Tugas Pendahuluan Proyek Akhir

PEMBUATAN DATA ANALYTICS PADA APLIKASI MATA GARUDA MENGGUNAKAN MULTI BRO IDS SENSOR

ABID FAMASYA ABDILLAH 2110131016

D4 TEKNIK INFORMATIKA
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA
2016

A. JUDUL PROYEK AKHIR

Pembuatan Data Analytics pada Aplikasi Mata Garuda Menggunakan Multi Bro IDS Sensor.

B. PENDAHULUAN

Sejak ditemukan pertama kalinya, internet telah digunakan oleh berbagai pihak sebagai sarana komunikasi massal karena sifatnya yang efisien dan universal. Data dari Statistica, perusahaan statistik terkemuka mencatat bahwa 3,5 milyar manusia telah terkoneksi aktif dengan internet, dimana 104 juta diantaranya berasal dari Indonesia. Seiring dengan penggunaan yang masif ini, ancaman terhadap pencurian informasi yang dikirimkan melalui internet juga meningkat secara signifikan. Akibatnya, para ahli keamanan jaringan harus menggunakan sebuah sistem pencegah serangan untuk meminimalisir dampak buruk serangan tersebut. Sistem ini banyak dikenal sebagai *Intrusion Detection System* (IDS).

Intrusion Detection System adalah sebuah sistem yang dapat digunakan untuk mendeteksi intrusi dalam sebuah sistem atau jaringan. Intrusi adalah sebuah aktivitas tidak sah atau tidak diinginkan yang mengganggu kerahasiaan, integritas dan atau ketersediaan dari informasi yang terdapat di sebuah sistem. Secara umum, sistem IDS berfungsi sebagai sebuah aplikasi yang memonitor lalu lintas data pada jaringan. Ketika terjadi suatu upaya intrusi, IDS akan menyimpannya kedalam *log file* yang hasilnya dapat dianalisa lebih lanjut.

Salah satu karakteristik infrastruktur internet di Indonesia adalah banyaknya jumlah NAP (*Network Access Point*) dan ISP (*Internet Service Provider*). Menurut data dari Kementrian Komunikasi dan Informatika pada 23 Oktober 2013, jumlah NAP adalah 30 penyelenggara, sedangkan ISP sebanyak 286 penyelenggara. Konsekuensinya, pemerintah kesulitan dalam monitoring trafik dan keamanan internet Indonesia. Mengatasi hal itu, aplikasi Mata Garuda yang telah dibangun sejak 2014 menjadi salah satu cara untuk mengatasi masalah monitoring trafik Indonesia.

Mata Garuda adalah sebuah aplikasi monitoring trafik yang menggunakan IDS yang dipasang di sensor-sensor NAP sebagai basisnya. Topologi awal aplikasi ini hanya dapat mendeteksi intrusi dalam satu jaringan dengan bantuan beberapa sensor. Sensor-sensor tersebut berfungsi untuk mengambil paket jaringan yang akan diolah

oleh *defense center*. Namun dengan bertambahnya sensor, maka trafik yang harus di-proses semakin besar. Akibatnya, daya komputasi di sisi *defense center* juga harus besar. Selain itu, dengan terjadinya perubahan model bisnis internet menuju komputasi awan (*cloud computing*), maka terjadi pula perubahan paradigma dimana sumberdaya komputasi infrastruktur jaringan dan server menjadi infrastruktur publik untuk digunakan bersama. Oleh karena itu, harus dilakukan pengambangan terhadap aplikasi Mata Garuda agar dapat menyesuaikan dengan model bisnis dan besarnya data yang akan diolah oleh aplikasi ini kedepannya.

C. PERUMUSAN MASALAH

Mata Garuda sebagai aplikasi utama dalam sistem *monitoring* trafik di Indonesia nantinya tidak bisa berdiri sendiri sebagai pusat pengolahan data insiden dan infrastruktur *monitoring* trafik. Dengan model *cloud computing* yang makin marak, pengembangan Mata Garuda tidak hanya sebagai pengolah data dari sekumpulan sensor NAP, tetap lebih menjadi *resource pool* dari aplikasi *agent* Mata Garuda yang tersebar di layanan *cloud computing*. Sehingga Mata Garuda harus bisa menyediakan platform analisa intrusi universal dengan data yang diolah dari berbagai aplikasi *agent* di banyak layanan tersebut.

D. TINJAUAN PUSTAKA

Tinjauan pustaka ini membahasa tentang teori-teori penunjang dalam penyelesaian proyek akhir ini. Beberapa teori penunjang tersebut adalah :

1. *Intrusion Detection System*

   Intrusion Detection System atau yang dikenal sebagai IDS adalah sebuah aplikasi yang memonitor jaringan dari aktivitas mencurigakan baik dari dalam maupun luar jaringan. Setiap aktivitas yang terdeteksi sebagai pelanggaran akan dilaporkan kepada Administrator melalui *Security Information and Event Management* (SIEM). Berdasarkan letak deteksinya ada dua tipe IDS yang ada, yaitu *Network-based IDS* (NIDS) dan *Host-based IDS* (HIDS). NIDS bekerja dengan cara melakukan monitoring terhadap jaringan, sedangkan HIDS bekerja hanya pada sistem jaringan internal suatu komputer, tidak pada paket jaringan eksternalnya. Sedangkan berdasarkan tipe deteksinya ada beberapa jenis, yaitu

signature-based, *anomaly-based* dan *hybrid detection*. Cara kerja IDS adalah dengan menganalisa paket data yang diterima dan hanya melakukan peringatan ketika ada aktivitas mencurigakan.

2. Bro

   Bro merupakan *Network-based Intrusion Detection System* (NIDS) yang juga bersifat *open source* dan populer. Dikembangkan oleh Vern Paxson pada 1998, Bro memiliki kemampuan untuk melakukan analisis *multi-layer*, *behavioral monitoring*, *policy enforcement*, dll. Perbedaan Bro dengan Snort terletak pada algoritma pendeteksiannya dimana Bro menggunakan *anomaly-based detection*. *Anomaly-based detection* menggunakan model statistik untuk mendeteksi serangan. Metode ini memiliki kelebihan dimana pendeteksian tidak terpaku pada signature yang didefinisikan sebelumnya, sehingga tipe-tipe serangan baru dapat terdeteksi. Selain itu, Bro akan menghasilkan log terhadap tiap paket data berdasar protokolnya dan juga mendukung penyesuaian terhadap file log mandiri. Namun kelemahannya, pendekatan dengan metode statistik ini memiliki *false alarm rate* yang cukup tinggi.

3. *Data analytics*

   Data analytics merupakan sebuah ilmu yang menggali data dengan tujuan untuk mendapatkan informasi terhadap data tersebut. *Data analytics* telah digunakan oleh berbagai perusahaan dan organisasi lain untuk membuat keputusan terbaik dengan cara membandingkan hipotesis dengan data yang ada. Selain itu, luaran yang diharapkan tidak hanya berguna untuk mencari pola dan pengetahuan dari sebuah data, namun juga mencari tahu apakah sebuah hipotesis benar atau tidak.

4. Apache Hadoop

   Apache Hadoop adalah *framework* perangkat lunak *open source* yang dapat digunakan sebagai *distributed storage* dan *distributed processing* data yang besar. Aplikasi ini populer digunakan karena sifatnya yang open source sehingga

siapa saja dapat berkontribusi dalam pengembangannya. Dalam basis frameworknya, Hadoop menggunakan HDFS (Hadoop Distributed File System) sebagai sistem penyimpanannya. HDFS ini dibangun berdasarkan konsep MapReduce dan Google File System yang dikembangkan Google. Sehingga kapabilitas Hadoop tidak perlu dilakukan lagi karena konsep yang dipakai telah terbukti efektif dalam aplikasi data skala besar. Hadoop sendiri menggunakan bahasa Java sebagai basis pengembangannya, meskipun juga memiliki *framework* pengembangan dengan bahasa lain seperti Python, Scala, dll.

E. PENELITIAN TERKAIT

Berikut adalah penelitian yang pernah dilakukan dan relevan dengan proyek akhir ini.

1. Bro: A System for Detecting Network Intruders in Real-Time oleh Vern Paxson [1]

   Penelitian ini merupakan publikasi awal terhadap Bro IDS yang dibuat oleh peneliti di Lawrence Berkeley National Laboratory, tempat dimana Bro dibuat. Dalam penelitian ini dijabarkan sistem dan komponen-komponen penyusun Bro secara menyeluruh sesuai dengan hierarkinya. Bagian-bagian tersebut dijelaskan sebagai berikut :

   (a) Libpcap

   Adalah system independent library yang digunakan sebagai *packet capture* di sebuah sistem. Bro menggunakan libpcap yang memiliki kemampuan *kernel level network monitoring* untuk melakukan pengambilan *network packet* dan melakukan inspeksi terhadap paket tersebut.

   (b) Event engine

   Merupakan komponen dari Bro yang berfungsi untuk memproses data TCP dan UDP yang telah di *capture* oleh libpcap untuk dilakukan *integrity check*. Jika ada sebuah kegagalan dalam *checking* tersebut, maka Bro akan membangkitkan *event* tertentu.

   (c) Policy script interpreter

Setelah *event engine* selesai memproses paket, maka *policy script interpreter* akan mengecek apakah ada *event* yang dibangkitkan. Jika ada, maka komponen ini akan memprosesnya dan menuliskan kedalam *log files*.

Berbagai macam *library*, bahasa pemrograman hingga cara kerja pemrosesan data lain dijelaskan secara terperinci oleh penulis. Sehingga penelitian ini telah menjadi salah satu acuan paling populer dan lengkap dalam pemrosesan data maupun *log analysis* menggunakan Bro.

2. IDS Log Analisis Menggunakan Hadoop dan Mahout untuk Data Mining Pada Mata Garuda oleh M. Hisyam, F. A. Saputra and J. Akhmad [2]

Penelitian ini mencoba untuk melakukan pemrosesan Snort *log file* dengan menggunakan prinsip *Big Data* di aplikasi Mata Garuda. Dengan sistem terdistribusi, metode *data mining* dilakukan terhadap data geolocation untuk mendapatkan lokasi serangan yang terjadi. Penulis menggunakan UDTF untuk melakukan *query* serta membandingkannya dengan *join query*. Sedangkan algoritma yang diterapkan dalam proses *mining* tersebut adalah K-means *clustering* untuk mendapatkan *cluster* dari GeoIP serangan. Hasilnya, UDTF mampu meruduksi waktu komputasi menjadi 0.08 detik daripada *join query* yang memakan waktu 3561 detik. Pada penelitian ini juga membuktikan bahwa penggunaan *distributed processing* untuk mengolah data Mata Garuda merupakan pilihan yang tepat karena data dapat diolah lebih cepat.

3. Inter-Domain Stealthy Port Scan Detection through Complex Event Processing oleh Leonardo Aniello, Giorgia Lodi dan Roberto Baldoni [3]

Penelitian ini berfokus pada pendeteksian *port scanning* yang biasanya terjadi pada skala *enterprise* dengan pendekatan *Complex Event Processing* (CEP). CEP adalah suatu metode yang melakukan suatu aksi berdasarkan berbagai hal yang terjadi sebelumnya. Penulis memfokuskan pada *port scanning* karena *port scanning* sendiri merupakan langkah awal dalam proses *system hacking*. Penulis membuat sebuah algoritma yang dinamakan R-SYN yang

melakukan deteksi terhadap (i) *half open connections*, (ii) *horizontal and vertical port scans* dan (iii) *entropy-based failed connections* yang kemudian menggabungkan ketiganya untuk memberi *rank* terhadap upaya SYN yang gagal. Ketika upaya tersebut melewati *threshold* tertentu, maka sistem ini akan memasukkan IP address *client* kedalam scanner list.

Setelah deteksi dilakukan, langkah selanjutnya yang dilakukan penulis adalah menggabungkan berbagai domain sebagai *data source* dari Esper, aplikasi yang dikembangkan penulis. Dari hasil penggabungan ini didapat performa algoritma yang dikembangkan penulis mampu mendeteksi upaya intrusi menggunakan berbagai dataset hingga 100%. Hasil analisa penulis juga menunjukkan bahwa *entropy correction* berpengaruh sangat besar dalam deteksi intrusi berdasar TCP flag ini.

4. Survey on Data Mining Techniques to Enhance Intrusion Detection oleh Deepthy K Denatious dan Anita John [4]

Pada paper ini dilakukan sebuah studi tentang metode-metode *data mining* yang bisa digunakan dalam pengolahan hasil deteksi intrusi. Pada awalnya penulis memaparkan proses yang dilakukan *intrusion system* secara umum yang meliputi : (i) *Data acquisition* sebagai proses pengambilan data, (ii) *Data preprocessor* untuk *data cleansing, integration* dan *reduction* (iii) *Data mining module* sebagai proses penyimpanan dan pengambilan *knowledge* dari data (iv) *Intrusion detection module* untuk mendeteksi intrusi (v) *Manager interface* module sebagai *interface* dan output akhir dari deteksi.

Di bagian akhir penelitian ini juga disebutkan tiga jenis teknik data mining (*classification, clustering*, dan *association rule*) serta bagaimana cara kerja metode tersebut dalam proses mining data intrusi. Penulis menyimpulkan bahwa metode *clustering* adalah metode yang paling cocok dalam intrusi mengingat besarnya data yang akan diolah oleh sistem dibanding klasifikasi.

5. Mendapatkan Dataset Rule Network dan Melakukan Ekstraksi Data Menggunakan Bro IDS oleh Mahbub, Ferry Astika Saputra dan Akhmad Alimudin [5]

Penelitian ini membahas tentang pembuatan dataset untuk penelitian intrusi internet. Karena keterbatasan data intrusi dalam dataset yang ada (KDD, DARPA, GUREKDDCUP), peneliti merasa perlu membuat dataset baru dengan fitur dan atribut yang diperbanyak. Penulis menggunakan Bro IDS dalam packet decoding, melakukan percobaan intrusi terkondisi lalu melakukan pelabelan terhadap data log dari paket tersebut menggunakan *Support Vector Machine*. Data yang berhasil diberi label merupakan data SSH dan FTP bruteforce. Pengujian yang dilakukan oleh penulis menggunakan WEKA menghasilkan *correctly classification* mencapai 92% dan semakin tinggi dengan bertambahnya fitur yang diolah. Namun dibandingkan dengan GUREKDDCUP, dataset yang dibuat penulis masih dibawah performa GUREKDDCUP. Hal ini diakibatkan karena data dari GUREKDDCUP jauh lebih banyak dan variatif dibanding dataset yang dihasilkan penulis.

F. TUJUAN PROYEK AKHIR

Tujuan yang ingin dicapai pada proyek akhir ini adalah membuat suatu modul *data analytics* pada aplikasi Mata Garuda yang menggunakan data *log* dari multi Bro-IDS.

G. KONTRIBUSI PROYEK AKHIR

Hasil dari proyek akhir ini dapat digunakan sebagai modul *data analytics* dalam aplikasi Mata Garuda. Sehingga aplikasi Mata Garuda tidak hanya mampu untuk menampilkan intrusi terhadap jaringan, namun juga dapat menyediakan platform terhadap analisa data intrusi dari beragam IDS sensor.

H. METODE PROYEK AKHIR

Untuk menyelesaikan proyek akhir ini langkah-langkah yang diambil ialah :

1. Studi Literatur

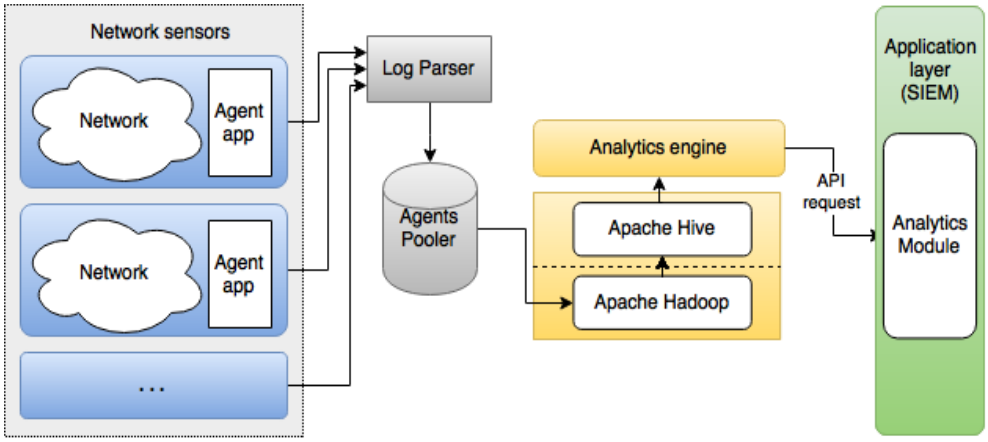
   Studi literatur merupakan langkah awal dalam pengerjaan proyek akhir ini. Tahapan ini merupakan tahap yang penting untuk memelajari teori-teori serta konsep teknis maupun non-teknis yang menunjang dalam pengerjaan proyek akhir ini.

2. Pengumpulan data

   Langkah selanjutnya dari penelitian ini adalah pengumpulan data log file dari Bro sebagai bahan analisa awal. Setelah data terkumpul, maka data tersebut akan diolah dengan teknik *text extraction* untuk membentuk algoritma dalam log parser.

3. Perancangan sistem

   Setelah tahap studi literatur, langkah yang diambil selanjutnya adalah melakukan perancangan sistem yang berfokus pada pembuatan *platform* analisa log file Bro dari banyak sensor (multisensor) ke dalam sistem Mata Garuda. Berikut adalah desain sistem yang akan akan dibuat.



Gambar 1. Desain sistem

Pada diagram sistem tersebut terdapat dua bagian utama dalam sistem, yaitu *network sensors* dan *application layer* (SIEM / *Security Information and Event Management*) dimana Mata Garuda berada. *Network sensors* adalah sekumpulan *host* jaringan dimana sensor Mata Garuda ditempatkan. Pada *network sensors* ini terdapat IDS yang akan mengolah *network packet* yang melewatinya. Diantara *network sensors* dan SIEM terdapat bagian untuk mengolah data sensor IDS. Setelah data diolah, maka akan ditampilkan pada *application layer* yang telah ditambahkan *analytics module*.

Pada awalnya, setiap paket data yang melewati jaringan *host* akan melalui

proses pemeriksaan oleh Bro yang telah terpasang pada tiap provider. Proses yang dilakukan oleh Bro ini mencakup *packet decoding* serta pemeriksaan oleh *detection engine* yang menghasilkan berbagai file *log* seperti yang dihasilkan Tabel 1. Hasil pemeriksaan oleh IDS tersebut akan dibaca oleh *agent app* pada *interval* waktu tertentu yang telah didefinisikan sebelumnya.

Tabel 1. Log file yang dihasilkan oleh Bro

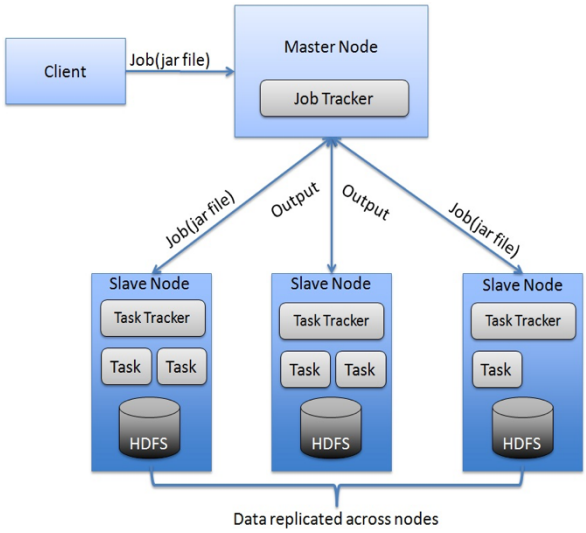
Network protocol log files			
Name	Description	Name	Description
conn.log	TCP/UDP/ICMP connections	ssh.log	SSH connections
dhcp.log	DHCP leases	ssl.log	SSL/TLS handshake info
dnp3.log	DNP3 requests and replies	syslog.log	Syslog messages
dns.log	DNS activity	files.log	File analysis results
ftp.log	FTP activity	notice.log	Bro notices
http.log	HTTP requests and replies	weird.log	Unexpected net-work-level activity
irc.log	IRC commands and responses	…	…

Meski banyak *log file* yang dihasilkan oleh Bro, namun pada sistem ini hanya akan mengambil *file* conn.log dan log lain yang bersesuaian untuk dianalisa. Hal ini dikarenakan karena *log file* tersebut merupakan file *log* mentah yang memuat seluruh paket data yang di *capture* oleh Bro sebelum diterapkan policy terhadap paket tersebut. Sistem ini tidak menggunakan log yang dihasilkan oleh notice.log untuk menghindari tidak konsistennya *notice rule* yang didefinisikan oleh administrator jaringan yang berbeda-beda.

Dalam interval waktu tertentu, agent app ini akan mengirimkan data menuju *agent pooler* dengan melewati *log parser*. *Log parser* adalah program yang

melakukan ekstraksi *log file* dengan konsep ETL (*extraction, transformation, loading*) untuk dilakukan normalisasi informasi yang diperoleh dari sensor dan mengubah bentuknya menjadi JSON untuk kemudahan pembacaan.

Agent pooler merupakan sebuah *service application* yang merupakan bagian dari Mata Garuda dan berfungsi untuk mengumpulkan (*pooling*) data yang dikirimkan oleh berbagai *agent apps* terlebih dahulu sebelum disimpan di *data storage*.



Gambar 2. Arsitektur distribusi Hadoop

Proses selanjutnya adalah menyimpan data menuju *data storage* (Hadoop) setelah data telah dilakukan normalisasi oleh *log parser* dan dikumpulkan oleh *agent pooler*. Pada Hadoop ini terjadi proses *batch processing* serta replikasi data menuju sejumlah slave server. Replikasi ini ditujukan untuk meningkatkan availibility data ketika dilakukan proses pengambilan data nanti.

Ketika end user ingin melakukan *request* data dari Mata Garuda, maka akan dilakukan *data ingestion* (pengambilan data) dari Hadoop dengan menggunakan Apache Hive sebagai *query language*. Pengambilan data dilakukan dengan melakukan API request terhadap *analytics engine*.

Pada analytics engine akan dilakukan proses *mining* untuk memperoleh suatu analisa dari data log *packet capture*. Pada awalnya *analytics engine* akan melakukan *loading* data dari Hadoop untuk mengambil data log yang tersimpan berdasarkan *packet timestamp*. Selanjutnya data tersebut ini diolah dengan algoritma tertentu untuk mendapatkan suatu prediksi intrusi. Prediksi intrusi yang dimaksud seperti *SSH bruteforce*, *FTP attack* dan *Denial of Service (DOS) attack*. Prediksi tersebut dapat diambil dilakukan berdasar analisa dari *conn_state field* dan *field-field* lain dari Bro *logs*. Dari data tersebut maka dapat diketahui berbagai analisa seperti waktu serangan, lokasi serta durasinya.

Dalam prosesnya, seluruh data dari *network sensors* akan digunakan dalam proses ini. Karena semakin banyak data yang diperoleh, gambaran pola intrusi yang didapat makin jelas. Hasil akhir dari proses *mining* ini akan ditampilkan pada *report* dan *data analytics* pada *application layer* agar diketahui oleh pihak-pihak yang berkepentingan.

4. Implementasi

   Langkah selanjutnya adalah melakukan implementasi sistem yang telah dirancang sebelumnya. Implementasi ini dilakukan pada server yang dibuat serupa dengan arsitektur Mata Garuda.

5. Pengujian dan analisa

   Pada tahap ini dilakukan pengujian terhadap implementasi desain sistem dengan menggunakan dataset yang disesuaikan dengan intrusi sesungguhnya. Pengujian dilakukan untuk mengetahui apakah model yang dibuat telah sesuai dengan keluaran yang diharapkan. Keluaran yang diharapkan dari sistem ini adalah terintegrasinya log file dari Bro serta modul data analytics berhasil dibuat. Selain keluaran, juga akan dilakukan testing untuk mengetahui performa dari implementasi.

I. JADWAL PELAKSANAAN

Kegiatan	Bulan												
	8	9	10	11	12	1	2	3	4	5	6	7	8
Proposal ProyekAkhir													
Pengambilan Data													
Implementasi/Coding													
Debugging													
Testing &Analisa													
Penyusunan Buku PA													
Sidang PA													

J. PERSONALIA PROYEK AKHIR

- Mahasiswa

  Nama : Abid Famasya Abdillah

  NRP : 2110131016

  Jurusan : Diploma IV - Teknik Informatika

  Agama : Islam

  Jenis kelamin : Laki-laki

- Dosen Pembimbing 1

  Nama : Ferry Astika Saputra, ST, M.Sc

  NIP : 197708232001121002

Departemen : Teknik Informatika

Bidang keahlian : Computer Network, Network Security

- Dosen pembimbing 2

Nama : Iwan Syarif, S.Kom., M.Sc., Ph.D.

NIP : 196904041995121002

Departemen : Teknik Informatika

Bidang keahlian : Data Mining, Machine Learning

K. PERKIRAAN BIAYA PROYEK AKHIR

No	Uraian	Jumlah	Harga Satuan(Rp)	Total (Rp)
1	Kertas A4 HVS 80gr	1 rim	40.000	40.000
2	Kertas A5	1 rim	40.000	40.000
3	Tinta Printer	4	25.000	100.000
4	Internet	8 bulan	100.000	700.000
5	Jilid	5	15.000	75.000
6	CD Dokumentasi	5	5.000	25.000
TOTAL PENGELUARAN				**980.000**

L. DAFTAR PUSTAKA

[1] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-Time," *Seventh USENIX Security Symp.,* 1998.

[2] M. Hisyam, F. A. Saputra and J. Akhmad, "IDS Log Analisis Menggunakan Hadoop dan Mahout untuk Data Mining pada Mata Garuda," in *Jurnal Teknik Informatika dan Komputer PENS,* Surabaya, 2015.

[3] L. Aniello, G. Lodi and R. Baldoni, "Inter-Domain Stealthy Port Scan Detection through Complex Event Processing," in *13th European Workshop on Dependable Computing*, Pisa, 2011.

[4] D. K. Denatious and A. John, "Survey on Data Mining Techniques to Enhance Intrusion Detection," *International Conference on Computer Communication and Informatics (ICCCI -2012),* 2012.

[5] Mahbub, F. A. Saputra and A. Alimudin, "Mendapatkan Dataset Rule Network dan Melakukan Ekstraksi Data Menggunakan Bro IDS," in *Jurnal Teknik Informatika dan Komputer PENS*, Surabaya, 2015.