SKRIPSI

PEMBANGUNAN PERANGKAT LUNAK ENKRIPSI DATA UNTUK PENAMBANGAN DATA

Himawan Saputra Utama

NPM: 2015730009

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2019

UNDERGRADUATE THESIS

DATA ENCRYPTION SOFTWARE DEVELOPMENT FOR DATA MINING

Himawan Saputra Utama

NPM: 2015730009

DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2019

LEMBAR PENGESAHAN

PEMBANGUNAN PERANGKAT LUNAK ENKRIPSI DATA UNTUK PENAMBANGAN DATA

Himawan Saputra Utama

NPM: 2015730009

Bandung, 16 Mei 2019

Menyetujui,

Pembimbing

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji	Anggota Tim Penguji
Pascal Alfadian, M.Comp.	Husnul Hakim, M.T.

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

PEMBANGUNAN PERANGKAT LUNAK ENKRIPSI DATA UNTUK PENAMBANGAN DATA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 16 Mei 2019

Meterai
Rp. 6000

Himawan Saputra Utama
NPM: 2015730009

ABSTRAK

Pada era dijital ini, semakin banyak data dijital yang beredar dalam dunia teknologi. Data tersebut terdiri dari berbagai macam informasi terbuka ataupun rahasia. Salah satu cara untuk menjaga privasi data rahasia adalah dengan melakukan enkripsi. Enkripsi adalah suatu proses perubahan bentuk data sedemikian rupa sehingga hanya dapat dikembalikan ke bentuk asalnya oleh orang-orang tertentu saja.

Namun, teknik enkripsi tersebut menimbulkan masalah baru. Data dijital yang beredar dalam dunia teknologi, bukanlah data yang memiliki fungsi untuk dibaca saja. Data tersebut adalah data mentah yang masih dapat diproses atau masih dapat dilakukan penambangan data. Dengan menggunakan teknik enkripsi umum, setiap kali data akan diproses, seluruh data harus didekripsi terlebih dahulu. Proses dekripsi ini, adalah proses yang tidak sebentar terutama bila data berukuran sangat besar dan pemrosesan sering dilakukan. Oleh karena itu, timbullah pemikiran teknik enkripsi yang dapat melakukan perhitungan tanpa perlu didekripsi terlebih dahulu.

Penelitian ini bertujuan untuk menganalisis teknik-teknik enkripsi yang ada untuk menyelesaikan masalah tersebut. Terdapat 3 skema enkripsi yang dianalisis pada penelitian ini, yaitu skema *Simple XOR*, *ElGamal homomorphic* dan *Order Preserving Encryption*. Setiap skema ini diimplementasikan dan diuji coba dilakukannya operasi fungsi-fungsi perhitungan dasar dalam penambangan data yaitu, penjumlahan (*sum*), perkalian (*multiply*), modus, median, mean, pencarian sebuah nilai dan nilai-nilai dalam interval, serta fungsi *k-means*. Uji coba dilakukan dengan cara mengkeksekusi fungsi-fungsi tersebut pada data asli dan data terenkripsi setiap skema. Hasil pada data terenkripsi kemudian didekripsi menggunakan skema yang bersangkutan dan hasilnya dibandingkan dengan hasil pada data asli.

Dengan skenario pengujian tersebut, berikut adalah hasil pengujian masing-masing skema. Pada skema *Simple XOR*, fungsi yang dapat dilakukan dan menghasilkan hasil yang valid adalah fungsi pencarian modus dan pencarian sebuah nilai. Sedangkan *k-means* memiliki tingkat kesamaan dengan hasil pada data asli di atas 50%. Pada skema *ElGamal*, fungsi *sum*, *multiply*, modus, dan pencarian sebuah nilai dapat dilakukan. Tetapi, fungsi-fungsi tersebut sepenuhnya bergantung pada jenis homomorfik dan faktor acak. Sedangkan untuk fungsi *k-means* dihasilkan tingkat kesamaan dengan hasil pada data asli hanya sekitar 20% saja.

Terakhir, pada skema *Order Preserving Encryption*, fungsi yang dapat dilakukan adalah fungsi modus, median, pencarian sebuah nilai ataupun nilai-nilai dalam interval. Sedangkan untuk fungsi *sum* dan mean tidak dapat ditentukan, tetapi fungsi penjumlahan beberapa data sebenarnya dapat dilakukan. Untuk fungsi *k-means* dihasilkan tingkat kesamaan antara 50% hingga 90%.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa operasi fungsi pemrosesan data terenkripsi dapat dilakukan bergantung pada teknik enkripsi yang digunakan. Masing-masing skema dari ketiga skema yang telah diuji coba memiliki kemampuan dan keterbatasannya sendiri. Namun, skema *Order Preserving Encryption* adalah teknik enkripsi terbaik untuk dilakukannya penambangan data tanpa memerlukan dekripsi terlebih dahulu.

Kata-kata kunci: Enkripsi, Penambangan Data, *Simple XOR*, *ElGamal*, *Homomorphic Encryption*, *Order Preserving Encryption*

ABSTRACT

In the era where everything is digitized, there is a great and still increasing amount of digital data traversing the net. Those data consist of many public and secret informations. One of many techniques in information security which is used to maintain private data secrecy is called encryption. Encryption is the process to encode data in such a way that could only be returned to its original form by specific people.

But, there is a new problem arise from encrypting data. The data from before, are not a read only data. These data are raw data and can be processed further or used for data mining. Using data encryption, every time data are going to be processed, all of the data must be decrypted first. This decryption process is not a quick process. It's a long process, especially if the amount of data is huge. Because of that problem, arise new ideas on how to do data processing without the need to decrypt all the encrypted data first.

This research is aiming to find such encryption techniques that is able to solve the problem. There are three encryption scheme found and analyzed, which is Simple XOR encryption scheme, homomorphic attribute in ElGamal scheme, and Order Preserving Encryption. Those three schemes are implemented and a few basic functions in data mining are tested using encrypted data of each scheme.

Sum, multiply, modes, median, mean, search and interval search, and k-means clustering are the test functions. Each of these functions are executed on plain data and encrypted data of each scheme. Results on encrypted data are then decrypted and the results are compared to the results on plain data.

Based on the previously explained testing scenario, here are the test results of each encryption scheme. Using Simple XOR scheme, functions with valid results are modes and searching a value. While k-means resulting in similarity rate with results on plain data above 50%. Using ElGamal scheme, sum, multiply, modes and searching a value resulting a valid results. But those results are dependent to the homomorphic type and random factor of the scheme. While k-means resulting in 20% similarity rate.

Finally, for Order Preseving Encryption scheme, modes, median, searching a or many values in interval have a valid results. For sum and mean functions, can not be determined, but addition of a few data can surely be done and resulting in valid results. While k-means resulting in 50 to 90% similarity rate.

According to the experiment results, it is possible to execute data mining functions over encrypted data without the need to decrypt it first, dependent to the encryption scheme. Each scheme of the 3 tested schemes, has its own ability and limitations. But, among the 3, Order Preserving Encryption is the best scheme for encrypted data processing.

Keywords: Encryption, Data Mining, Simple XOR, ElGamal, Homomorphic Encryption, Order Preserving Encryption

Dipersembahkan untuk Tuhan Yesus Kristus, keluarga tercinta, dosen-dosen, teman-teman seperjuangan, segala pihak yang terlibat dalam penulisan skripsi ini, serta diri sendiri

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yesus Kristus, karena karunia-Nya sehingga skripsi berjudul "Pembangunan Perangkat Lunak Enkripsi Data untuk Penambangan Data" ini dapat selesai dengan baik. Selama penulisan skripsi ini, penulis banyak menghadapi kendala dan berbagai masalah, namun dengan berbagai dorongan dari lingkungan sekitar penulis, akhirnya skripsi ini dapat diselesaikan penulis. Oleh karena itu, penulis ingin mengungkapkan rasa terima kasih kepada orang-orang disekitar penulis, yaitu:

- Keluarga yang selalu memberikan doa dan dukungan mental serta materiil.
- Ibu Mariskha Tri Adithia selaku pembimbing skripsi yang telah memberikan dukungan, bimbingan serta kesabaran dalam menghadapi penulis selama penulisan skripsi ini.
- Bapak Pascal Alfadian dan Bapak Husnul Hakim selaku penguji yang telah memberikan kritik dan saran yang membuat skripsi ini lebih baik lagi.
- Segenap dosen, staf Tata Usaha, dan Pekarya yang terlibat dalam kegiatan pembelajaran selama penulisan skripsi ini.
- Seluruh teman-teman mahasiswa yang telah memberikan dorongan, masukan dan bantuan dalam penulisan skripsi ini.
- Segenap pihak lain yang ikut terlibat dalam penulisan skripsi ini.

Bandung, Mei 2019

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Saat ini, kehidupan manusia semakin hari semakin berorientasi kepada teknologi. Dalam kehidupan berteknologi terdapat begitu banyak data dijital yang dapat bersifat publik ataupun rahasia. Maka dari itu, keamanan informasi di era dijital ini seringkali menjadi fokus utama. Keamanan informasi merupakan sebuah bidang pembelajaran yang secara khusus membahas mengenai kerahasiaan data dan cara untuk menjamin kerahasiaan data. Salah satu cara yang digunakan dalam keamanan informasi adalah dengan melakukan enkripsi terhadap data yang bersifat rahasia. Teknik pengamanan data dengan cara melakukan enkripsi secara khusus terdapat dalam bidang kriptografi dari keamanan informasi. Enkripsi merupakan proses perubahan bentuk data menjadi bentuk sedemikian rupa sehingga hanya orang-orang tertentu saja yang dapat mengembalikannya ke bentuk aslinya [1]. Proses perubahan bentuk data kembali menjadi data aslinya disebut juga dengan dekripsi.

Data dijital merupakan sekumpulan data yang masih dapat diproses untuk mendapatkan suatu informasi penting. Salah satu bidang yang mempelajari teknik-teknik untuk melakukan ekstraksi informasi penting dari suatu kumpulan data adalah bidang penambangan data. Penambangan data biasanya dilakukan terhadap data yang disimpan dalam suatu mesin yang dimiliki pemilik data. Tetapi, dengan semakin berkembangnya *cloud storage*, banyak orang yang menyimpan datanya dalam *cloud storage* yang dimiliki oleh pihak lain yang dipercaya sang pemilik data. Tujuan utama dari *cloud storage* adalah demi kemudahan akses. Walaupun pihak pemilik *cloud storage* merupakan pihak yang dipercaya, tetap saja data yang disimpan dalam *cloud storage* biasanya berupa data yang tidak dienkripsi. Maka dari itu, keamanan informasi juga menjadi fokus banyak orang dalam menyimpan data dalam *cloud storage*.

Permasalahan tersebut dapat diselesaikan dengan cara melakukan enkripsi terhadap data terlebih dahulu sebelum disimpan dalam *cloud storage*. Tetapi, solusi tersebut menimbulkan permasalahan baru, yaitu, setiap kali seseorang akan mengolah data miliknya dalam *cloud storage*, ia harus pertama-tama mengambil seluruh data miliknya, melakukan dekripsi, lalu akhirnya mengolah data hasil dekripsi tersebut. Proses tersebut merupakan proses yang lama, terutama jika perhitungan sering dilakukan. Maka dari itu, timbullah pemikiran untuk menghilangkan proses tersebut. Salah satu cara yang telah dicetuskan adalah dengan melakukan perhitungan atau penambangan data terhadap data yang masih terenkripsi tanpa perlu dilakukan dekripsi terlebih dahulu.

Terdapat beberapa teknik enkripsi yang dapat diterapkan untuk menyelesaikan permasalahan tersebut. Salah satunya adalah teknik *order preserving encryption*. *Order preserving encryption* merupakan suatu teknik enkripsi yang mampu menghasilkan data terenkripsi yang terurut berdasarkan data aslinya [5]. Keterurutan data ini memungkinkan dilakukannya fungsi-fungsi basis data, seperti fungsi pencarian, *sum*, *indexing*, dsb.

Teknik lain yang dapat digunakan adalah teknik *homomorphic encryption*. *Homomorphic encryption* adalah teknik enkripsi yang memungkinkan dilakukannya komputasi terhadap data yang terenkripsi [4]. Komputasi tersebut berbentuk suatu fungsi terhadap data terenkripsi. Teknik ini akan menghasilkan hasil komputasi dalam bentuk yang terenkripsi. Hasil komputasi tersebut, jika

didekripsi akan menghasilkan hasil yang sama sebagaimana jika komputasi dilakukan pada data yang tidak terenkripsi.

Selain itu juga, akan diuji coba suatu skema enkripsi sederhana yaitu *simple XOR*. Skema ini adalah skema yang mengenkripsi data menggunakan suatu kunci konstan. Proses dekripsinya pun sama dengan proses enkripsi. Skema sederhana ini juga akan diuji coba dan dianalisis kemungkinannya untuk dilakukan operasi fungsi terhadap hasil enkripsi skema ini.

Pada skripsi ini, akan dibuat sebuah perangkat lunak yang dapat melakukan enkripsi terhadap suatu kumpulan data dan melakukan penambangan data terhadap data terenkripsi tersebut tanpa perlu melakukan dekripsi terlebih dahulu. Teknik *order preserving encryption*, *homomorphic encryption*, dan *simple XOR encryption* adalah teknik yang akan digunakan untuk membangun perangkat lunak. Perangkat lunak yang dibangun berfokus utama untuk melakukan enkripsi data dalam sebuah basis data menggunakan teknik enkripsi yang dipilih dan melakukan operasi penambangan data terhadap data terenkripsi tersebut.

1.2 Rumusan Masalah

Rumusan-rumusan masalah dari skripsi ini adalah sebagai berikut:

1. Bagaimana cara penyiapan data terenkripsi yang siap untuk ditambang?
2. Bagaimana cara melakukan penambangan data terhadap data yang terenkripsi?
3. Bagaimana cara mengimplementasikan teknik *order preserving encryption*, *homomorphic encryption* atau *simple XOR* beserta teknik penambangan data untuk membangun perangkat lunak?

1.3 Tujuan

Berdasarkan pada rumusan masalah, maka berikut adalah tujuan yang ingin dicapai dari skripsi ini, yaitu:

1. Mempelajari teknik *order preserving encryption*,*homomorphic encryption*, dan *simple XOR* untuk melakukan enkripsi data.
2. Mempelajari penambangan data terhadap data yang terenkripsi.
3. Membangun perangkat lunak yang mengimplementasikan teknik *order preserving encryption*, *homomorphic encryption* atau *simple XOR* beserta teknik penambangan data yang akan digunakan.

1.4 Batasan Masalah

Seperti yang telah disebutkan pada Subbab 1.1, berikut adalah batasan-batasan masalah dari penelitian ini:

- Struktur basis data yang digunakan dalam skripsi ini adalah basis data yang tersimpan secara lokal.
- Seluruh data yang diproses dalam skripsi ini adalah data yang bersifat numerik bilangan bulat atau data yang dapat direpresentasikan sebagai bilangan bulat.
- Penelitian ini berpusat pada apakah operasi fungsi dapat dilakukan terhadap hasil enkripsi, oleh karena itu, kekuatan masing-masing skema tidak akan dianalisis dan bukan merupakan titik berat penelitian ini.

1.5 Metodologi

Metodologi penelitian yang digunakan dalam skripsi ini adalah metodologi eksperimental. Dalam skripsi ini, akan digunakan dan diuji coba teori-teori yang mungkin dapat menyelesaikan masalah dari skripsi ini. Berikut adalah langkah-langkah metodologi penelitian yang dilakukan:

1. Melakukan studi literatur terkait teori-teori yang diperlukan.
2. Mengimplementasikan skema enkripsi *simple XOR*, *ElGamal*, dan *order preserving encryption*.
3. Melakukan eksperimen operasi fungsi penambangan data terhadap data terenkripsi masing-masing skema yang telah diimplementasikan.

1.6 Sistematika Pembahasan

Penelitian ini ditulis dalam beberapa topik pembahasan yang disusun dengan sistematika sebagai berikut:

Bab 1 Pendahuluan
Bab ini menguraikan tentang latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.

Bab 2 Landasan Teori
Bab ini menguraikan tentang teori-teori yang digunakan dalam penelitian ini. Bagian ini terdiri dari beberapa bagian, bagian pertama menguraikan tentang dasar-dasar dalam kriptografi. Bagian selanjutnya menjelaskan mengenai dasar dalam penambangan data dan beberapa operasi yang umum dilakukan. Bagian terakhir menguraikan tentang beberapa kriptosistem yang dirasa dapat digunakan dalam penelitian ini, seperti *order preserving encryption*, *homomorphic encryption*, dan *simple XOR*.

Bab 3 Analisis Masalah
Penguraian masalah yang dihadapi, serta menganalisis teori-teori yang dapat digunakan. Menggunakan analisis untuk menentukan kebutuhan dasar yang digunakan dalam pembangunan perangkat lunak.

Bab 4 Perancangan Perangkat Lunak
Menguraikan rancangan perangkat lunak untuk mengimplementasikan teori-teori yang digunakan.

Bab 5 Implementasi dan Pengujian
Bab ini akan berisi proses implementasi perangkat lunak berdasarkan analisa dan desain dalam perancangan menggunakan bahasa pemrograman Java ver. 8 dan menggunakan JavaFX untuk tampilan, MySQL ver. 15.1 untuk basis data, serta aplikasi MATLAB R2017b yang digunakan untuk melakukan *clustering k-means*. Selain itu juga akan dijabarkan mengenai pengujian dari perangkat lunak yang telah dibangun.

Bab 6 Kesimpulan dan Saran
Berisi kesimpulan dan saran dari hasil penelitian ini.