Penggunaan Teori Kombinatorial dalam CAPTCHA

Gilbran Imami, 13509072
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13509072@std.stei.itb.ac.id

***Abstrak*—Kemajuan zaman yang pesat dan berkembangnya teknologi internet mendorong sebagian orang untuk melakukan kejahatan dengan menggunakan teknologi komputer, kejahatan *cyber* tersebut telah merugikan banyak sekali pihak dengan besar kerugian yang tidak sedikit. Namun hal ini mendorong teknologi dari sisi keamanan turut berkembang pesat. Salah satu contoh teknologi di bidang keamanan adalah penggunaan CAPTCHA. CAPTCHA sendiri merupakan teknologi yang terlihat simpel namun sangat ampuh untuk menangkal *hacker-hacker* yang ingin menyusupi berbagai situs-situs internet yang memiliki celah di bidang keamanan. Prinsipnya cukup sederhana, yaitu dengan mengkombinasikan huruf dan angka menjadi satu atau dua buah kata yang tidak memiliki makna. Kemudian pengguna akan mengetikkan kata-kata tersebut dan server akan mengecek kebenarannya. Dan kata yang ditunjukkan oleh CAPTCHA biasanya memiliki tampilan yang sedikit sulit dibaca karena dilakukan proses distorsi dan melekuk teks sedemikian rupa. Tapi tetap saja para hacker terus mencari cara untuk dapat membuat program yang dapat membaca CAPTCHA tersebut. Untuk itu, kombinasi tidak hanya dilakukan untuk membuat kata, tapi juga untuk mencari variasi apa saja yang bisa dilakukan untuk membuat teks sulit dibaca oleh program komputer buatan para *hacker*.**

***Kata Kunci*—Captcha, kombinatorial, keamanan.**

I. PENDAHULUAN

Di zaman modern ini penggunaan internet sudah menjadi bagian hidup dari setiap orang yang hampir tidak bisa dipisahkan lagi. Pengguna internet di seluruh dunia terdiri dari berjuta-juta orang meliputi berbagai bidang seperti lembaga, pemerintah, akademisi, bisnis, maupun perseorangan. Penggunaan internet tersebut sangat beragam, seperti komunikasi data dan media musik, gambar, video; berkirim email, menulis blog, maupun sekedar *browsing* dan *download.*

Penggunaan internet yang global dan multifungsi tersebut mengundang banyak sekali terjadinya kejahatan *cyber*. Kejahatan *cyber* yang dilakukan pihak-pihak yang tidak bertanggung jawab menyebabkan banyak sekali kerugian, baik kerugian yang tidak menimbulkan banyak dampak, maupun kerugian yang dapat menyebabkan seseorang kehilangan uang milyaran rupiah.

Beberapa contoh kejahatan *cyber* yang berbahaya dan sedang tren di beberapa negara termasuk Indonesia, yaitu:

1. **Kasus *spoofing/phishing* berkombinasi dengan *Malware***
   Situs web yang ditempel dengan situs palsu ini berisi formulir bank palsu dan juga terdapat *malware* yang akan menyerang di sisi *end user* yang membuka *URL Phishing* tersebut.
   Sistem yang diserang di antaranya yang berbasis Windows hingga Open Source seperti RedHat, Linux, dan lainnya.
2. **Kasus *Scam* (Penipuan) mengatasnamakan institusi pemerintah**
   ID-CERT (*Indonesia Computer Emergency Response Team*) menerima laporan dari sebuah kelompok *anti fraud* di Eropa yang menyampaikan keluhan tentang adanya dugaan email scam yang beredar di Eropa mengatasnamakan institusi pemerintah Indonesia dan meminta bantuan ID-CERT melakukan investigasi lebih jauh tentang hal ini.
3. ***IP Address* pemerintah yang digunakan untuk melakukan *Network Incident* ke luar/dalam negeri, seperti melakukan *DDOS Attack, Probing* bahkan hingga *Flooding***
   Laporan terbanyak untuk sektor pemerintah ini justru datang dari *IP Address* dan situs web yang digunakan oleh kalangan pendidikan di bawah kemdiknas. Untuk situs web, banyak laporan tentang adanya situs web .sch.id dan ac.id yang mengalami serangan *cyber*.
4. **Kasus *spoofing/phishing* ke bank di Indonesia dan Malaysia**
   Kasus terbanyak yang dilaporkan ke ID-CERT dalam masalah *spoofing/phishing* ini adalah situs web perbankan di Indonesia yang dipalsukan serta dibuat mirip dengan aslinya.
   Umumnya situs yang dipalsukan adalah dengan nama domain generik (.com, dan .net). Sedangkan untuk bank dengan nama domain .co.id, hampir belum pernah ada laporan yang masuk.
   Selain bank di Indonesia, hal yang sama juga menimpa situs perbankan di Malaysia dan Eropa yang justru dipalsukan dan ditempeli disitus web maupun IP Address organisasi di Indonesia.

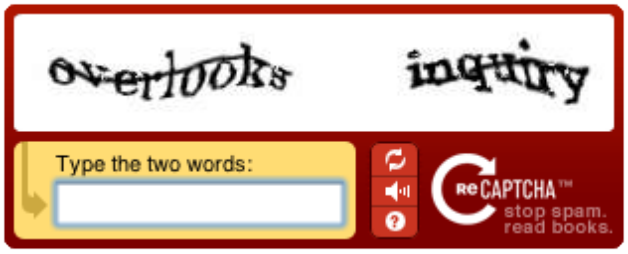
Contoh kejahatan cyber berbahaya yang perlu disoroti adalah pencurian identitas seseorang dan penjebolan server suatu lembaga. Yang dimaksud pencurian identitas adalah seseorang dapat memiliki akses ke halaman

pribadi seseorang yang di dalamnya berisi informasi-informasi penting yang pribadi. Informasi-informasi seseorang yang sering dicuri yaitu password, PIN ATM, kartu kredit, dan informasi penting lainnya.

Sedangkan yang dimaksud dengan penjebolan server suatu lembaga adalah saat pihak tertentu mendapatkan akses ke suatu server lembaga, dan mengganggu atau bahkan merusak fungsionalitas server tersebut. Contohnya adalah server lembaga yang menyediakan jasa email. Seorang *hacker* dapat masuk ke server lembaga tersebut, menyisipkan program jahat seperti *bot* atau *malware*, dan kemudian membebani server tersebut.

Contoh kasus pembobolan tersebut pernah terjadi pada server situs raksasa Google. Para pembajak diduga berasal dari China dan berhasil menggunakan Google untuk mendapatkan akses ke rekening pejabat senior AS, Korea, dan pemerintahan lainnya.

Untuk meminimalisir kejadian tersebut, digunakanlah CAPTCHA. CAPTCHA adalah program yang dapat men-*generate* dan menilai suatu tes yang dapat dijawab oleh manusia, dan tidak oleh program komputer. Biasanya berupa teks terdistorsi yang merupakan kombinasi acak.



Gambar 1. Salah satu contoh CAPTCHA

II. Dasar Teori

Dalam bab ini akan dibahas mengenai apa itu CAPTCHA, sejarah dan penggunaan CAPTCHA hingga sekarang, dan teori kombinatorial yang menjadi salah satu dasar penggunaan CAPTCHA.

A. Pengertian CAPTCHA

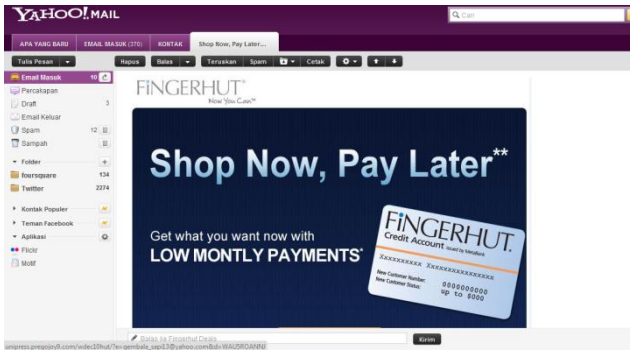
CAPTCHA merupakan singkatan dari *Completely Automated Public Turing Test To Tell Computers and Humans Apart* atau tes otomatis untuk membedakan manusia dan komputer. CAPTCHA adalah sejenis tes Turing untuk melihat respon yang dilakukan komputer sebagai langkah untuk memastikan apakah respon tersebut dilakukan oleh manusia atau bukan.

Proses yang dilakukan biasanya melibatkan satu komputer (suatu server) yang menanyakan pengguna untuk mengisi tes yang simpel. Tes tersebut merupakan kombinasi yang di-*generate* oleh komputer dan kemudian dilihat hasilnya apakah benar atau tidak. Karena diasumsikan bahwa komputer lain tidak akan bisa menjawab tes CAPTCHA, maka siapapun yang berhasil menjawab adalah manusia. Untuk itu, CAPTCHA sering dianggap sebagai tes Turing terbalik. Karena justru mesinlah yang memberikan pertanyaan dengan targetnya adalah manusia.

Yang dimaksud dengan tes Turing adalah tes kemampuan mesin untuk menunjukkan tingkah laku yang cerdas. Tes ini diperkenalkan oleh Alan Turing pada tahun 1950 dalam papernya yang berjudul *Computing Machinery and Intellegence*.

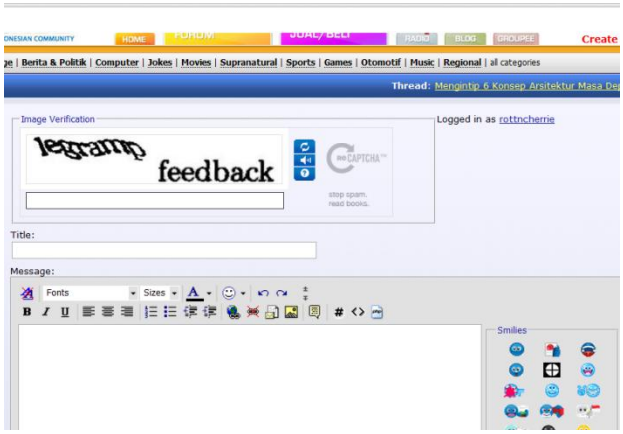
Istilah "CAPTCHA" sendiri diperkenalkan pada tahun 2000 oleh Louis von Ahn, Manuel Blum, Nicholas J. Jopper, dan John Langfold yang semuanya berasal dari Carnegie Mellon University. CAPTCHA adalah akronim yang kata-katanya berbasis pada kata "capture" tetapi memiliki singkatan khusus yaitu "***C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part*". Carnegie Mellon University telah mencoba untuk membuat *trademark* istilah tersebut, namun aplikasi trademark tersebut di abaikan pada 21 April 2008.

Pada masa ini CAPTCHA sering digunakan untuk melindungi sistem yang dapat ditembus oleh *spam*, seperti layanan webmail dari Gmail, Yahoo! Mail, dan Hotmail. Hingga beberapa tahun yang lalu, hampir semua dari layanan tersebut terjangkit tipe serangan yang spesifik dari program jahat (*bot*) yang melakukan proses pendaftaran ribuan akun email setiap menitnya. Hal tersebut melumpuhkan server-server penyedia layanan tersebut selama beberapa waktu. Hingga solusinya ditemukan, yaitu penggunaan CAPTCHA yang membatasi pihak-pihak untuk dapat mendaftar.



Gambar 2. Email yang terkena spam.

CAPTCHA juga digunakan untuk meminimalisir postingan otomatis ke blog, forum, dan wiki, yang biasanya menghasilkan promosi komersil, ataupun penyalahgunaan lain.



Gambar 3. CAPTCHA yang harus diisi sebelum memposting di forum kaskus

Kebanyakan bloggers mungkin cukup kenal dengan program yang mengajukan komentar palsu, dengan tujuan yang biasanya meningkatkan peringkat situs webnya pada *search engine*. Dengan menggunakan CAPTCHA, hanya manusia yang dapat memasukkan komentar pada blog. Dengan demikian, tidak perlu membuat pengguna mendaftar dahulu sebelum bisa memasukkan komentar, seperti yang terjadi pada beberapa waktu lalu, sebelum CAPTCHA menjadi populer seperti sekarang.

CAPTCHA juga menawarkan solusi terhadap email *worms* dan *spam* dengan konsep: "saya hanya menerima email jika saya tahu bahwa pengirimnya adalah manusia." Beberapa perusahaan telah menggunakan ide ini.

Jika suat situs web membutuhkan perlindungan dari penyalahgunaan, maka sangat direkomendasikan menggunakan CAPTCHA ini. Ada banyak implementasi dari CAPTCHA, dan beberapa diantaranya memiliki keunggulan-keunggulan dari yang lainnya. Berikut ini adalah rekomendasi untuk setiap CAPTCHA:

- Aksesibilitas
  CAPTCHA harus dapat diakses. CAPTCHA berbasis semata-mata pada membaca teks, atau persepsi visual lain, dapat membuat seseorang dengan kelainan visual untuk mendapatkan akses. Untuk itu, setiap implementasi CAPTCHA harus dapat memperbolehkan orang yang kelainan tadi dapat juga mengisi CAPTCHA, contohnya, dengan memberikan pilihan untuk audio CAPTCHA.
- Keamanan Gambar
  Gambar atau teks harus didistorsi secara acak sebelum dapat ditunjukkan ke pengguna. Banyak implementasi dari CAPTCHA menggunakan teks yang tidak terdistorsi. Implementasi tersebut dapat ditembus oleh program. Sebagai contoh, CAPTCHA dibawah ini semua dapat ditembus menggunakan teknik pemrosesan gambar.

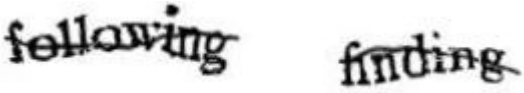
Gambar 4. Contoh CAPCHA yang dapat ditembus.

- Keamanan Skrip
  Membuat CAPTCHA tidaklah mudah. Sebagai tambahan untuk membuat gambar tidak bisa dibaca komputer, sistem harus memastikan tidak ada cara mudah untuk menembusnya pada level skrip. Contoh umum dari ketidakamanan aspek ini antara lain: (1) Sistem yang memberi jawaban kepada CAPTCHA dalam teks sederhana sebagai bagian dari form. (2) Sistem dimana solusi dari CAPTCHA dapat digunakan berulang kali.

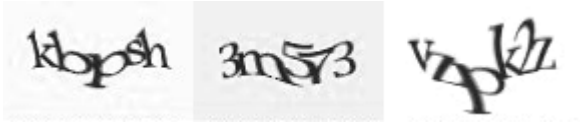
Berbagai jenis CAPTCHA dari waktu ke waktu:

1. CAPTCHA pada awal mula, di *generate* oleh program EZ-Gimpy, dan pernah digunakan oleh Yahoo!. Tetapi telah ada teknologi yang bisa membaca dan menembus CAPTCHA tipe ini.

2. CAPTCHA modern, dibandingkan dengan hanya membuat latar terdistorsi dan teks berlekuk dengan level tinggi, lebih fokus kepada penambahan pembuatan segmentasi yang sulit dengan menambahkan garis bersudut.



3. Cara lain untuk membuat segmentasi yang sulit adalah dengan menggabungkan simbol-simbol bersama, seperti yang digunakan Yahoo! Pada saat ini.



B. Teori Kombinatorial

Kombinatorial adalah cabang matematika yang mempelajaari pengaturan susunan objek-objek. Yang dimaksud dengan pengaturan disini adalah bagaimana

objek-objek dapat dikombinasikan dalam berbagai susunan atau urutan yang menghasilkan output yang berbeda.

Secara umum, ada dua kaidah utama dalam teori kombinatorial, yaitu:

1. Kaidah Perkalian
   Jika terdapat kejadian P dan Q dimana P dan Q dilakukan bersamaan, maka banyaknya kejadian yang mungkin sama dengan P x Q.
2. Kaidah Penjumlahan
   Jika terdapat kejadian P dan Q dimana P dan Q dilakukan tidak bersamaan (dalam kondisi yang berbeda), maka banyaknya kejadian yang mungkin sama dengan P + Q.

Selain dua kaidah diatas, ada dua cara untuk menghitung dalam teori kombinatorial, yaitu permutasi dan kombinasi.

1. Permutasi
   Permutasi adalah salah satu bentuk umum dari kombinatorial. Permutasi dari n elemen adalah jumlah kemungkinan urutan r buah elemen yang dipilih dari n buah elemen, dengan r ∈ n, yang dalam hal ini, pada setiap kemungkinan urutan tidak ada elemen yang sama.
   Jika terdapat suatu untai abjad *abcd*, maka untai itu dapat dituliskan kembali dengan urutan yang berbeda: *acbd*, *dacb*, dan seterusnya. Selengkapnya ada 24 cara menuliskan keempat huruf tersebut dalam urutan yang berbeda satu sama lain.

abcd	abdc	acbd	acdb	adbc	adcb
bacd	badc	bcad	bcda	bdac	bdca
cabd	cadb	cbad	cbda	cdab	cdba
dabc	dacb	dbac	dbca	dcab	dcba

   Terkadang kita hanya ingin menyusun ulang sejumlah elemen saja, tidak semuanya. Permutasi ini disebut permutasi-*k* dari *n* benda. Pada contoh untai *abcd*, maka permutasi-2 dari *abcd* (yang semuanya ada 4 unsur) adalah sebanyak 12:
   ab ac ad
   ba bc bd
   ca cb cd
   da db dc
   Sedangkan permutasi-3 dari untai yang sama adalah sebanyak 24:
   abc abd acb acd adb adc
   bac bca bad bda bcd bdc
   cab cba cad cda cbd cdb
   dab dba dac dca dbc dcb
   Banyaknya kemungkinan permutasi seperti ini adalah

$$P_k^n = \frac{n!}{(n-k)!}$$

2. Kombinasi
   Kombinasi adalah bentuk khusus dari permutasi. Jika dalam permutasi urutan kemunculan diperhitungkan, maka dalam kombinasi, urutan kemunculan diabaikan.
   Kombinasi *r* dari sebuah himpunan *S*, berarti dari himpunan *S* diambil elemen sebanyak *r* untuk dijadikan sebuah himpunan baru.
   Banyaknya kombinasi *r* dari sebuah himpunan berisi *n* elemen dapat dihitung tanpa harus memperhatikan isi dari himpunan tersebut. Besarnya dinyatakan dengan fungsi:

$$C_r^n = \frac{n!}{r!(n-r)!}$$

   Fungsi C_r^n dalam banyak literatur dinyatakan juga dengan notasi $\binom{n}{r}$.

III. PEMBAHASAN

Secara spesifik disini akan dibahas mengenai penggunaan kombinatorial dalam menentukan CAPTCHA berbentuk teks. Standar yang digunakan pada reCAPTCHA, yaitu salah satu jasa penyedia CAPTCHA yang populer, adalah dengan menggunakan dua kata CAPTCHA yang masing-masing memiliki panjang maksimal sepuluh kata.

Pertama kita hitung apabila satu kata terdiri dari satu karakter, maka kemungkinan yang ada adalah:

C = 26 (huruf) + 10 (angka)
C = 36 kemungkinan

Jika satu kata terdiri dari dua karakter, maka kemungkinannya:

C = 36×36
C = 1.296 kemungkinan

Jika satu kata terdiri dari tiga karakter, maka kemungkinannya:

C = 36×36×36
C = 46.656 kemungkinan

Jika satu kata terdiri dari empat karakter, maka kemungkinannya:

C = 36×36×36×36
C = 1.679.616 kemungkinan

Jika satu kata terdiri dari lima karakter, maka kemungkinannya:

C = 36×36×36×36×36
C = 60.466.176 kemungkinan

Jika satu kata terdiri dari enam karakter, maka kemungkinannya:

C = 36×36×36×36×36×36
C = 2.781.444.096 kemungkinan

Jika satu kata terdiri dari tujuh karakter, maka kemungkinannya:

C = 36×36×36×36×36×36×36
C = 100.131.987.456 kemungkinan

Jika satu kata terdiri dari delapan karakter, maka kemungkinannya:

C = 36×36×36×36×36×36×36×36
C = 3.604.751.548.416 kemungkinan

Jika satu kata terdiri dari sembilan karakter, maka

kemungkinannya:

C = 36×36×36×36×36×36×36×36×36×36

C = 129.771.055.742.976 kemungkinan

Jika satu kata terdiri dari sepuluh karakter, maka kemungkinannya:

C = 36×36×36×36×36×36×36×36×36×36

C = 4.671.758.006.747.136 kemungkinan

Karena dalam satu kata kemungkinan jumlah hurufnya tidak tentu (1 hingga 10), maka total semua kemungkinan dalam satu kata adalah =

$$\begin{aligned} C_{kata} &= 36 + 1.296 + 46.656 + 1.679.616 \\ &\quad + 60.466.176 + 2.781.444.096 \\ &\quad + 100.131.987.456 \\ &\quad + 3.604.751.548.416 \\ &\quad + 129.771.055.742.976 \\ &\quad + 4.671.758.006.747.136 \\ &\quad = 4805236787984244 \end{aligned}$$

$$C_{kata} = 4805236787984244$$

Karena ada dua kata, maka kemungkinan tersebut dikalikan dengan dua:

$$\begin{aligned} C_{total} &= 4805236787984244 \times 2 \\ &= 9610473575968488 \end{aligned}$$

Sehingga kemungkinan CAPTCHA yang ada adalah sebesar 9.610.473.575.968.488 kemungkinan.

IV. KESIMPULAN

CAPTCHA adalah salah satu cara untuk dapat mengamankan suatu situs internet dengan mencegah situs tersebut disusupi oleh program yang tidak diinginkan. CAPTCHA bertujuan untuk membedakan apakah prngguna tersebut merupakan manusia atau program komputer. Cara CAPTCHA bekerja yaitu server akan memilih kombinasi huruf dan angka secara acak, menyusunnya, dan kemudian memilih kombinasi variasi distorsi dan modifikasi seperti apa yang dapat digunakan agar teks tersebut tidak mudah dibaca program. Teori kombinasi memungkinkan kita untuk dapat menghitung segala kombinasi yang dapat disusun oleh CAPTCHA jenis teks. Namun dengan berkembangnya teknologi, variasi dan jenis CAPTCHA semakin banyak setiap harinya, membuat kemungkinan banyaknya CAPTCHA menjadi tidak terbatas.

REFERENSI

http://en.wikipedia.org/wiki/CAPTCHA
Tanggal Akses: 10 Desember 2012, 1 4:00 WIB

http://www.captcha.tv/
Tanggal Akses: 10 Desember 2012, 1 6:00 WIB

http://www.detikinet.com/read/2011/09/23/160857/1729122/323/5-aksi-cyber-crime-yang-paling-disorot-id-cert
Tanggal Akses: 10 Desember 2012, 1 9:00 WIB

Munir, Rinaldi. "Diktat Kuliah IF2091 Struktur Diskrit", STEI, ITB, 2008.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 12 Desember 2012

Gilbran Imami, 13509072