Analisis Frekuensi pada Teks Bahasa Indonesia Dan Modifikasi Algoritma Kriptografi Klasik

Galih Andana - 13507069

Program Studi Teknik Informatika, Institut Teknologi Bandung
Jalan Ganesha 10 Bandung
E-mail: xugalz_@hotmail.com

ABSTRAK

Saat ini sudah sangat banyak algoritma kriptografi modern yang ditemukan, tetapi algoritma kriptografi klasik pun masih sering dipakai untuk mengenkripsi suatu plainteks. Sebagai kriptanalis kita diharpkan untuk tidak melupakan teknik-teknik enkripsi maupun analisis kunci untuk algoritma tersebut.

Salah satu teknik pemecahan kunci di kriptografi klasik adalah analisis frekuensi, yaitu teknik untuk memperkirakan tingkat kemunculan huruf-huruf pada cipherteks dan disubstitusikan dengan huruf-huruf yang sering muncul pada dunia nyata. Namun, pada saat ini daftar frekuensi kemunculan huruf-huruf yang tersedia hanya untuk teks berbahasa Inggris saja.

Oleh sebab itu, pada makalah ini akan dibahas mengenai tingkat frekuensi kemunculan huruf-huruf pada teks Bahasa Indonesia yang dibutuhkan untuk melakukan analisis algoritma kriptografi klasik. Selain itu, pada makalah ini juga akan dibahas mengenai bagaimana cara memodifikasi algoritma kriptografi klasik, dalam hal ini *Vigenere Cipher,* agar tahan menghadapi serangan analisis frekuensi.

Kata kunci: *Klasik, Analisis Frekuensi, Modifikasi Vigenere, Indonesia.*

1. PENDAHULUAN

1.1. ALGORITMA ENKRIPSI KLASIK

Pada kriptografi klasik ada dua macam cara enkripsi yang dilakukan. Yang pertama adalah **transposisi**. Transposisi adalah mengubah susunan huruf pada plaintext sehingga urutannya berubah. Contoh yang paling sederhana adalah mengubah suatu kalimat dengan menuliskan setiap kata secara terbalik :

Plaintext : IBU AKAN DATANG BESOK PAGI
Ciphertext : UBI NAKA GNATAD KOSEB IGAP

Contoh transposisi yang sedikit lebih sulit adalah plaintext yang disusun dalam kelompok huruf yang terdiri dari beberapa kolom huruf, misalnya 5 kolom huruf :

I	B	U	A	K
A	N	D	A	T
A	N	G	B	E
S	O	K	P	A
G	I			

Tabel 1 - Transposisi

menjadi : IAASG BNNOI UDGK AAABP AKTEA. Ukuran kolom pada tabel ini dapat diubah-ubah sesuai kesepakatan antara pengirim dan penerima.

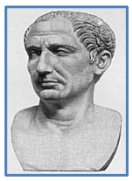
Cara kedua adalah cara **substitusi** yaitu setiap huruf pada plaintext akan digantikan dengan huruf lain berdasarkan suatu cara atau rumus tertentu. Ada dua macam substitusi yaitu *polyalphabetic substitution cipher* dan *monoalphabetic substitution cipher*. Pada *polyalphabetic substitution cipher*, enkripsi terhadap satu huruf yang sama bisa menghasilkan huruf yang berbeda sehingga lebih sulit untuk menemukan pola enkripsinya. Sedangkan pada *monoalphabetic substitution cipher*, satu huruf tertentu pasti akan berubah menjadi huruf tertentu yang lain.[1] Seperti contohnya ***Caesar-Cipher.***

Cipher klasik ini sangat mudah untuk dipecahkan. Banyak dari cipherteks klasik dapat didekripsi bahkan jika sang penyerang cukup mengetahui cipherteksnya saja. Hal ini menyebabkan mereka disebut rentan oleh *ciphertext-only-attack.* Cara penyerangannya dapat dengan menggunakan *brute-force attack*, yaitu mencoba semua kemungkinan kunci pada ke 26 huruf yang ada, maupun dengan **analisis frekuensi,** karena keseringan munculnya suatu huruf pada bahasa plainteks. Namun, hal tersebut telah diatasi oleh *polyalphabetic substitution cipher* seperti ***Vigenere-Cipher.*** Tapi tetap saja, metode Vigenere standar ini masih dapat dipecahkan dengan Metode **Kasiski.**

[1] sistem-keamanan-komputer.blogspot.com/2009/04

1.1.1. CAESAR-CIPHER

Dalam kriptografi, *Caesar-Cipher*, yang juga dikenal sebagai sandi Caesar, sandi geser, kode Caesar atau pergeseran Caesar, adalah salah satu teknik enkripsi yang paling sederhana dan dikenal luas. Ini adalah jenis penyandian substitusi di mana setiap huruf pada plaintext digantikan oleh sebuah huruf pada posisi lain dengan jarak yang tetap di alfabet. Sebagai contoh, dengan pergeseran 3, A akan digantikan oleh D, B akan menjadi E, dan seterusnya. Metode ini dinamai Caesar karena Julius Caesar menggunakannya untuk berkomunikasi dengan para jenderal.



Langkah enkripsi oleh sandi Caesar sering dijadikan bagian dari metode yang lebih kompleks, seperti Vigenere-Cipher, dan masih digunakan untuk siste aplikasi modern yaitu **ROT13**[2]. Seperti semua sandi substitusi alfabet tunggal lainnya, sandi Caesar dapat dengan mudah dipecahkan dan praktis tidak ada komunikasi yang menawarkan keamanan dengannya.

1.1.2. VIGENERE-CIPHER

Vigenere-Cipher adalah metode enkripsi teks alfabetik dengan menggunakan rangkaian berbagai *Caesar-Cipher* berdasarkan huruf-huruf dari kata kunci. Ia merupakan bentuk sederhana dari substitusi polyalphabetic.

Vigenere (dalam French : [viʒnɛːʁ]) cipher telah digambarkan kembali beberapa kali oleh orang yang berbeda. Metode ini awalnya digambarkan oleh ***Giovan Battista Bellaso*** pada tahun 1553 dalam buku *La cifra del. Sig. Giovan Battista Bellaso*, kemudian dibahas kembali oleh ***Blaise de Vigenere*** pada abad ke-19, dan sekarang justru dikenal luas sebagai "Sandi Vigenère".

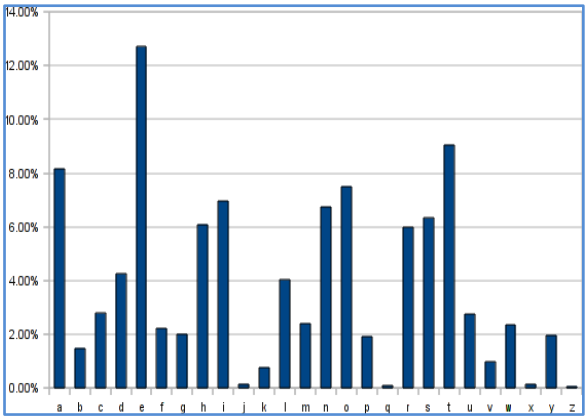
Kelebihan cipher ini adalah mudah untuk dipahami dan diterapkan, hal itu sering muncul bagi pemula yang sedang mencoba memecahkannya; Vigenere dikenal sebagai 'sandi yang tak terpecahkan' (bahasa Perancis : *le chiffre indéchiffrable*). Akibatnya, banyak orang telah mencoba menerapkan skema enkripsi yang pada dasarnya adalah Vigenere cipher, hanya untuk membuatnya pecah.[3]

Dalam Caesar-Cipher, setiap huruf dari alfabet digeser sepanjang beberapa tempat, misalnya pergeseran 3, A akan menjadi D, B akan menjadi E dan seterusnya. Sandi Vigenère terdiri dari beberapa sandi Caesar dalam urutan dengan nilai pergeseran yang berbeda.

1.2. ANALISIS FREKUENSI

Dalam kriptoanalisis, analisis frekuensi adalah studi tentang frekuensi huruf atau kelompok huruf dalam sebuah ciphertext. Metode ini digunakan sebagai bantuan untuk memecahkan cipher klasik.

Analisis frekuensi dibuat berdasarkan pada kenyataan bahwa, dalam setiap penggunaan bahasa tulisan, karakter-karakter tertentu dan kata dari kombinasi huruf terjadi dengan frekuensi yang berbeda-beda. Selain itu, ada karakteristik penyebaran huruf yang kira-kira sama untuk hampir semua sampel bahasa tertentu. Misalnya pada bahasa Inggris, E cenderung menjadi sangat umum, sedangkan X sangat jarang. Demikian pula, ST, NG, TH, dan QU adalah pasangan karakter yang sering muncul (disebut bigrams atau digraf), sedangkan QJ dan NZ sangat jarang. Frase "ETAOIN SHRDLU" mewakili ke-12 karakter yang paling sering muncul dari teks bahasa Inggris biasa.



Bagan 1- Analisis Frekuensi untuk Bahasa Inggris[4]

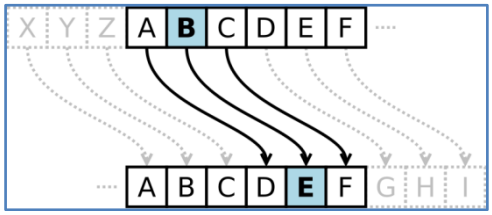
Dalam beberapa cipher, sifat-sifat seperti bahasa alami plaintext yang tersimpan dalam ciphertext, dan pola-pola ini memiliki potensi untuk dieksploitasi dalam serangan ciphertext-only karena kita dapat memprediksi karakter apa yang paling sering muncul di cipherteks yang bersangkutan. Dan karakter ini berkorespondensi dengan karakter yang paling sering muncul di Bahasa Plainteksnya.

2. PROSES ENKRIPSI DAN ANALISIS FREKUENSI STANDAR

2.1. CAESAR CIPHER

Transformasi di sini dapat diibaratkan dengan penyelarasan dua alfabet; sandi alfabet adalah alfabet biasa diputar kiri atau kanan sesuai jumlah posisi. Sebagai contoh, di sini terdapat sandi Caesar menggunakan rotasi 3 tempat ke kiri (parameter pergeseran = 3, digunakan sebagai kunci).

[2] Wobst, Reinhard. 2001. *Cryptology Unlocked*. Wiley.
[3] Smith, Laurence D. 1943. "Substitution Ciphers". *Cryptography the Science of Secret Writing: The Science of Secret Writing*.
[4] www.stealthcopter.com/blog/2010/01



Gambar 1 - Caesar Cipher

Plain : ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher : DEFGHIJKLMNOPQRSTUVWXYZABC

Ketika mengenkripsi, seseorang melakukan *Look-up* dari setiap huruf dari pesan dalam Tabel Plain dan menuliskan huruf yang sesuai dalam Tabel Cipher. Sedangkan cara memecahkannya dilakukan secara terbalik.

Ciphertext:
WKH TXLFN EURZQ IRA MXPSV RYHU WKH ODCB GRJ.
Plaintext:
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.

Teknik enkripsi ini juga dapat direpresentasikan dengan menggunakan aritmatika modular dengan mengubah huruf-huruf menjadi angka-angka, sesuai dengan skema, 'A'=0, 'B'= 1, ..., 'Z'=25. Enkripsi dari karakter x oleh pergeseran n digambarkan secara matematis sebagai :

$$E_n(x) = (x + n)\ mod\ 26$$

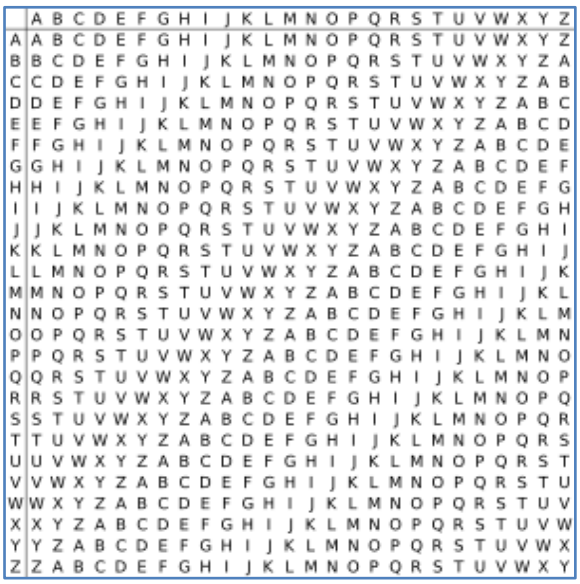
Sedangkan untuk Dekripsi dirumuskan sebagai :

$$D_n(x) = (x - n)\ mod\ 26$$

(Ada beberapa definisi untuk operasi Modulo. Dalam contoh di atas, hasilnya adalah dalam kisaran 0 ... 25. Namun, jika x + n atau x - n tidak berada dalam jangkauan 0 ... 25 (misalnya kode ASCII) kita harus mengurangi atau menambahkan sesuai kode terkecilnya sehingga proses tetap berada di modulo 26. Hal ini dilakukan dengan sama sepanjang pesan, sehingga sandi Caesar digolongkan sebagai jenis substitusi monoalphabetik, lawan dari substitusi polialphabetik.

2.2. VIGENERE CIPHER

Untuk mengenkripsi dengan algoritma Vigenere klasik, kita dapat menggunakan tabel abjad yang disebut sebagai *tabula recta*, *Vigenere Square*, atau tabel Vigenere. Tabel ini terdiri dari alfabet yang ditulis 26 kali pada baris yang berbeda, setiap alfabet bergeser ke kiri siklis dibandingkan dengan abjad sebelumnya, sesuai dengan 26 kemungkinan sandi Caesar. Pada titik yang berbeda dalam proses enkripsi, penyandian menggunakan alfabet pada urutan yang berbeda dari salah satu baris. Alfabet yang digunakan pada setiap titik tergantung pada kata kunci yang berulang.



Gambar 2 - Tabel Vigenere

Sebagai contoh, anggaplah bahwa plaintext yang akan dienkripsi adalah:

UTS KRIPTOGRAFI DIGANTIKAN MAKALAH

Pengirim pesan memilih kata kunci dan mengulanginya sampai sesuai dengan panjang plaintext, misalnya, kata kunci yang dipilih adalah "CIRCULAR":

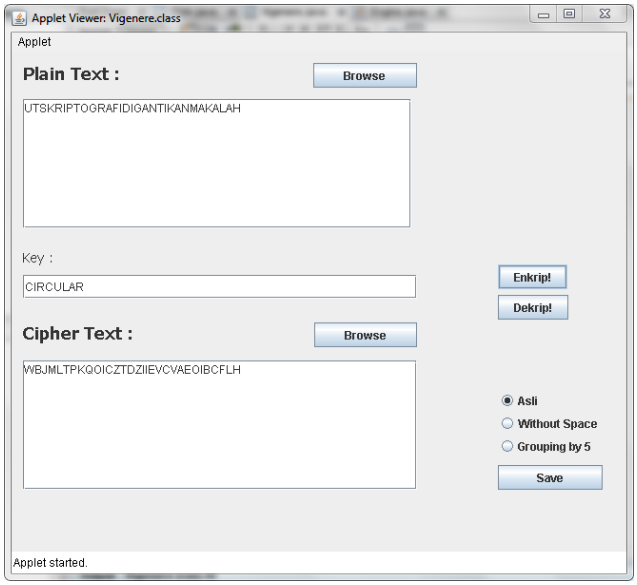
CIR CULARCIRCUL ARCIRCULAR CIRCULA

Huruf pertama dari plaintext, 'U', dienkripsi menggunakan alfabet pada baris 'C', yang merupakan huruf pertama dari kunci. Hal ini dilakukan dengan melihat karakter di baris 'C' dan kolom 'U' dari *Vigenere Square*, yaitu 'W'. Demikian pula, untuk karakter kedua plaintext, kita menggunakan karakter kedua pula dari kunci; huruf di baris 'I' dan kolom 'T' adalah 'B'. Dan sisa dari plaintext dienkripsi dengan cara yang sama:

Plaintext :
UTSKRIPTOGRAFIDIGANTIKANMAKALAH

Kunci :
CIRCULARCIRCULARCIRCULARCIRCULA

Ciphertext :
WBJMLTPKQOICZTDZIIEVCVAEOIBCFLH



Gambar 3 - Hasil Enkripsi Vigenere

Dekripsi dilakukan dengan pergi ke baris dalam tabel sesuai dengan kunci, menemukan posisi huruf ciphertext di baris ini, dan kemudian menggunakan kolom label sebagai plaintext. Sebagai contoh, pada baris 'C' (dari 'C'- IRCULAR), karakter sandi 'W' muncul pada kolom 'U', yang merupakan huruf plaintext pertama. Selanjutnya kita pergi ke baris 'I' (dari C-'I'-RCULAR), dan diperoleh karakter sandi 'B' di kolom 'T', yang merupakan karakter kedua plainteks itu.

Vigenere juga dapat dilihat secara aljabar. Jika huruf A - Z adalah dianggap sebagai angka 0-25, dan setelah penambahan dilakukan modulo 26, maka enkripsi Vigenere dapat ditulis :

$$C_i \equiv P_i + K_i (mod\ 26)$$

dan dekripsi sebagai :

$$P_i \equiv C_i - K_i (mod\ 26)$$

2.3. ANALISIS FREKUENSI BAHASA INGGRIS

Secara sederhana pengertian cipher substitusi adalah setiap huruf dari plaintext digantikan dengan yang lain, dan huruf tertentu di dalam plaintext akan selalu diubah menjadi huruf yang sama dalam ciphertext. Misalnya, jika dalam bahasa Ingrris semua huruf e berubah menjadi huruf X, sebuah pesan chipertext akan berisi banyak huruf X, dan hal ini akan memberi petunjuk pada kriptanalis bahwa X mewakili e.

Penggunaan dasar analisis frekuensi adalah pertama-tama menghitung frekuensi huruf ciphertext dan kemudian mengasosiasikan dengan menebak huruf plaintextnya. Menurut hasil analisis sebelumnya pada Bahasa Inggris, X sesuai dengan e dalam plaintext, tetapi ini tidak pasti; t dan juga sangat umum dalam bahasa Inggris, sehingga X akan menjadi salah satu dari mereka juga. Namun, kecil mungkinan X di sini menjadi plaintext z atau q yang kurang umum. Jadi, cryptanalyst mungkin perlu mencoba beberapa kombinasi pemetaan antara cipherteks dan huruf plainteks.

Untuk penggunaan yang lebih kompleks, kita dapat juga mempertimbangkan statistik pasangan huruf (digrams), triplet (trigram), dan seterusnya. Hal ini dilakukan untuk memberikan lebih banyak informasi kepada kriptanalis, misalnya, Q dan U hampir selalu terjadi bersama-sama dalam urutan dalam bahasa Inggris, meskipun Q sendiri sangat jarang.

Selain itu dapat dimungkin bahwa plainteks yang dikirm tidak menunjukkan distribusi frekuensi yang diharapkan sang analis. Hal ini biasanya terjadi ketika ukuran pesan pendek sebab pesan pendek cenderung menunjukkan lebih banyak variasi. Dapat juga disebabkan karena si pengirim sengaja membuat teka-teki teks secara sengaja untuk estetika. Sebagai contoh, seluruh novel telah ditulis bahwa menghilangkan huruf "e" sama sekali - suatu bentuk sastra yang dikenal sebagai lipogram.

3. ANALISIS FREKUENSI BAHASA INDONESIA DAN MODIFIKASI VIGENERE

3.1. ANALISIS BERBAGAI JENIS TEKS

Analisis frekuensi yang ada sudah ditemukan sekarang, hanya terbatas pada bahasa Inggris saja. Banyaknya distribusi informasi membuat kebutuhan enkripsi dan dekripsi di Indonesia juga menjadi hal yang vital dan wajib dipertimbangkan. Hal ini juga menuntut ditemukannya tabel frekuensi untuk teks Bahasa Indonesia. Berikut ini adalah hasil analisis untuk berbagai jenis teks Bahasa Indonesia dengan menggunakan program Java Applet yang telah penulis buat.

3.1.1. TEKS BACAAN FIKSI

Berikut ini adalah hasil analisis frekuensi teks bacaan dengan jenis cerpen, yang berjudul "Peradilan Rakyat"[5].

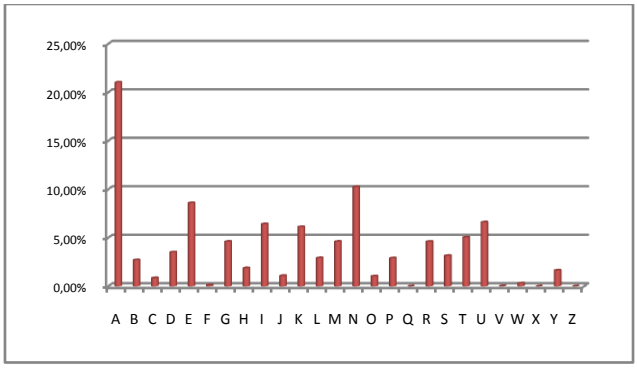
Gambar 4 - Daftar Frekuensi pada Cerpen

[5] http://kumpulan-cerpen.blogspot.com/

Dari sini kita dapat mengetahui bahwa huruf ‘a’ paling sering muncul di sini dengan frekuensi 2444 kali.

Huruf	Frekuensi	Persentasi
A	2444	21,03 %
B	314	2,70 %
C	99	0,85 %
D	408	3,51 %
E	998	8,59 %
F	13	0,11 %
G	536	4,61 %
H	218	1,88 %
I	746	6,42 %
J	126	1,08 %
K	712	6,13 %
L	339	2,92 %
M	536	4,61 %
N	1191	10,25 %
O	121	1,04 %
P	337	2,90 %
Q	0	0,00 %
R	535	4,60 %
S	366	3,15 %
T	586	5,04 %
U	768	6,61 %
V	2	0,02 %
W	36	0,31 %
X	0	0,00 %
Y	191	1,64 %
Z	0	0,00 %
Total	11622	100%

Tabel 2- Frekuensi Huruf pada “Peradilan Rakyat”



Bagan 2 - Analisis Frekuensi Cerpen “Peradilan Rakyat”

3.1.2. TEKS BACAAN NON-FIKSI

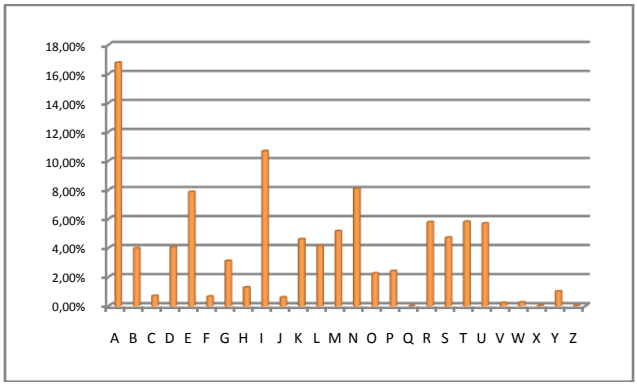
Berikutnya, analisis frekuensi juga dilakukan terhadap bacaan non-fiksi yang disadur dari wikipedia tentang “Bumi” dengan menggunakan program yang sama.

Gambar 5 - Analisis Frekuensi Teks Non-Fiksi

Dan dari sini kita memperoleh data bahwa huruf ‘a’ adalah yang paling sering muncul, sedangkan huruf ‘q’ sama sekali tidak pernah muncul.

Huruf	Frekuensi	Persentasi
A	1116	16,77 %
B	265	3,98 %
C	47	0,71 %
D	270	4,06 %
E	523	7,86 %
F	44	0,66 %
G	207	3,11 %
H	86	1,29 %
I	711	10,69 %
J	40	0,60 %
K	307	4,61 %
L	276	4,15 %
M	344	5,17 %
N	538	8,09 %
O	150	2,25 %
P	161	2,42 %
Q	0	0,00 %
R	385	5,79 %
S	314	4,72 %
T	387	5,82 %
U	379	5,70 %
V	15	0,23 %
W	17	0,26 %
X	1	0,02 %
Y	68	1,02 %
Z	2	0,03 %
Total	6653	100%

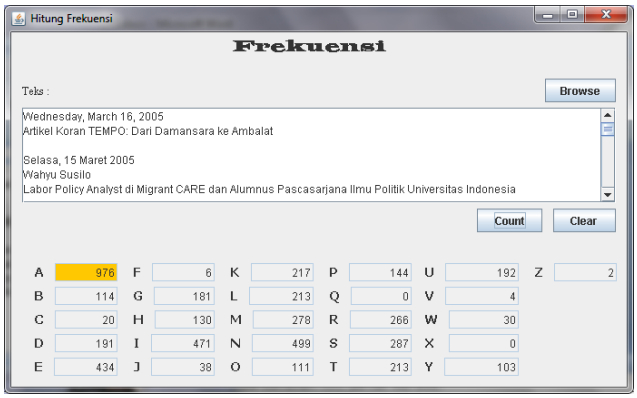
Tabel 3 - Frekuensi Huruf pada Teks “Bumi”



Bagan 3 – Analisi Frekuensi Teks Non-Fiksi "Bumi"

3.1.3. TEKS ARTIKEL KORAN

Berikut ini, penulis juga mencoba melampirkan hasil analisis frekuensi yang dilakukan terhadap sebuah artikel di koran Tempo yang berjudul "Dari Damansara Ke Ambalat" dengan menggunakan program yang sama.

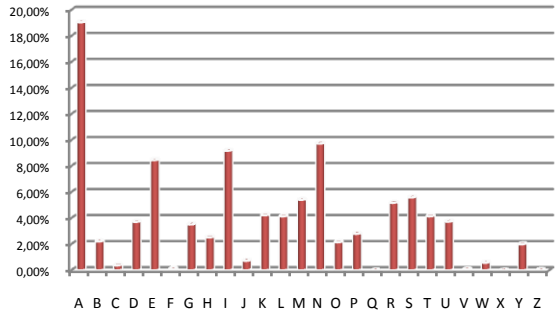


Gambar 6 - Analisis Frekuensi Artikel Koran

Dari sini dapat terlihat kembali, bahwa frekuensi huruf 'a' menduduki peringkat teratas dengan jumlah 976 kali, dan di sini huruf yang tidak pernah muncul adalah huruf 'q' dan 'x'.

Huruf	Frekuensi	Persentasi
A	976	19,06 %
B	114	2,23 %
C	20	0,39 %
D	191	3,73 %
E	434	8,48 %
F	6	0,12 %
G	181	3,54 %
H	130	2,54 %
I	471	9,20 %
J	38	0,74 %
K	217	4,24 %
L	213	4,16 %
M	278	5,43 %
N	499	9,75 %
O	111	2,17 %
P	144	2,81 %
Q	0	0,00 %
R	266	5,20 %
S	287	5,61 %
T	213	4,16 %
U	192	3,75 %
V	4	0,08 %
W	30	0,59 %
X	0	0,00 %
Y	103	2,01 %
Z	2	0,04 %
Total	**5120**	**100 %**

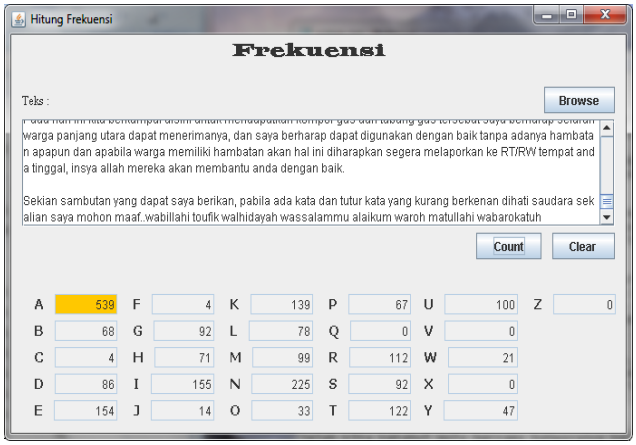
Tabel 4 - Frekuensi Huruf pada Artikel "Dari Damansara Ke Ambalat"



Bagan 4 - Analisis Frekuensi Arttikel "Dari Damansara Ke Ambalat"

3.1.4. TEKS PIDATO FORMAL

Selain pada teks lisan, analisis frekuensi dapat juga dilakukan terhadap sebuah teks pidato formal. Hal ini karena teks pidato juga merupakan bentuk rancangan komunikasi verbal yang menggunakan Bahasa Indonesia dengan baik dan benar.

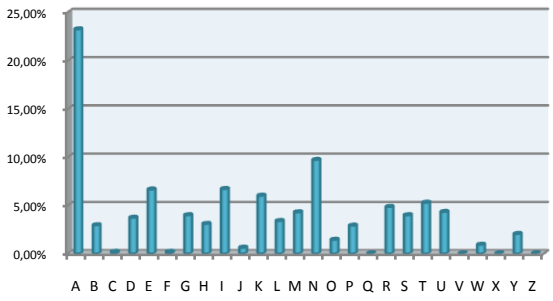


Gambar 7 - Analisis Frekuensi Teks Pidato

Dari sini kita dapat melihat hal yang sama, yaitu huruf ‘a’ merupakan huruf yang paling dominan muncul dan kemudian diikuti oleh huruf ‘N’ dan huruf ‘I’. Hal ini akan dipergunakan untuk menentukan urutan huruf yang paling sering muncul dalam teks berbahasa Indonesia.

Huruf	Frekuensi	Persentasi
A	539	23,21 %
B	68	2,93 %
C	4	0,17 %
D	86	3,70 %
E	154	6,63 %
F	4	0,17 %
G	92	3,96 %
H	71	3,06 %
I	155	6,68 %
J	14	0,60 %
K	139	5,99 %
L	78	3,36 %
M	99	4,26 %
N	225	9,69 %
O	33	1,42 %
P	67	2,89 %
Q	0	0,00 %
R	112	4,82 %
S	92	3,96 %
T	122	5,25 %
U	100	4,31 %
V	0	0,00 %
W	21	0,90 %
X	0	0,00 %
Y	47	2,02 %
Z	0	0,00 %
Total	**2322**	**100 %**

Tabel 5 - Frekuensi Huruf pada Teks Pidato Formal

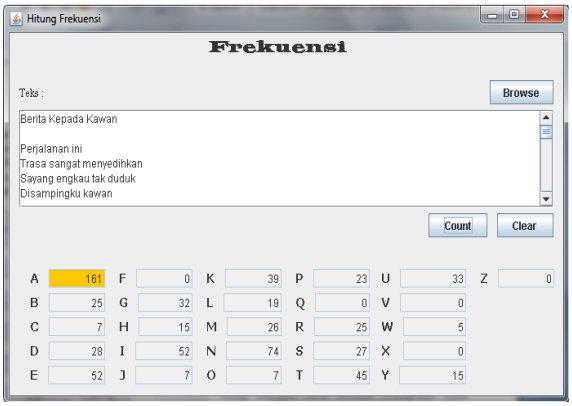


Bagan 5 - Analisis Frekuensi Teks Pidato Formal

Dari keempat sample yang diteliti di atas, sebenarnya dapat diambil kesimpulan awal secara induksi, bahwa huruf yang memiliki tingkat kemunculan paling tinggi pada teks Bahasa Indonesia adalah huruf ‘a’.

3.1.5. LIRIK LAGU BAHASA INDONESIA

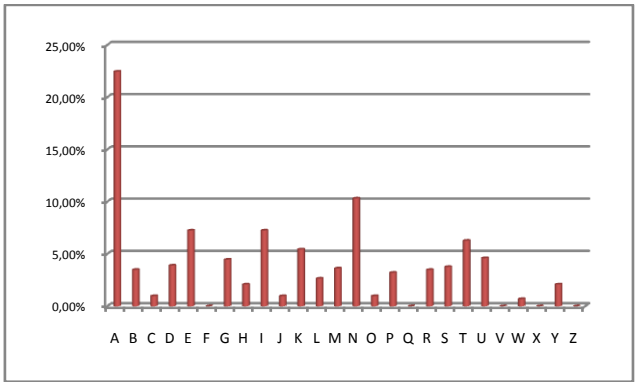
Lagu dalam Bahasa Indonesia juga merupakan sebuah sampel yang layak untuk diuji tingkat frekuensinya karena lagu merupakan salah satu sampel dari jenis karya seni yang biasanya sering didengarkan orang. Berikut hasil analisis pada Lagu Berita Kepada Kawan, karangan Ebiet G. Ade.



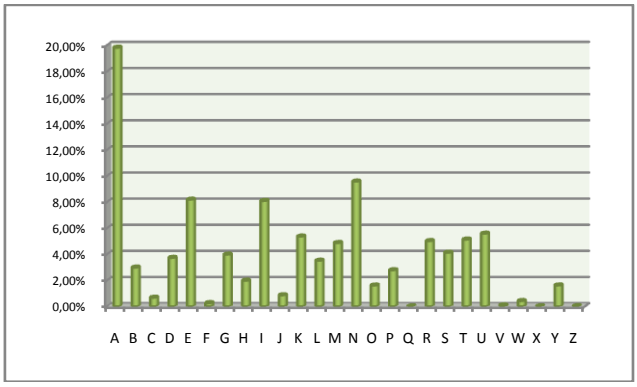
Gambar 8 - Analisis Frekuensi Lirik Lagu

Huruf	Frekuensi	Persentasi
A	161	22,45%
B	25	3,49%
C	7	0,98%
D	28	3,91%
E	52	7,25%
F	0	0,00%
G	32	4,46%
H	15	2,09%
I	52	7,25%
J	7	0,98%
K	39	5,44%
L	19	2,65%
M	26	3,63%
N	74	10,32%
O	7	0,98%
P	23	3,21%
Q	0	0,00%
R	25	3,49%
S	27	3,77%
T	45	6,28%
U	33	4,60%
V	0	0,00%
W	5	0,70%
X	0	0,00%
Y	15	2,09%
Z	0	0,00%
Total	**717**	**100 %**

Tabel 6 - Frekuensi Huruf pada Lagu



Bagan 6 - Analisis Frekuensi Lagu



Bagan 7 - Analisis Frekuensi Keseluruhan

3.1.6. ANALISIS FREKUENSI TOTAL

Dari ke lima data di atas, dapat disusun sebuah analisis frekuensi keseluruhan pada beberapa jenis teks di Bahasa Indonesia dengan bentuk sebagai berikut.

Huruf	Frekuensi	Persentasi
A	5236	19,81%
B	786	2,97%
C	177	0,67%
D	983	3,72%
E	2161	8,18%
F	67	0,25%
G	1048	3,96%
H	520	1,97%
I	2135	8,08%
J	225	0,85%
K	1414	5,35%
L	925	3,50%
M	1283	4,85%
N	2527	9,56%
O	422	1,60%
P	732	2,77%
Q	0	0,00%
R	1323	5,00%
S	1086	4,11%
T	1353	5,12%
U	1472	5,57%
V	21	0,08%
W	109	0,41%
X	1	0,00%
Y	424	1,60%
Z	4	0,02%
Total	**26434**	**100 %**

Tabel 7 - Tabel Frekuensi Total

Dari sini dapat kita bentuk suatu 10 urutan teratas frekuensi kemunculan huruf sebagai berikut.

Huruf	Persentasi
A	19,81%
N	9,56%
E	8,18%
I	8,08%
U	5,57%
K	5,35%
T	5,12%
M	4,85%
S	4,11%
G	3,96%

Tabel 8 - Urutan Frekuensi Tertinggi

3.2. MODIFIKASI ALGORITMA VIGENERE

Dari pengetahuan di atas, serangan terhadap suatu algoritmakriptografi klasik akan mudah dilakukan, terutama dengan menganalisis frekuensi kemunculan huruf dan menganalogikannya sebagai ‘a’ pada substitusi monoalphabethik atau penerapan yang sama untuk Metode Kasiski. Hal ini mendorong penulis untuk memodifikasi algoritmaVigenere menjadi algoritma yang sulit untuk dipecahkan yaitu dengan menggunakan bilangan **euler** dan **phi** (GOLDEN RATIO).

Pembangkitan kunci ini menggunakan perkalian secara matematis dengan bilangan **Euler+Phi** kemudian setiap angka hasil perkalian akan dikelompokkan sesuai dengan panjang digit dari kunci.

Sebagai contoh:

Plaintext: KILL KING TONIGHT OR DIE (20 karakter)

Kunci (dalam huruf) : BUDI

Kunci (dalam angka) : 1 20 3 8 (panjang angka = 5 digit)

Kemudian ambil bilangan euler dan phi dengan panjang sesuai dengan plaintext (yaitu 20 digit):

Euler : 2.7182818284590452353[6]

Phi : 1.6180339887498948482

Euler + Phi =
2718281828459045235 3+1618033988749894848 2
= 4336315817208940083 5

Berikutnya, kunci dalam bentuk angka dikalikan bilangan euler + phi adalah:

12038 x 4336315817208940083 5 =
521389941229568537759873 0

Dari sini kita pisahkan lagi sesuai dengan panjang angka kunci (5 digit) menjadi:

5 21 3 8 9 94 1 2 2 95 6 8 5 37 7 5 9 87 3 0

Kemudian kita ubah menjadi bentuk huruf lagi (mod 26) sehingga menjadi

F V D I J Q B C C R G I F L H F J J D A

Sehingga kunci yang akan digunakan dalam melakukan encrypt adalah: FVDIJQBCCRGIFLHFJJDA

Maka didapatkan hasil sebagai berikut:

Plaintext : KILL KING TONIGHT OR DIE

Kunci : FVDI JQBC CRGIFLH FJ JDA

Ciphertext : PDOT TYOI VFTQLSA TA MLE

Kekuatan dari metode ini adalah pengelompokan hasil perkalian antara kunci dengan bilangan euler dan phi yang sangat sulit untuk diterka serta bentuk acak dari kunci yang sangat panjang, sehingga akan menyulitkan kriptanalisis untuk menyerang dengan menggunakan metode Kasiski maupun analisis frekuensi.

IV. KESIMPULAN

Dalam teks Bahasa Indonesia, urutan karakter yang paling sering muncul adalah **ANEIUKTMSG.**

Karena metode kriptografi klasik hanya terdiri dari 2 jenis transposisi dan substitusi, *attacker* akan dapat menyerang cipherteks dengan mudah apabila panjang kunci diketahui, padanan huruf untuk analisis frekuensi diketahui, dan jenis bahasa diketahui.

Oleh sebabitu, untuk mencegah serangan dengan *known-ciphertext-attack* pada algoritma kriptografi klasik, kita perlu melakukan pengamanan dengan cara membuat kunci sepanjang plainteks, tidak menggunakan saluran komunikasi umum untuk mengirimkan kunci, tidak menggunakan kunci yang sama berulang kali, dan mengenkripsi tanpa pola dan format pada plainteks.

Untuk melakukan enkripsi dengan panjang kunci sepanjang plainteks, kita dapat menggunakan bilangan

Euler : 2.7182818284590452353

dan

Phi : 1.6180339887498948482

[6] http://www.mu.org/~doug/exp/100000.html

REFERENSI

[1] R. Munir, Diktat Kuliah IF5054 Kriptografi, Program Studi Teknik Informatika Institut Teknologi Bandung, 2006.

[2] Livio, Mario. "The golden ratio and aesthetics"

[3] http://sistem-keamanan-komputer.blogspot.com/ 2009/04/kriptografi-klasik.html 17:00

[4] http://www.aaipul.co.cc/2009/05/contoh-sambutan.html

[5] http://www.lyricsty.com/lyrics/e/ebiet_g_ade /berita_kepada_kawan.html