application is needed to get information from OSINT Source, an information provider for web defacement attacks which is the result of reports from hackers. The information obtained is then processed using ELK Stack to facilitate analysis in the form of visualization on the Dashboard.

Keywords : Defacement, Cyber Security, ELK Stack, Web Scraping, E-Government

Web defacement merupakan serangan yang dikategorikan sebagai serangan siber yang berpotensi. Dampak dari serangan *web defacement* adalah proses pengembalian ke kondisi semula yang memerlukan mitigasi khusus. Pemantauan serangan *web defacement* dapat dilakukan dari dua sisi, yaitu sisi internal dan eksternal. Pada penelitian ini pemantauan dan analisis serangan *web defacement* dari sisi eksternal. Oleh sebab itu diperlukan aplikasi khusus untuk mendapatkan informasi dari *OSINT Source* penyedia informasi serangan *web defacement* yang merupakan hasil laporan dari peretas. Informasi yang didapatkan selanjutnya diolah menggunakan ELK Stack untuk mempermudah analisis dalam bentuk visualisasi pada *Dashboard*.

Kata Kunci : *Defacement*, Keamanan Siber, *ELK Stack, Web Scraping, E-Government*

1. PENDAHULUAN

Penetrasi pengguna internet di Indonesia meningkat dari tahun ke tahun. Sejak 10 tahun yang lalu pengguna selalu naik minimal ± 10.000.000 pengguna setiap tahunnya. Pada tahun 2018 mengalami peningkatan ± 27.900.000 pengguna dibandingkan tahun 2017 (Asosiasi Penyelenggara Jasa Internet Indonesia, 2019). Tindak kejahatan melalui media digital atau teknologi informasi dan komunikasi pun selaras dengan pertumbuhan pengguna internet, selalu mengalami peningkatan penerimaan barang bukti elektronik dan/atau digital di Kepolisian (Hariyadi, Winarno, & Luthfi, 2016).

Tindak kejahatan berupa serangan siber sepanjang tahun 2018 dengan objek pantau situs web sebanyak 16.939 serangan. Situs web dengan domain .go.id mendapat serangan terbanyak dibanding *Country Code Top Level Domain* (ccTLD) .id lainnya. Tabel 1 menunjukan persentase pemantauan insiden situs web pada tahun 2018 dengan ccTLD .id. Bentuk insiden yang terjadi pada situs web adalah *web defacement* (Indonesia Security Incident Response Team on Infrastructure / Coordination Center, 2019). Pada penelitian sebelumnya serangan *web defacement* pada situs web milik pemerintah juga pada urutan pertama berdasarkan observasi pada bulan Januari sampai dengan Juli 2014 (Mantra, 2015).

Tabel 1. Pemantauan Insiden Situs Web Tahun 2018

No	ccTLD	Persentase
1	.go.id	30.75 %
2	.ac.id	28.38 %
3.	.sch.id	12.58 %
4.	.co.id	10.92 %
5.	.id	8.25 %
6.	.or.id	2.96 %
7.	.desa.id	2.76 %
8.	.web.id	2.56 %
9.	.my.id	0.53 %
10.	.mil.id	0.11 %
11.	.biz.id	0.08 %
12.	.net.id	0.08 %
13.	.ponpes.id	0.03 %

Defacement pada situs web atau *web defacement* dapat diartikan tindakan mengubah tampilan halaman situs yang tidak semestinya oleh orang yang tidak memiliki otoritas (Romagna & Hout, 2017). Masih menurut Rogmana dan Hout serangan *web defacement* merupakan serangan yang berpotensi karena memerlukan biaya untuk memperbaikinya. Penelitian ini fokus pada pemantauan serangan *web defacement* dengan studi kasus situs web pemerintah yaitu yang memiliki domain .go.id supaya pihak pemerintah memperhatikan pembiayaan setelah pengembangan sebuah sistem berbasis web.

2. KAJIAN PUSTAKA

2.2. TOP-LEVEL *DOMAIN*

Domain Name System (DNS) merupakan sistem yang berfungsi mengkoversi nama domain yang mudah diingat ke dalam bentuk IP Address dengan melakukan permintaan informasi ke sistem yang memiliki hierarki dan tersebar. Adanya DNS maka memudahkan menghubungkan sumber daya komputasi baik melalui internet maupun jaringan internal (Mockapetris, 1987). Sistem hierarki DNS tertinggi yang disebut *Root* yang melakukan pendelegasian tanggung jawab ke *Top-Level Domain (TLD)*. Contoh *Top-Level Domain* diantaranya: .com, .net, .org, .info, .online, dan .id. Organisasi nirlaba yang mengelola DNS dan IP adalah Internet Corporation for Assigned Names and Numbers (ICANN). Adapun daftar basis data *Root Zone*[1] dikelola Internet Assigned Numbers Authority (IANA), sebuah departemen di bawah ICANN (Wang, Zhang, & Xu, 2018). Berdasarkan surat Dirjen APTEL Kementerian Komunikasi dan Informatika Nomor BA–343/DJAT/MKOMINFO/6/2007 pengelolaan *Top-Level Domain* .id diserahkan dari Dirjen APTEL ke Pengelola Nama Domain Internet Indonesia (Pandi). Pandi tidak hanya mengelola *Top-Level Domain* .id, sub domain dua tingkat dibawahnya juga dikelola oleh Pandi. Adapun sub domain dua tingkat tersebut adalah .co.id, .ac.id, .or.id, .go.id, .my.id, .web.id, .biz.id,
.net.id, .mil.id, .sch.id, .desa.id, dan .ponpes.id (Pengelola Nama Domain Internet Indonesia, n.d.).

1 . ftp://ftp.rs.internic.net/domain/root.zone

2.3. WEB SCRAPING

Web scraping merupakan kode yang memanfaatkan teknik untuk melakukan ekstraksi informasi dengan sumber suatu halaman situs web. Informasi yang didapatkan dapat disimpan pada suatu berkas dengan format, open document spreadsheet, Microsoft Excel, Comma Separated Value, Structured Query Language, Extensible Markup Language, atau berkas teks. Kode tersebut maka disebut *web scraper* yang dibuat menggunakan bahasa pemrograman tertentu (Mishra & Pujari, 2011). Pada penelitian ini bahasa yang digunakan untuk melakukan *scraping* situs web adalah Python. Istilah lain dari *web scraping* juga dikenal sebagai *screen scraping*, *web data extraction* atau *web harvesting* (Jain & Kasbe, 2018).

2.4. PUSTAKA PYTHON

Dipilihnya bahasa pemrograman Python pada penelitian ini karena ketersediaan pustaka yang mendukung untuk proses *web scraping*. Adapun pustaka Python yang digunakan dalam penelitian ini sebagai berikut:

1. *Requests*, pustaka yang penggunaan mudah dan sederhana dalam pemanfaatan *HTTP Persistent Connection.* Hal ini memudahkan dalam berkomunikasi dengan *web service* dengan berbagai metode yang digunakan dalam penelitian ini, yaitu GET untuk mengambil data dari halaman web (Mehak, Zafar, Aslam, & Bhatti, 2019).
2. CSV, pustaka manipulasi data yang dikhususkan pada berkas berformat CSV. Berkas berformat CSV merupakan dokumen bertipe MIME Text sesuai standar RFC 4180. Dipilih berkas format CSV untuk memudahkan pengolahan data (Nastiti, Hariyadi, & Fazlurrahman, 2019).
3. *Argparse*, pustaka yang berfungsi melakukan parsing argumen suatu masukan, perintah tambahan dan pilihan suatu perintah yang berbasis *Command Language Interpreter*.
4. *BeatifulSoup*, fungsi dari pustaka ini mengambil informasi pada halaman situs web baik dalam bentuk HTML ataupun XML. Proses pengambilan informasi menggunakan pendekatan pohon dari Document Object Model (DOM) (Mehak et al., 2019).
5. Datetime, hasil dari *web scraping* berupa berkas CSV yang memanfaatkan pustaka CSV yang dikombinasikan dengan pustaka ini untuk penamaan berkas berdasarkan waktu pengambilan atau *scraping*.
6. Cookie, untuk meminimalisir mengubah kode web scraping dengan memanfaatkan pustaka ini untuk memisahkan berkas yang berisi informasi dari Cookie situs web.

2.5. ELK STACK

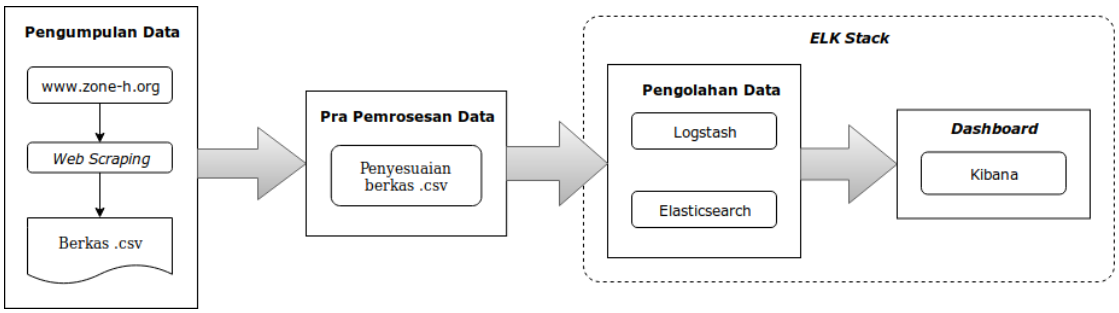
ELK Stack yang merupakan kependekan dari Elasticsearch, Logstash, Kibana sebagai satu kesatuan sebuah sistem dengan fungsi melakukan sebuah analisis dengan visualisasi untuk mempermudah pengguna. Sesuai kependekannya ELK Stack terdiri dari tiga komponen yaitu (Prakash, Kakkar, & Patel, 2016):

1. Elasticsearch merupakan pengindeks konten dari sebuah mesin pencarian situs web yang pencariannya dan daftar informasinya memanfaatkan arsitektur RESTful sebagai JSON diatas protokol HTTP. Proyek pengembangan Elasticsearch dibawah naungan Apache's Lucene Project dengan lisensi Apache License versi 2 yang mudah untuk diadaptasi tanpa biaya tinggi.
2. Logstash merupakan alat pengelola sebuah catatan berupa log atau berkas teks lainnya yang selanjutnya untuk diolah oleh Elasticsearch.
3. Kibana berfungsi untuk mengolah informasi dari Logstash berdasarkan dari pengindeksan dari Elasticsearch dalam bentuk visualisasi yang mempermudah pengguna.

3. DESAIN SISTEM

Pada penelitian ini desain sistemnya terbagi menjadi 4 bagian, yaitu Pengumpulan Data, Pra Pemrosesan Data, Pengolahan Data, dan *Dashboard*. Gambar 3.1 menunjukan desain sistem dari bagian pengumpulan data sampai dengan visualisasi analisis serang siber. Berdasarkan Tabel 1 serangan siber berupa *web defacement* dialami oleh pemerintah dengan *Top-Level Domain* .go.id. Oleh sebab itu pada bagian Pengumpulan Data sumber data yang digunakan khusus pada *Top-Level Domain* dari situs web www.zone-h.org. Metode yang digunakan untuk

mendapatkan informasi serangan pada situs web www.zone-h.org adalah *web scraping*. Hasil dari *web scraping* disimpan dalam berkas berformat .csv.



Gambar 3.1. Desain Sistem

Berkas .csv hasil dari *web scraping* perlu dilakukan validasi pada bagian Pra Pemrosesan Data. Hal ini bertujuan untuk mendapatkan data yang rapi berdasarkan kaidah atau kolom-kolom yang telah ditetapkan. Beberapa kasus, informasi yang di-scraping ada yang tidak sesuai dengan kolom-kolom yang telah ditetapkan. Bagian Pengolahan Data dan Dashboard terintegrasi dengan ELK Stack. Berkas .csv yang telah sesuai dengan kaidah maka diproses oleh Logstash dan Elasticsearch untuk dibuatkan pengindeksan. Selanjutnya visualisasi serangan siber pada situs web pemerintah dengan Top-Level Domain .go.id melalui Kibana.

4. HASIL DAN PEMBAHASAN

Beberapa serangan *web defacement* dilaporkan oleh penyerang ke situs web www.zone-h.org. Serangan berasal dari berbagai negera sehingga untuk memisahkan serangan *web defacement* ke pemerintah menggunakan sub-domain dari TLD, yaitu .go.id. Untuk menarik informasi serangan *web defacement* pada www.zone-h.org terlebih dahulu menelusuri pola dari *Uniform Resource Locator* (URL) dengan kata kunci situs web pemerintah Indonesia. Adapun URL yang digunakan untuk mendapatkan serangan *web defacement* pada situs wes pemerintah di Indonesia adalah http://www.zone-h.org/archive/filter=1/special=1/domain=go.id/fulltext=1/. Selain mencatat URL yang diperlukan dalam penelitian ini adalah *Cookies* saat mengakses URL tersebut.

Web scraper[2] yang dikembangkan menggunakan bahasa pemrograman Python membutuhkan beberapa pustaka. Oleh sebab itu sebelum menjalankan aplikasi *web scraper* terlebih dahulu installasi pustaka-pustaka yang diperlukan. Berdasarkan penelusuran kata kunci .go.id didapatkan daftar situs-situs web pemerintah Indonesia yang ter-*deface* dengan penyajian per halaman web dua puluh lima situs web pemerintah Indonesia. Oleh sebab itu proses *scraping* diperlukan juga nomor lembar halaman situs web. Penggunaan aplikasi *web scraper* menggunakan perintah *python makaboro.py –start [Nomor Halaman Terdepan] --stop [Nomor Halaman Terakhir] –output [Penanda Berkas]*. Dari perintah tersebut akan mendapatkan berkas berformat *zoneH-Tahun-Bulan-Tanggal-Penanda.csv*. *Tahun-Bulan-Tanggal* merupakan waktu pengambilan data. Penanda berfungsi sebagai pemberian versi pengambilan data secara manual.

Berkas .csv yang didapatkan tidak dapat langsung dilakukan analisis. Berkas tersebut perlu dilakukan penyesuaian, diantaranya memastikan bahwa setiap kolom terisi dengan informasi yang benar dan mengisi kolom Institusi yang belum bisa didapatkan saat *scraping*. Tabel 2 sebagian contoh hasil *web scraping* dalam bentuk berkas .csv. Informasi yang didapatkan diantaranya Halaman Terdeface, Attacker, Tebas Index, Kejadian, Arsip. Sedangkan nama institusi pemerintah belum bisa didapatkan.

2 https://github.com/orangmiliter/makaboro

Tabel 2: Hasil Scraping

Institusi	Halaman Terdeface	*Attacker*	Tebas Index	Kejadian	Arsip
	perpustakaan.pn-medankota.go.i...	Vijune15	T	2019-03-07 00:00:00	www.zone-h.org/mirror/id/32252173
	wondamakab.go.id/galau.htm	Astra	T	2019-03-07 00:00:00	www.zone-h.org/mirror/id/32252807
	www.blh.badungkab.go.id/k.htm	Mr.Yagami	T	2019-03-07 00:00:00	www.zone-h.org/mirror/id/32252501
	takalarkab.go.id/lol.php	MR.5T1Y0	T	2019-03-06 00:00:00	www.zone-h.org/mirror/id/32251085
	oku.sumsel.polri.go.id/readme....	KURD ELECTRONIC TEAM	T	2019-03-05 00:00:00	www.zone-h.org/mirror/id/32248515
	pa-taliwang.go.id	by_dadaş	Y	2019-03-05 00:00:00	www.zone-h.org/mirror/id/32248369

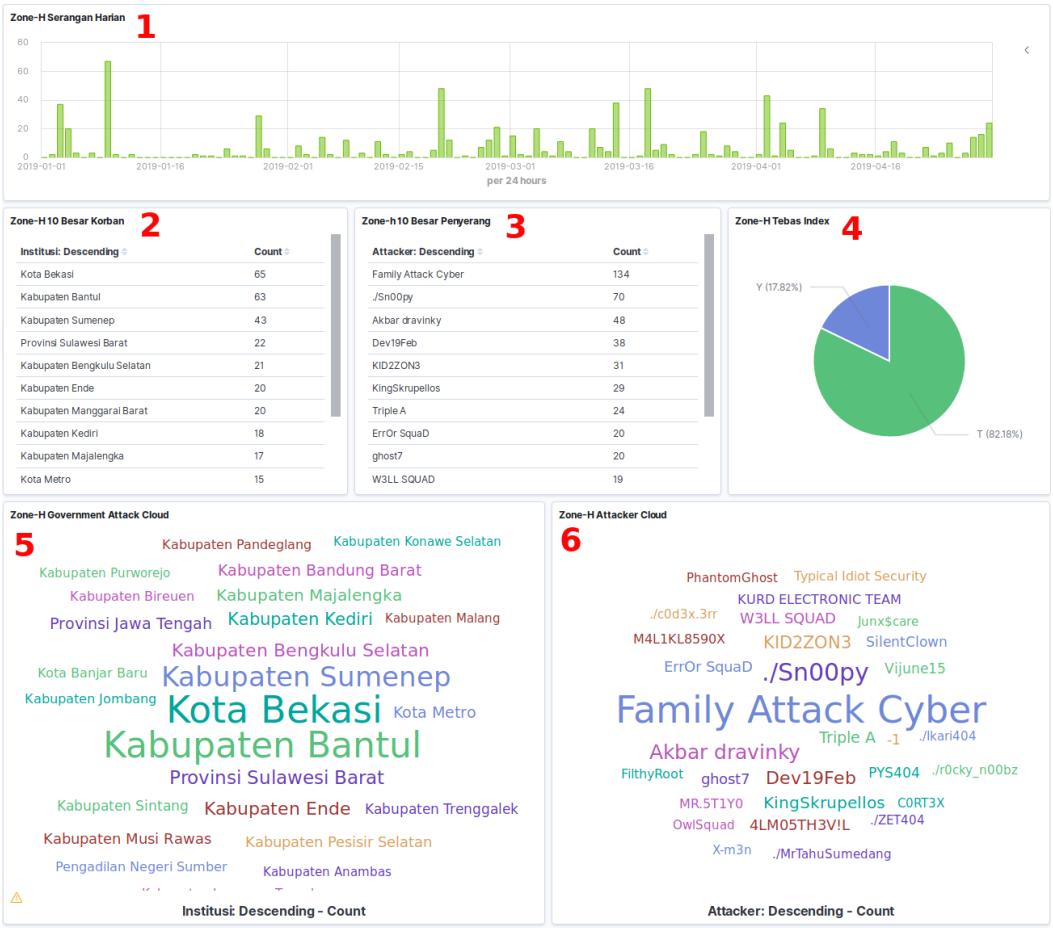
Setelah berkas .csv sudah ter-*scrap* proses selanjutnya adalah melakukan penyesuaian berkas .csv. Terlihat pada Tabel 2 kolom Insitusi masih kosong. Oleh sebab itu kolom Institusi diisi manual berdasarkan kaidah nama institusi pemerintah. Tabel 3 menunjukan proses dari hasil penyesuaian bekars pada kolom institusi.

Tabel 3. Hasil Penyesuaian Berkas

Institusi	Halaman Terdeface	*Attacker*	Tebas Index	Kejadian	Arsip
Pengadilan Negeri Kota Meda	perpustakaan.pn-medankota.go.i...	Vijune15	T	2019-03-07 00:00:00	www.zone-h.org/mirror/id/32252173
Kota Wondaman	wondamakab.go.id/galau.htm	Astra	T	2019-03-07 00:00:00	www.zone-h.org/mirror/id/32252807
Kabupaten Badung	www.blh.badungkab.go.id/k.htm	Mr.Yagami	T	2019-03-07 00:00:00	www.zone-h.org/mirror/id/32252501
Kabupaten Takalar	takalarkab.go.id/lol.php	MR.5T1Y0	T	2019-03-06 00:00:00	www.zone-h.org/mirror/id/32251085
Kepolisian	oku.sumsel.polri.go.id/readme....	KURD ELECTRONIC TEAM	T	2019-03-05 00:00:00	www.zone-h.org/mirror/id/32248515
Pengadilan Agama Taliwang	pa-taliwang.go.id	by_dadaş	Y	2019-03-05 00:00:00	www.zone-h.org/mirror/id/32248369

Berkas yang telah disesuaikan diunggah ke mesin ELK Stack untuk diolah menggunakan Logstash dan Elasticsearch yang selanjutnya ditampilkan dalam bentuk *Dashboard* menggunakan Kibana. Pada *Dashboard* yang tampak pada Gambar 4.1 terdiri dari enam bagian, yaitu grafik serangan harian *web defacement* (1), sepuluh besar institusi pemerintah yang menjadi korban *web defacement* (2), sepuluh besar penyerang (3), grafik yang menunjuk

persentase dampak serangan tebas index (4), visualisasi cloud situs web pemerintah yang terkena dampak (5), dan visualisasi *cloud* penyerang situs web pemerintah (6).



Gambar 4.1. Dashboard Serangan Web Defacement

Analisis dan visualisasi pada Gambar 4.1 menggunakan ELK Stack dengan sumber data yang diambil dari 1 Januari 2019 sampai dengan 30 April 2019. Serangan web defacement tidak selalu dilaporkan setiap hari oleh penyerang. Dalam hal ini penyerang ada dalam bentuk kelompok ataupun individu. Pada situs www.zone-h.org tidak mengklasifikasikan peretas menjadi kelompok atau individu. Dalam rentang 1 Januari 2019 sampai dengan 30 April 2019 institusi pemerintah yang lebih banyak diserang adalah Kota Bekas sedangkan pelaku yang paling sering melakukan penyerangan adalah grup Family Attack Cyber.

Tabel 4. Sepuluh Besar Institusi Terdampak Web Defacement

No	Institusi Terdampak	Jumlah Serangan
1	Kota Bekas	65
2	Kabupaten Bantul	63
3	Kabupaten Sumenep	43
4	Provinsi Sulawesi Barat	22
5	Kabupaten Bengkulu Selatan	21
6	Kabupaten Ende	20
7	Kabupaten Manggarai Barat	20
8	Kabupaten Kediri	18
9	Kabupaten Majalengka	17
10	Kota Metro	15

Tabel 5. Sepuluh Besar Penyerang

No	Penyerang	Kategori penyerang	Jumlah Serangan
1	Family Attack Cyber	Grup	134
2	./Sn00py	Individu	70
3	Akbar dravinky	Individu	48
4	Dev19Feb	Individu	38
5	KID2ZON3	Individu	31
6	KingSkrupellos	Grup	29
7	Triple A	Individu	24
8	ErrOr SquaD	Grup	20
9	ghost7	Grup	20
10	W3LL SQUAD	Grup	19

Dampak serangan *web defacement* terbagi menjadi dua, yaitu Tebas Index dan Serangan Sub Halaman Web. Tebas Index merupakan istilah yang sering digunakan oleh para penyerang dengan dampak serangan halaman depan terganti dengan halaman yang tidak semestinya. Tentu halaman tersebut telah dipersiapkan oleh penyerang dengan berbagai pesan.

5. DAFTAR PUSTAKA

Asosiasi Penyelenggara Jasa Internet Indonesia. (2019). *Penetrasi dan Perilaku Pengguna Internet Indonesia 2018*. Jakarta.

Hariyadi, D., Winarno, W. W., & Luthfi, A. (2016). Analisis Konten Dugaan Tindak Kejahatan Dengan Barang Bukti Digital Blackberry Messenger. *Teknomatika STMIK Jenderal Achmad Yani Yogyakarta*, *9*(1), 81–89. Diambil dari http://teknomatika.stmikayani.ac.id/wp-content/uploads/2017/01/Teknomatika-9-1-8-Hariyadi-Analisis-Barang-Bukti-Digital-BBM.pdf

Indonesia Security Incident Response Team on Infrastructure / Coordination Center. (2019). *Indonesia Cyber Security Monitoring Report 2018*. Jakarta.

Jain, A., & Kasbe, A. (2018). Fake News Detection. *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2018*, 1–5. https://doi.org/10.1109/SCEECS.2018.8546944

Mantra, I. (2015). Indonesia Web Defacement Attacks Analysis for Anti Web Defacement. *Jurnal TICOM*, *3*(3).

Mehak, S., Zafar, R., Aslam, S., & Bhatti, S. M. (2019). Exploiting Filtering Approach with Web Scrapping for Smart Online Shopping. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies, iCoMET 2019*, 1–5. https://doi.org/10.1109/ICOMET.2019.8673399

Mishra, D., & Pujari, N. (2011). Cross-domain query answering: Using Web scrapper and data integration. *2011 2nd International Conference on Computer and Communication Technology, ICCCT-2011*, 27–32. https://doi.org/10.1109/ICCCT.2011.6075193

Mockapetris, P. V. (1987). *RFC 1035: Domain Names - Implementation and Specification*.

Nastiti, F. E., Hariyadi, D., & Fazlurrahman. (2019). TelegramBot : Crawling Data Serangan Malware dengan Telegram. *Journal of Computer Engineering System and Science*, *4*(1). https://doi.org/10.24114/cess.v4i1.11436

Pengelola Nama Domain Internet Indonesia. (n.d.). Tentang PANDI. Diambil 1 Februari 2019, dari https://pandi.id/profil/tentang-pandi/

Prakash, T., Kakkar, M., & Patel, K. (2016). Geo-Identification of Web Users through Logs using ELK Stack. In *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016* (hal. 606–610). https://doi.org/10.1109/CONFLUENCE.2016.7508191

Romagna, M., & Hout, N. J. Van Den. (2017). Hacktivism and website defacement : Motivations, capabilities and potential threats. *27th Virus Bulletin International Conference*, (October).

Wang, M., Zhang, Z., & Xu, H. (2018). DNS Configurations and Its Security Analyzing via Resource Records of the Top-Level Domains. In *Proceedings of the International Conference on Anti-Counterfeiting, Security and Identification, ASID* (Vol. 2017-Octob, hal. 21–25). https://doi.org/10.1109/ICASID.2017.8285736