PENGGUNAAN ALGORITMA K-MEANS PADA DATALOG HONEYNET UNTUK PROFILING SERANGAN DDOS

THE USE OF K-MEANS ALGORITHM ON DATALOG HONEYNET FOR PROFILING DDOS ATTACKS

Winda Andrini Wulandari
Program Studi Magister Teknik Informatika Universitas Islam Indonesia
Jalan Kaliurang KM 14,5, Umbulmartani, Ngemplak, Kabupaten Sleman,
Daerah Istimewa Yogyakarta 55584
e-mail: windayswuland@gmail.com

ABSTRAK

Komputer awan (cloud computing) menjadi primadona karena keunggulanya, bersifat flexibel, dan dapat digunakan secara bersama pada berbagai aspek dan bidang usaha. Menurut Arbor Network bahwa teknologi awan sangat rentan terhadap serangan DDoS dan frekuensi serangan meningkat tajam di beberapa tahun terakhir. Penelitian ini hasil kerjasama dengan Universitas AWN dan Indonesia Honeypot Project (IHP) Kementrian Komunikasi dan Informatika. Hasilnya mengusulkan bahwa honeynet sebagai sebuah cara untuk melindungi dan memantau keamanan jaringan public cloud computing, serta menganalisis datalog-nya melalui pendekatan metode k-means untuk mengenali cyber profilling serangan DDoS berdasarkan timestamp yang terekam di dalam data.

Kata Kunci*: Public Cloud Computing, Honeypot, K - Means Clustering, Cyber Crime, Distributed Denial of Service (DDoS).*

ABSTRACT

cloud computing are admired for excellence, flexible benefits, and can be used together in various aspects and areas of business. According to Arbor Network that cloud technology is very vulnerable to DDoS attacks and the frequency of attacks has increased sharply in recent years.This research is the result of cooperation with AWN University and Indonesia Honeypot Project (IHP) Ministry of Communications and Informatics. The results suggest that honeynet as a way to protect and monitor the security of public cloud computing networks, and analyze its datalog through the k-means method approach to recognize cyber profilling DDoS attacks based on the timestamp recorded in the data.

Keywords*: Public Cloud Computing, Honeypot, K - Means Clustering, Cyber Crime, Distributed Denial of Service (DDoS).*

I. PENDAHULUAN

Pada tahun 2015**,** *server* utama di jaringan *public* Universitas AWN mengalami *downtime* yang cukup lama disebabkan membanjirnya *bandwidth* yang mempengaruhi *Quality of Servic*-kualitas pelayanan (*QoS*) dan *Service Level Aggrement*-persetujuan tingkat layanan (*SLA*) dalam jaringan tersebut.

Dalam seminar nasional yang bertema "*Smart City For Smart* Indonesia" pada bulan Oktober 2017 di Universitas Gajah Mada di Yogyakarta, para pakar dan praktisi IT mengatakan bahwa, sedang

dalam proses membangun kota pintar di beberapa kota besar di Indonesia yang dianggap lebih murah, mudah diakses, mendukung keefektifitan kinerja, proses evaluasi, dokumentasi, dan pengarsipan agar mudah dikelola secara intensif oleh pusat pemerintahan [2].

Dilansir dari *Arbor Network* bahwa teknologi *cloud* sangat rentan terhadap serangan DDoS, frekuensinya meningkat tajam di beberapa tahun terakhir, dan biasanya ditargetkan ke *port* terutama layanan *web* HTTTP tipe GET seperti TCP SYN, ICMP (*Internet Control Message Protocol*), NTP (*Network Time Protocol*) atau SNMP (*Simple Network Management Protocol*) yang menimbulkan dampak kerugian ekonomi dan hilangnya kepercayaan terhadap ISP (*Intenet Service Provider*) [12].

Tentu saja hal tersebut menimbulkan kekhawatiran sehingga [6] mengusulkan *honeypot* sebagai sebuah cara yang praktis untuk melindungi aset dari serangan dan penyalahgunaan, serta teknik *data mining* menjadi langkah yang tepat untuk memprediksi dan mengetahui *vulnerabilty*-kerentanan terhadap arus anomali yang mencurigakan di dalam *traffic* jaringan berdasarkan data *log honeypot*. Data *mining* dapat membantu dalam menilai kesamaan karakteristik, prediksi [5], dan menemukan pola, menguraikan penemuan dalam sumber data yang mencakup *database*, dan gudang data pada sistem dinamis. Di dalam data mining dikenal sebuah metode *clustering*-kekelompokan, *clustering* adalah proses pengelompokkan beberapa data ke dalam *cluster*-kelompok atau gugus yang memiliki kemiripan paramater dan objek satu sama lain dan yang tidak memiliki sifat yang sangat tidak mirip dengan obyek lain maka berada pada *cluster* yang berbeda [11].

II. LANDASAN TEORI

A. *Honeynet/ Honeypot*

Honeynet adalah sistem yang terdiri dari beberapa kumpulan *honeypot*. *Honeypot* adalah sistem umpan untuk mengumpulkan informasi *attacker* dari penyerang dengan cara menunggu, memantau setiap aktivitas *attacker* yang memulai interaksi, mengumpulkan sebanyak – banyaknya data yang dikenali sebagai sebuah serangan untuk melindungi sistem dan jaringan [8].

B. *Public Cloud Computing*

NIST (*National Institute of Standar and Technology*) di dalam *draft* – nya yang berjudul the NIST *Defitinition of Cloud Computing* (Peter Meel dan Timothy Grance) mendefinisikan *Cloud computing* sebagai sebuah perkembangan teknologi layanan yang mudah dikonfigurasi, memiliki penyimpanan, dan sumber daya yang mudah disesuaikan dengan kebutuhan [9].

C. *DDoS*

Pada dasarnya DDoS adalah sama yang membedakan adalah terletak dari model penyerangannya, DoS merupakan model serangan yang berasal dari satu sumber saja, berbeda halnya dengan serangan DDoS yang dilakukan secara beramai-ramai dengan menggunakan ratusan bahkan ribuan komputer untuk meruntuhkan layanan yang ada [3]

D. *K – Means Clustering*

Clustering adalah proses mengelompokkan masing – masing *cluster* yang mewakili kesamaan perilaku objek data dan kemiripan karakteristik. *K – Means clustering* mengelompokkan data secara tegas dengan cara meminimalkan jarak antara suatu data atau kelompok data terhadap titik *cluster (centroid)* data. Secara umum metode *K – Means clustering* dianalogikan sebagai berikut [10]:

1. Inisialisasi: menentukan nilai K sebagai *centroid* dan matrik ketidakmiripan (jarak) yang diinginkan. Jika perlu, tetapkan ambang batas perubahan objektif dan ambang batas perubahan posisi tersebut.

$$\boldsymbol{v} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_1}}{\boldsymbol{n}} \; ; \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{1})$$

2. Bangkitkan *centroid* awal secara *random* dari objek – objek yang tersedia sebanyak k *cluster*, lalu menghitung *centroid* dengan menggunakan rumus berikut :

$$v = \frac{\sum_{i=1}^{n} x_1}{n} \quad ; i = 1, 2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

untuk,

v : *centroid* pada *cluster*.
xi : objek ke - i.
n : banyaknya objek / jumlah objek yang menjadi anggota *cluster*.

3. Hitung jarak setiap objek ke masing – masing dari masing *centroid*. Untuk menghitung jarak antara objek dengan *centroid* menggunakan *euclidean distance* dengan rumus sebagai berikut:

$$d\,(x - y) = ||x - y|| = \sqrt{\sum_{j=1}^{n} (xi - yi)^2} \quad ; i = 1, 2, \ldots\ldots\ldots\ldots\ldots (3)$$

untuk,

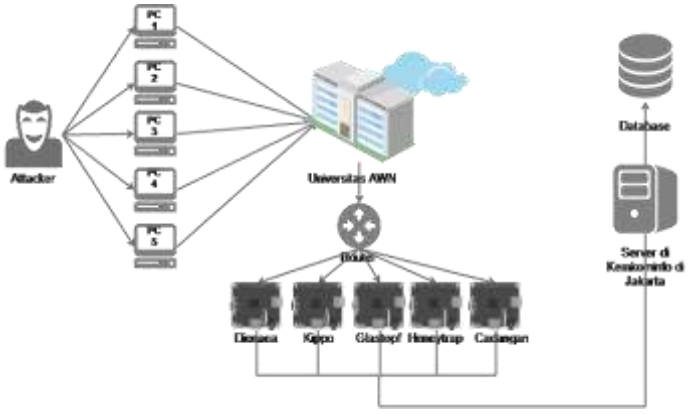
xi : objek x ke - i
yi : daya y ke - i
n : banyaknya objek.

a. Mengalokasikan masing – masing objek ke dalam *centroid* yang paling terdekat.
b. Melakukan iterasi dan tentukan posisi *centroid* baru dengan menggunakan persamaan (2).
c. Ulangi langkah tiga dan empat hingga kondisi konvergen tercapai dengan beberapa syarat berikut :
   - Perubahan fungsi objektif sudah diambang bawah batas yang ditentukan.
   - Tidak ada data yang berpindah *cluster*.
   - Perubahan *centroid* sudah di bawah ambang batas yang sudah ditetapkan.
   - Perubahan *centroid* sudah dibawah ambang batas yang sudah ditetapkan.

III. METODOLOGI PENELITIAN

E. *Raspberry Honeypot* sebagai Media Penyimpanan *Public Cloud*

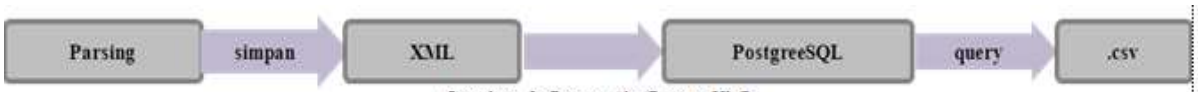
Honeypot data dikatakan sebagai sebuah media penyimpanan *cloud* karena secara virtual fungsinya dapat menyimpan data *storage file log* dari hasil tindakan *cyber space* tersebut. Gambar 1 menjelaskan bahwa *Honeypot* yang berada pada jaringan *public* ini merupakan *cloud computing* bertipe IAAS (*Infrastructure AS A Service*) yang terhubung dari Universitas AWN ke *server* pusat milik (*Indonesian Honeypot Project*) IHP di (Kemkominfo) Jakarta. *Raspberry honeypot* yang berjumlah lima buah ini mengemulasikan beberapa *port* yang terdiri dari RbHp1 (Dionea : port 21, 69, 80, 445, dan 145), RbHp2 (Kippo : port 22), RbHp3 (Glastopf : port 80), RbHp4 (Honeytrap seperti proxy, membuka layanan SMTP), RbHp5 (cadangan *honeypot* yang disesuaikan dengan kebutuhan, biasanya digunakan untuk cadangan *spam*). *Honeypot* akan mem-*filter* setiap aliran *traffic*, maka yang bersifat anomali akan direkam oleh *honeypot*, namun *traffic* yang dianggap tidak berbahaya akan sampai ke *server* asli.



Gambar 1. Arsitektur Jaringan Honeypot Public Cloud Computing

F. Pengumpulan Data

Gambar 2 menjelaskan bahwa masing – masing *sensor raspberry honeypot* dilengkapi dengan *database SQLite* yang dikofigurasikan dengan *Extensible Messaging and Presence Protocol* (XMPP) yang dikoneksikan ke sebuah *chatting room server* menggunakan *script python*. Maka ketika data serangan masuk ke *honeypot* akan di – *parsing* kemudian diproses oleh XML lalu disimpan ke format *PostgreeSql* setelah itu di *query* ke format *file* .csv. XMPP adalah bahasa *markup language* yang memungkinkan beberapa aplikasi, web, pesan instan dapat berkomunikasi dan berkolaborasi secara *real time***]**. Sehingga setiap ada serangan masuk ke *honeypo*t akan muncul notifikasi yang berupa pesan yang masuk ke *chatting room server*.



Gambar. 2. Bagan Alir Proses XML

Di dalam *datalog honeypot* terdiri dari 11 atribut log berupa: *number connection*, *connection transport (protocol), ip destination*, *local port (port* yang dituju), *ip_source*, *remote_port* (nomor unik dari *device attacker*), *remote_hostname* (*service sharing*), *country name* (asal negara), dan *country short* (inisial negara sumber serangan).

Menggunakan *K – Means* untuk mengkategorikannya berdasarkan tiga pembagian waktu (pagi, siang, dan malam) Waktu Indonesia Barat (WIB). Penggunaan *K – Means* diharapkan dapat menciptakan sebuah sistem yang memberikan visual waktu terjadinya serangan DDoS, dan informasi *cyber profilling* sebagai antisipasi mencegah kerentanan jaringan terhadap *cyber crime dashboard* berbasis *Hypertext Prepocessor* (PHP).

K – Means Clustering

K – Means Clustering mengelompokkan data secara tegas dengan cara meminimalkan jarak antara suatu data atau kelompok data terhadap titik *cluster (centroid)* data menggunakan rumus persamaan *euclidean distance* pada persamaan-1 menggunakan alat dan bahan yang dibutuhkan sebagai berikut :

a) Hardware
   a. *Processor Intel*(R) *Core* (TM) i5 - 7200U CPU @ 2.50GHz 2.71 GHz.
   b. *RAM* 4.00 GB
   c. *System Type* 64 – *Bit Operating System*

b) Software
   a. *Operating System Windows* 10

b. *Tools*
XAMPP
Notepad ++

TABEL I
DATA AWAL

Asal Negara	
Asal Negara	**Waktu (Jam)**
China	3
Indonesia	**1**
Canada	7
Australia	17
German	**12**
Nigeria	1
Mesir	20
Jepang	**20**
Iraq	9
Brunei	10
Singapore	23

TABEL II
CENTROID RANDOM

Centroid	*Asal Negara*	*Centroid Random*
C1	Indonesia	1
C2	German	13
C3	Jepang	20

TABEL III
HITUNG JARAK (EUCLIDEAN DISTANCE)

Asal Negara	C1	C2	C3
China	2	10	17
Indonesia	0	12	19
Canada	6	6	13
Australia	16	4	3
German	11	1	8
Nigeria	0	12	19
Mesir	19	7	0
Jepang	19	7	0
Iraq	8	4	11
Brunei	9	3	10
Singapore	22	10	3

TABEL IV
PENGALOKASIAN NILAI KEDALAM KE MASING – MASING CENTROID

Asal Negara	C1	C2	C3	Keanggotaan *Centroid*
China	2	10	17	C1
Indonesia	0	12	19	C1
Canada	6	4	13	C2
Australia	16	4	3	C3
German	11	1	8	C2
Nigeria	0	12	19	C1
Mesir	19	7	0	C3
Jepang	19	7	0	C3
Iraq	8	7	11	C2
Brunei	9	3	10	C2
Singapore	22	10	3	C3

Tabel IV (sambungan)

Asal Negara	Keanggotaan	C1	C2	C3
China	C1	3		
Indonesia	C1	1		
Canada	C1	7		
Australia	C3			17
German	C2		12	
Nigeria	C1	1		
Mesir	C3			20
Jepang	C3			20
Iraq	C2		9	
Brunei	C2		10	
Singapore	C3			23
Rata – rata :		**3**	**10.33333333**	**20**

TABEL IV
TITIK PUSAT CENTROID ITERASI PERTAMA

Pusat *Centroid Pertama*	
C1	3
C2	10.33333333
C3	20

TABEL V
HASIL NILAI CENTROID ITERASI PERTAMA

Asal Negara	Keanggotaan *Centroid*	C1	C2	C3
China	C1	3		
Indonesia	C1	1		
Canada	C2		7	
Australia	C3			17
German	C2		12	
Nigeria	C1	1		
Mesir	C3			20
Jepang	C3			20
Iraq	C2		9	
Brunei	C2		10	
Singapore	C3			23
Rata – rata :		**1.666666667**	**9.5**	**20**

TABEL VI
TITIK PUSAT CENTROID ITERASI KEDUA

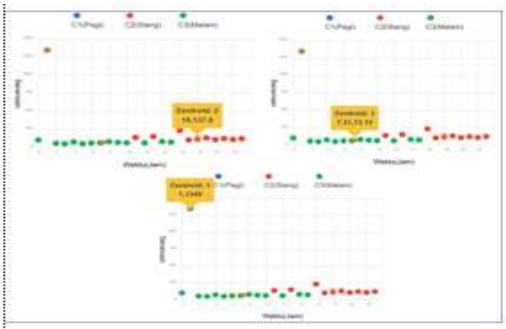
Pusat *Cluster Baru*	
C1	3
C2	10.33333333
C3	20

TABEL VII
HASIL NILAI PUSAT CENTROID KEDUA

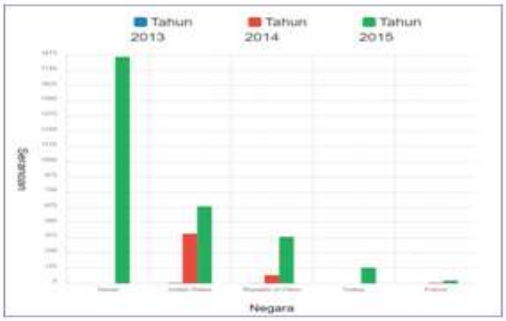
Asal Negara	Keanggotaan *Centroid*	C1	C2	C3
China	0	3		
Indonesia	2	1		
Canada	4		7	
Australia	14			17
German	9		12	
Nigeria	2	1		
Mesir	17			20
Jepang	17			20
Iraq	6		9	
Brunei	7		10	
Singapore	20			23
Rata – rata :		**1.666666667**	**9.5**	**20**

G. *Analisa dan Hasil*

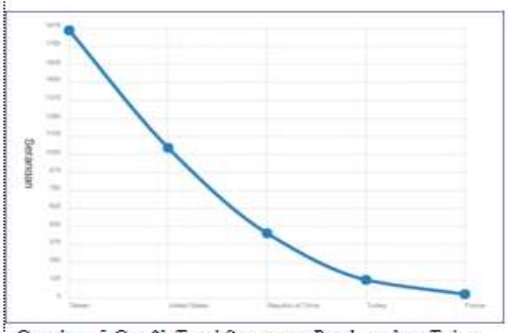
Gambar-3 menjelaskan tentang tiga titik *centroid* waktu C1 (pagi) = 1, C2 (Siang) = 12, dan C3 (malam) = 20. Visual grafik menunjukkan bahwa serangan DDoS memuncak pada pukul 01:00 pada waktu pagi (dini) hari dengan jumlah 1349 serangan.



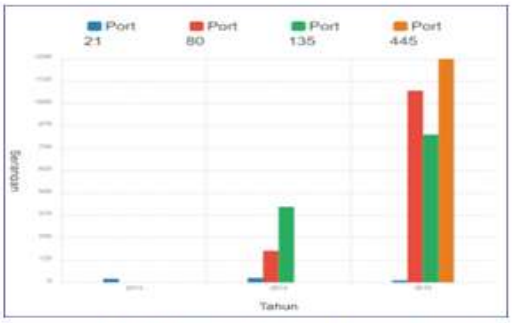
Gambar. 3 Visualisasi K Means



Gambar. 4 Grafik Jumlah Serangan Berdasarkan Negara



Gambar. 5 Grafik Total Serangan Berdasarkan Tahun



Gambar. 6 Grafik Total Serangan Berdasarkan Tahun

D. Kesimpulan

Public cloud computing merupakan teknologi jaringan yang paling rentan terhadap serangan DDoS. DDoS merupakan jenis serangan yang bekerja secara terus menerus untuk meruntuhkan layanan dan membanjiri *bandwitdh*. Penulis mengusulkan *honeypot* yang bersifat seperti server *mirror* (menerima, melindungi, merekam setiap aktivitas *attacker*, dan menyimpan ke *datalog)*. Kemudian datalog (kurun waktu dari 2013 hingga 2015 yang diambil 1 hari dari masing - masing tahun) ini dianalisis menggunakan *K – Means Clustering* untuk mengenali serangan DDoS dan *cyber profilling* berdasarkan *timestamp*. Dikategorikan kedalam tiga titik *centroid* C1(pagi), C2 (malam), dan C3 (siang) Waktu Indonesia Barat (WIB) ke dalam visual grafik dalam *dashboard*. Hasil menunjukkan bahwa serangan DDoS semakin meningkat secara signifikan pada tahun 2015 dari kurun waktu sebanyak 1250 serangan dari negara Taiwan yang tertuju ke *port destination* (Indonesia). Berdasarkan

time stamp serangan DDoS banyak diketahui masuk pada jam 01:00 pagi (dini hari) pada *port* UDP 445 (*service* SMB).

DAFTAR PUSTAKA

[1] Charles Lim, Mario Marcello, Andrew Japar, Joshua Tommy, & I Eng Kho. (2014). Development of Distributed Honeypot Using Raspberry Pi. *International Conference on Information, Communication Technology and System*, (January 2016).

[2] Gunawan. (2017). *Smart City for Smart Indonesia.* Yogyakarta: Fakultas Ekonomi dan Bisnis Universitas Gajah Mada.

[3] Hidayat, J. (2014). *CEH (Certified Ethical Hacker) 500% Illegal.* Yogyakarta: Jasakom.

[4] Hikmatyar, M., Prayudi, Y., & Riadi, I. (2017). Network Forensics Framework Development using Interactive Planning Approach. *International Journal of Computer Applications*, *161*(10), 41–48. https://doi.org/10.5120/ijca2017913352

[5] Hussein, M. K., Bin Zainal, N., & Jaber, A. N. (2016). Data security analysis for DDoS defense of cloud based networks. *2015 IEEE Student Conference on Research and Development, SCOReD 2015*, 305–310. https://doi.org/10.1109/SCORED.2015.7449345

[6] Jiang, C. B., Liu, I. H., Chung, Y. N., & Li, J. S. (2016). Novel Intrusion Prediction Mechanism Based on Honeypot. *INTERNATIONAL JOURNAL OF NETWORK MANAGEMENT Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/nem.1923*, 1-20

[7] Li, Z., Pan, H., Liu, W., Xu, F., Cao, Z., & Xiong, G. (2017). A network attack forensic platform against HTTP evasive behavior. Journal of Supercomputing, 73(7), 3053–3064. https://doi.org/10.1007/s11227-016-1924-3

[8] Mahajan, S., Adagale, A. M., & Sahare, C. (2016). Intrusion Detection System Using Raspberry PI Honeypot in Network Security. *International Journal of Scientific and Engineering Research- IJESC*, *6*(3), 2792–2795. https://doi.org/10.4010/2016.651

[9] Pratama, I. P. (2014). *Smart City Beserta Cloud Computing dan Teknologi - Teknologi Pendukung Lainnya.* Bandung: Informatika.

[10] Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB.* Yogyakarta: ANDI.

[11] Ramadhan, A., & Efendi, Z. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling, 18–19

[12] Somani, G., Gaur, M. S., Sanghi, D., Conti, M., Rajarajan, M., & Buyya, R. (2017). Combating DDoS attacks in the cloud: Requirements, trends, and future directions. *IEEE Cloud Computing*, *4*(1), 22–32. https://doi.org/10.1109/MCC.2017.14

[13] Sulianta, F. (2016). *Komputer Forensik Melacak Kejahatan Digital.* Yogyakara: ANDI.