Peningkatan Keamanan Siber dengan Mendeteksi Tautan Berbahaya atau Palsu Menggunakan Pencocokan *String*

Benjamin Sihombing - 13522054
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail (gmail): 13522054@std.stei.itb.ac.id

***Abstract*—Peningkatan digitalisasi telah memicu pertumbuhan signifikan dalam teknologi internet dan web, yang mendukung berbagai aspek kehidupan seperti penyebaran informasi, perdagangan, komunikasi, dan pekerjaan. Namun, kemajuan ini juga menghadirkan peluang bagi kejahatan siber, sehingga keamanan siber menjadi isu yang sangat penting. Salah satu ancaman siber yang umum adalah tautan palsu atau berbahaya yang menyerupai tautan asli, yang dapat menyebabkan pencurian data, penyebaran virus, serangan ransomware, dan phishing. Makalah ini membahas metode untuk meningkatkan keamanan siber dengan mendeteksi tautan berbahaya atau palsu menggunakan teknik pencocokan string. Pencocokan string memungkinkan verifikasi tautan secara akurat dan andal melalui program komputer, yang dapat mengidentifikasi serangan homograf yang sulit dikenali oleh mata manusia. Studi ini mengimplementasikan algoritma pencocokan string Boyer-Moore untuk memverifikasi tautan, dengan membandingkan URL yang berpotensi berbahaya dengan database URL asli yang dikenal. Pendekatan ini memastikan deteksi berbagai skenario serangan, termasuk tautan http biasa, penambahan karakter, variasi domain, dan serangan homograf IDN, sehingga secara signifikan meningkatkan keamanan tautan dan keamanan siber.**

Keywords—keamanan siber, pencocokan string, tautan berbahaya, serangan homograf, algoritma Boyer-Moore, verifikasi URL.

I. Pendahuluan

Dewasa ini, digitalisasi terjadi di berbagai hal. Salah satunya adalah teknologi internet dan *website* yang sangat berkembang. Banyak aspek kehidupan ditopang oleh teknologi internet dan *website* seperti media informasi, tempat berjualan, tempat berkomunikasi, dan bahkan tempat bekerja. Namun, segala sesuatu pasti memiliki celah sama seperti internet dan *website*. Celah tersebut bisa saja dimanfaatkan oleh oknum-oknum yang tidak bertanggung jawab untuk melakukan kejahatan. Maka dari itu, proteksi perlu dilakukan untuk menghindari dampak negatif dari kejahatan tersebut.

Keamanan siber menjadi salah isu penting di tengah digitalisasi tersebut. Internet dan *website* bisa menjadi sumber kejahatan-kejahatan baru seperti pencurian data pribadi hingga penyebaran virus ataupun hal-hal berbahaya lainnya. Salah satu strategi kejahatan yang ada saat ini adalah pemberian tautan palsu atau berbahaya yang mirip dengan tautan yang asli dan aman. Berikut ini contoh dari tautan palsu dan tautan asli, **googlé.com** (palsu) dan **google.com** (asli). Tautan tersebut memanfaatkan homograf dari tautan alsi. Dengan membuka link palsu, pengguna bisa terkena berbagai hal seperti pencurian data, serangan virus/*worms*, serangan *ransomware*, atau *phishing*.

Salah satu cara menghindari tautan palsu atau berbahaya adalah mengecek atau mendeteksi keaslian dan keamanan *website*. Pengecekan tautan bisa menjadi salah satu solusi untuk mengecek keaslian dan keamanan *website*. Pengecekan tautan bisa dilakukan dengan metode pencocokan *string*. Dengan pencocokan *string*, pengecekan tautan akan lebih akurat dan pasti karena dilakukan oleh program komputer. Komputer bisa mendeteksi tautan-tautan palsu *homograf* yang sulit dideteksi mata manusia secara langsung.

II. Teori Dasar

A. Keamanan Siber

Keamanan siber atau yang sering disebut dengan *cyber security* adalah praktik melindungi sistem, jaringan, dan program dari serangan digital. Serangan siber biasanya berupa pengaksesan, pengubahan, penghambatan, atau perusakan informasi sensitif di dalam suatu perangkat. Dengan hal ini, perangkat kurang bisa diandalkan dan bahkan tidak bisa digunakan. Ujung dari penyerangan ini biasanya adalah untuk mendapatkan keuntungan sepihak dengan memeras uang dari pihak pengguna. Keamanan siber melibatkan berbagai teknologi, proses, dan praktik yang dirancang untuk melindungi jaringan, perangkat, program, dan data dari serangan, kerusakan, atau akses tidak sah.

B. Bentuk Serangan Siber

Semakin berkembangnya digitalisasi, semakin banyak teknologi yang muncul. Semakin banyak pula celah untuk

oknum-oknum untuk memanfaatkannya. Berikut ini beberapa contoh serangan siber yang sering terjadi:

- *Malware*

  Malware adalah suatu *software* yang berbahaya untuk suatu perangkat. Malware terbagi menjadi beberapa jenis. Berikut ini beberapa jenis *Malware*:

  1. Virus
  2. Worms
  3. Trojan
  4. Ransomware

- *Phishing*

  Istilah *phishing* diambil dari kata bahasa inggris "*fishing*". Mirip seperti asal katanya, *phishing* berarti memancing korban untuk memberikan data yang sensitif ataupun privasi. Salah satu bentuk *phishing* yang sering terjadi adalah meminta korban untuk mengisi data pada halaman *login*. Dengan cara tersebut, oknum bisa mendapatkan *username* dan *password* pengguna dan menggunakannya untuk hal-hal yang tidak diinginkan.

- *Man in the Middle Attack*

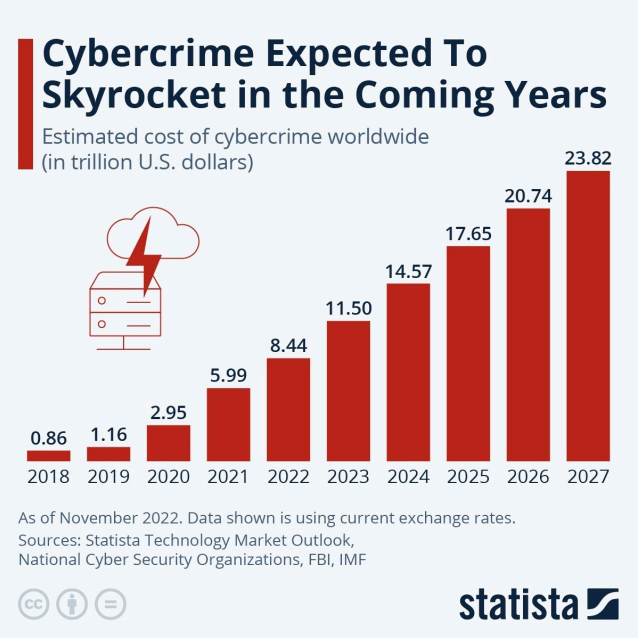
  Serangan *Man in the Middle* adalah serangan siber yang menyusup ke dalam suatu komunikasi. Serangan ini akan berpura-pura menjadi jaringan komunikasi asli sehingga data dikirim ke jaringan ini. Dengan lewatnya data melalui jaringan palsu ini, data tersebut bisa dicuri dan dimanfaatkan oleh oknum tertentu untuk melakukan hal yang tidak diinginkan.

- Serangan Denial of Service

  Serangan Denial of Service adalah serangan yang dilakukan untuk membebani suatu server, sistem, ataupun *website*. Biasanya serangan ini memberikan banyak perintah atau tugas ke suatu server di waktu yang bersamaan.

C. *Dampak Serangan Siber*

Serangan siber bisa menimbulkan banyak dampak negatif. Serangan siber bisa mengakses, mengubah, menghambat, atau merusak informasi sensitif di dalam suatu perangkat tanpa izin. Selain itu, tujuan utama dari serangan siber adalah memeras korban untuk mendapatkan keuntungan sebesar-besarnya. Menurut, World Economic Forum pada tanggal 10 Januari 2024, serangan siber bisa memakan biaya sebesar $23,84 triliun pada tahun 2027 yang sebelumnya hanya sebesar $8,44 triliun pada tahun 2022.



Gambar 1. Prediksi Biaya Dampak Kejahatan Siber (sumber: https://www.weforum.org/agenda/2024/01/cybersecurity-cybercrime-system-safety/)

D. *Modus Kejahatan Tautan Palsu*

Malware, *phishing, Man in the Middle,* dan *Denial of Service* adalah bentuk serangan siber pada tahapan lanjut atau akhir. Namun, sebelum serangan siber di atas bisa terjadi, suatu perangkat harus terhubung ke suatu jaringan atau perangkat lain. Selain itu, membuka suatu *website* atau mengunduh sesuatu hal bisa menjadi awal dari serangan siber. Maka dari itu, setiap *website* yang dibuka harus dipastikan *website* yang asli dan aman.

Dengan semakin maraknya media sosial, sering kali oknum memberikan tautan-tautan palsu dan berbahaya ke segala media. Biasanya tautan-tautan tersebut disertai dengan pesan-pesan yang persuasif dan menarik. Salah satu topik pesan yang sering dipakai adalah tentang uang, hadiah, bonus, dan hal-hal gratis lainnya. Dengan pesan tersebut, oknum mengharapkan penerima pesan menekan tautan palsu yang dikirim sehingga mereka bisa langsung melanjutkan serangannya.

Pada awalnya, tautan palsu tersebut masih berbentuk aneh dan mencurigakan. Namun, semakin sadarnya orang-orang terhadap tautan-tautan palsu tersebut, oknum pun tetap mencari akal untuk dapat mengelabui korban-korban. Oknum-oknum sekarang berusaha untuk menciptakan tautan palsu yang mirip dengan tautan asli dari perusahaan/lembaga/badan ternama.

Banyak cara yang dilakukan oleh oknum untuk dapat membuat tautan *website* yang mirip dengan tautan *website* asli yang terkenal. Berikut ini beberapa cara untuk membuat tautan yang mirip dengan tautan asli:

1. Penggunaan *shorten link*

Dengan menggunakan *shorten link*, tautan palsu akan terlihat lebih rapi. Contoh dari layanan *shorten link* adalah bit.ly dan tinyurl.com. Berikut contoh penggunaan *shorten link,* bit.ly/gratis-hape.

2. Penambahan sedikit karakter

   Oknum akan menambahkan 1 huruf pada tautan palsunya. Contoh dari cara ini adalah tokopediia.com/iphone-gratis. Tautan tersebut mirip dengan tautan tokopedia.com.

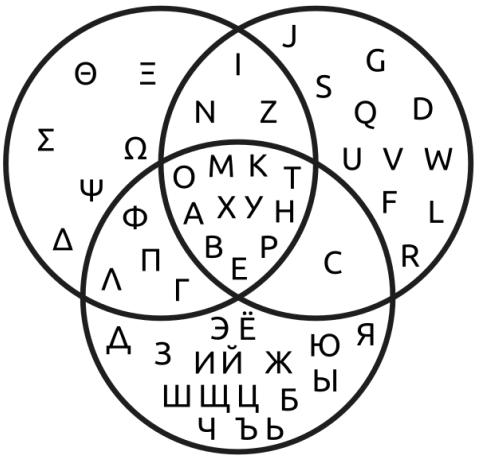
3. Penggunaan domain yang berbeda

   Oknum akan menggunakan domain yang berbeda dengan tautan *website* asli. Contoh dari cara ini adalah shopee.net/pulsa-gratis. Tautan tersebut menggunakan domain yang berbeda dengan *website* asli shopee.co.id.

4. Penggunaan *punycode* (*homograph attack*)

   Punycode adalah metode pengkodean yang digunakan untuk membuat tautan yang bisa mencakup hampir semua karakter di dunia selain ASCII. Dengan ini, aksara-aksara seperti aksara Korea, Jepang, Mandarin, dan Kiril bisa digunakan menjadi tautan suatu *website*.

   Oknum-oknum biasanya menggunakan karakter dari aksara selain latin yang dengan mirip dengan karakter di aksara latin. Para oknum memanfaatkan hal ini untuk menciptakan tautan palsu yang mirip (secara visual) dengan tautan *website* asli. Bahkan, dengan cara ini, ada tautan yang benar-benar sama secara visual, namun berbeda secara sistem komputer. Metode ini biasanya juga disebut *internationalized domain name* (IDN) *homograph attack.*



Gambar 2. Diagram venn Alfabet Latin, Yunani, dan Kiril (sumber: https://www.bitdefenhttps://www.weforum.org/agenda/2024/01/cybersecurity-cybercrime-system-safety/der.com/blog/businessinsights/homograph-phishing-attacks-when-user-awareness-is-not-enough/)

Berikut ini beberapa contoh tautan palsu yang memanfaatkan *punycode*:

Gambar 3. Contoh 1 *Homograph Attack* (sumber: https://www.bca.co.id/id/informasi/awas-modus/2023/06/26/07/22/waspada-website-palsu-menggunakan-abjad-karakter-yang-mirip-dengan-aslinya)

Brand	What the user sees	The Punycode
Adidas	adıdas.de	http://xn--addas-o4a.de/
Aerlingus	aerlıngus.com	xn--aerlngus-j80d.com
Aerlingus	aerlingus.com	xn--aelingus-of0d.com
Air France	airfrance.com	xn--airfrnce-rx0d.com
British Airways	britishairways.com	xn--britishairays-541g.com
British Airways	britishairways.com	xn--britishirways-of2g.com

Gambar 4. Contoh 2 *Homograph Attack* (sumber: https://www.jamf.com/blog/punycode-attacks/)

Target	Homograph	Punycode encoding
amazon.com	amazon.com	xn--amaon-x59a.com
amazon.com	amazon.com	xn--amazn-p29a.com
amazon.com	amɑzon.com	xn--amzon-1jc.com
amazon.com	amàzon.com	xn--amzon-sqa.com
amazon.com	àmäzón.com	xn--mzn-8kav5g.com
amazon.com	amɑzøn.com	xn--amzn-ira28r.com

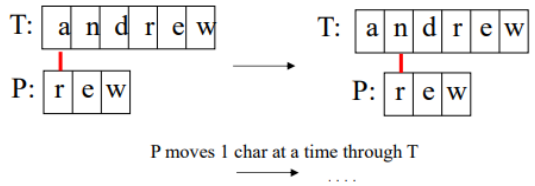
Gambar 5. Contoh 3 *Homograph Attack* (sumber: https://aws.amazon.com/blogs/machine-learning/infoblox-inc-built-a-patent-pending-homograph-attack-detection-model-for-dns-with-amazon-sagemaker/)

E. Pencocokan String

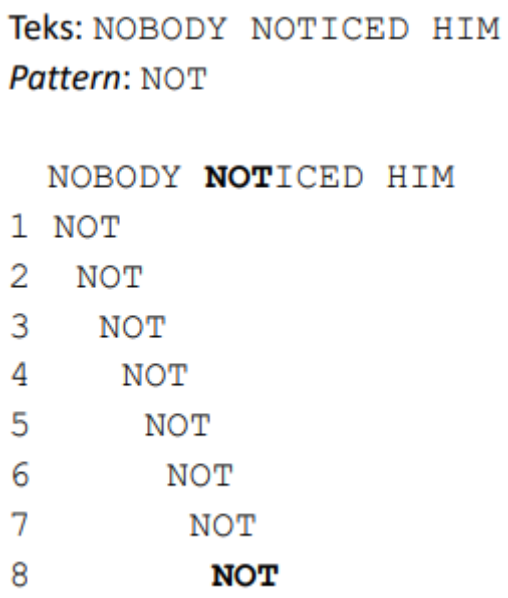
Pencocokan *string* atau *string matching* atau *pattern matching* adalah salah satu algoritma penting di bidang komputer. Pada pencocokan *string*, ada istilah *text* dan *pattern*. *Text* adalah *string* yang tempat dicarinya *pattern*. Sedangkan, *pattern* adalah *string* yang ingin dicari di dalam suatu *text*. Panjang *text* harus lebih besar atau sama dengan panjang *pattern*. Terdapat beberapa algoritma pencocokan *string*. Berikut ini tiga contoh algoritma pencocokan *string*:

- Algoritma *Brute Force*

  Algoritma *Brute Force* adalah algoritma paling sederhana untuk melakukan pencocokan *string*. Setiap karakter di *pattern* akan dikomparasikan dengan karakter di *text* yang sejajar dengan-nya. Jika ada karakter yang tidak sama, *pattern* digeser sebanyak 1 karakter dan dilakukan komparasi ulang setiap karakter.



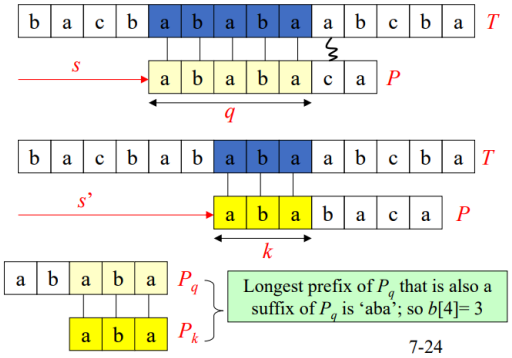
Gambar 6. Ilustrasi 1 Algoritma *Brute Force*



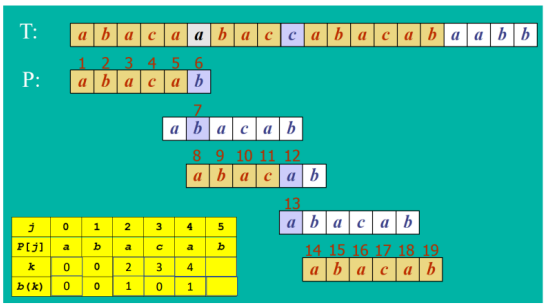
Gambar 7. Ilustrasi 2 Algoritma *Brute Force*

- Algoritma Knuth-Morris-Pratt

  Algoritma Knuth-Morris-Pratt adalah algoritma pattern matching yang menelusuri teks dari kiri ke kanan. Algoritma ini memiliki kemiripan dengan algoritma brute force. Tetapi berbeda dengan algoritma brute force, pergeseran pattern pada algoritma Knuth-Morris-Pratt dilakukan dengan lebih baik/cerdas. Ada sebuah fungsi pinggiran (border) yang menghasilkan suatu list yang bernama Longest Prefix Suffix. LPS akan menyimpan data prefiks terpanjang yang juga merupakan suffix untuk setiap posisi dalam pola. List ini akan digunakan untuk melakukan pergeseran yang lebih baik.



Gambar 8. Ilustrasi Algoritma Knuth-Morris-Pratt

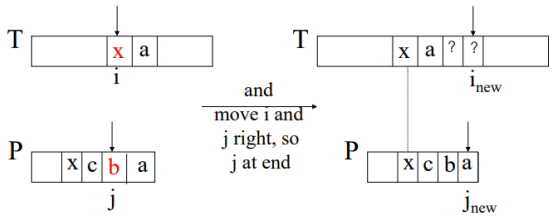


Gambar 9. Ilustrasi Algoritma Knuth-Morris-Pratt versi lain

- Algoritma Boyer-Moore

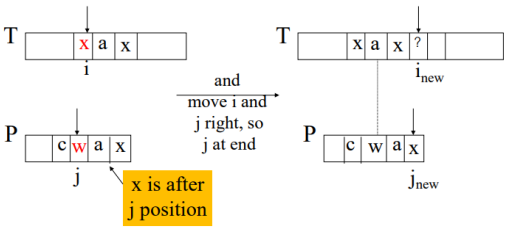
  Algoritma Boyer-Moore adalah algoritma pattern matching yang berlandaskan teknik looking-glass dan character-jump. Teknik looking-glass adalah teknik mencari pattern di dalam text dengan melakukan komparasi secara mundur (dari kanan ke kiri). Teknik character-jump adalah teknik pemindahan karakter yang dibandingkan jika karakter yang sedang dibandingkan tidak sama (*miss-match*). Berikut ini 3 skenario yang mungkin terjadi:

  1. Jika karakter saat ini dari text (misalnya x) ada di sebelah kiri karakter saat ini dari pattern, geser pattern dengan menyejajarkan x dari text dan pattern dan lakukan komparasi ulang dari ujung kanan pattern.



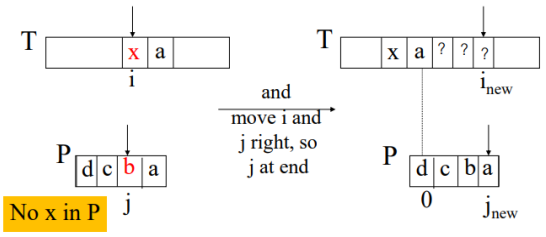
Gambar 10. Skenario Miss-Match 1 Boyer-Moore

2. Jika karakter saat ini dari text (misalnya x) tidak ada di sebelah kiri dan ada di sebelah kanan karakter saat ini dari pattern, geser pattern ke kanan sepanjang 1 dan lakukan komparasi ulang dari ujung kanan pattern.

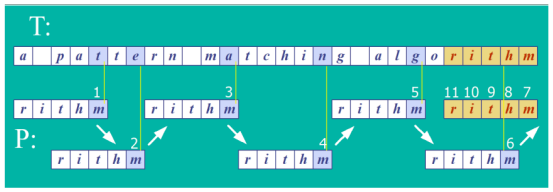


Gambar 11. Skenario Miss-Match 2 Boyer-Moore

3. Jika dua skenario di atas tidak terjadi, geser pattern hingga ujung kiri pattern sejajar dengan setelah karakter saat ini di text dan lakukan komparasi ulang dari ujung kanan pattern.



Gambar 12. Skenario Miss-Match 3 Boyer-Moore

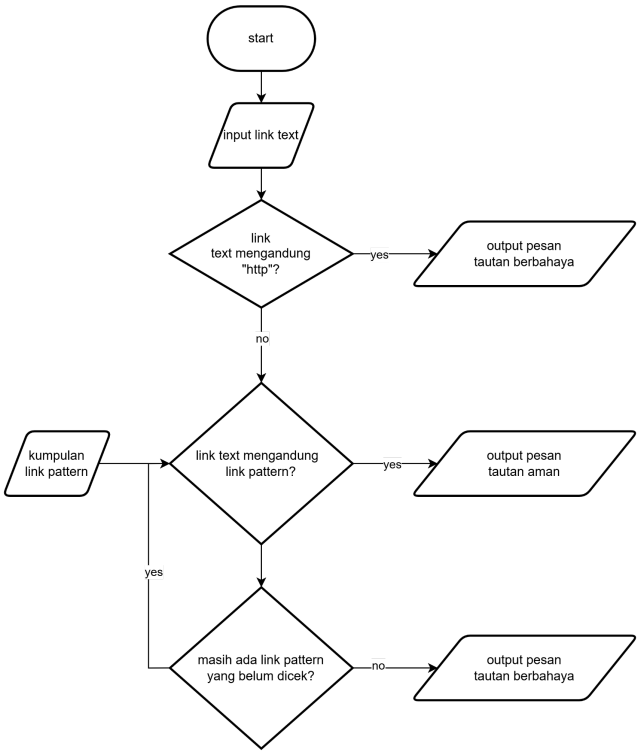


Gambar 13. Ilustrasi Algoritma Boyer-Moore

III. Implementasi

Tautan palsu akan menjadi *text* pada pencocokan *string*. Sedangkan, tautan *website* asli akan menjadi *pattern*. Untuk mendapatkan *pattern-pattern*, akan dicari beberapa tautan *website* asli yang terkenal. Selanjutnya, pencocokan *string* menggunakan algoritma Boyer-Moore akan dilakukan. Pattern juga akan mencakup "https://" dan "http://" untuk mengecek keamanan website. Jika tautan mengandung *pattern* tautan, tautan tersebut asli dan aman. Jika tautan mengandung "http://" dan tidak mengandung "https://", tautan tersebut tidak aman. Berikut ini alur proses program:

1. Meminta input tautan *text*.
2. Mengecek apakah *text* mengandung "http://".
3. Jika iya, program menampilkan pesan tautan tersebut berbahaya dan program berhenti.
4. Mengecek apakah *text* mengandung *pattern*.
5. Jika iya, program menampilkan pesan tautan tersebut aman dan program berhenti.
6. Jika tidak, kembali lakukan proses 4 sampai semua *pattern* telah dicek.



Gambar 13. Alur Program

IV. Pengujian dan Pembahasan

A. *Pengujian*

Untuk pengujian, tautan website asli yang bisa dicek adalah "google.com", "facebook.com", "twitter.com", "linkedin.com", "instagram.com", "reddit.com", "youtube.com", "amazon.com", "wikipedia.org", "github.com", "stackoverflow.com", "medium.com", "netflix.com", "spotify.com", "apple.com", "microsoft.com", "tumblr.com", "pinterest.com", "quora.com", "wordpress.com", "tokopedia.com", "shopee.co.id".

- Pengujian pertama (pengujian biasa)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/User
sihom/AppData/Local/Programs/Python/Python312/python.exe "
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: amazon.com
Tautan aman
```

- Pengujian kedua (pengujian https)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/User
sihom/AppData/Local/Programs/Python/Python312/python.exe "
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: https://google.com
Tautan aman
```

- Pengujian ketiga (pengujian https)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/Users/
sihom/AppData/Local/Programs/Python/Python312/python.exe "c:
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: http://google.com
Tautan berbahaya!
```

- Pengujian keempat (karakter tambahan)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/Users/
sihom/AppData/Local/Programs/Python/Python312/python.exe "c:
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: tokopediia.com/iphone-gratis
Tautan berbahaya!
```

- Pengujian kelima (domain berbeda)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/Users/
sihom/AppData/Local/Programs/Python/Python312/python.exe "c:
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: shopee.net/pulsa-gratis
Tautan berbahaya!
```

- Pengujian keenam (*homograph* mirip)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/Users/
sihom/AppData/Local/Programs/Python/Python312/python.exe "c:
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: médium.com
Tautan berbahaya!
```

- Pengujian ketujuh (*homograph* identik)

```
PS C:\Kuliah\Tingkat 2\Semester 4\Stima\makalah> & C:/Users/
sihom/AppData/Local/Programs/Python/Python312/python.exe "c:
/Kuliah/Tingkat 2/Semester 4/Stima/makalah/CheckLink.py"
Masukkan text: apple.com/iphone-gratis
Tautan berbahaya!
```

*keterangan: karakter a yang digunakan berasal dari aksara Kiril.

B. Pembahasan

Dari hasil pengujian di atas, program ini telah bisa mencegah tautan palsu atau berbahaya dari beberapa skenario. Skenario yang ditangani oleh program ini adalah:

1. Skenario biasa
2. Skenario *https* dan *http*
3. Skenario karakter tambahan
4. Skenario domain berbeda
5. Skenario *homograph* mirip
6. Skenario *homograph* identik

V. KESIMPULAN

Makalah ini telah membahas berbagai macam cara oknum untuk membuat tautan palsu. Dari hal tersebut, solusi dilakukan dengan cara membuat program yang menerapkan pencocokan *string* dengan algoritma Boyer-Moore. Dari hasil pengujian, program ini sudah bisa mendeteksi tautan yang palsu atau berbahaya berdasarkan skenario-skenario yang mungkin dilakukan oleh oknum.

VI. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada:
1. Tuhan Yang Maha Esa
2. Ibu Dr. Nur Ulfa Maulidevi, S.T., M.Sc

TAUTAN GITHUB PROGRAM

https://github.com/Bbennn/CheckLink/tree/main

REFERENSI

[1] Munir, Rinaldi. 2024. "Pencocokan String (String matching/pattern matching", https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf. Diakses pada 12 Juni 2024.

[2] World Economic Forum. 2024. "2023 was a big year for cybercrime – here's how we can make our systems safer", https://www.weforum.org/agenda/2024/01/cybersecurity-cybercrime-system-safety/. Diakses pada 12 Juni 2024.

[3] BCA. 2023. "Waspada Website Palsu Menggunakan Abjad yang Mirip Aslinya", https://www.bca.co.id/id/informasi/awas-modus/2023/06/26/07/22/waspada-website-palsu-menggunakan-abjad-karakter-yang-mirip-dengan-aslinya. Diakses pada 12 Juni 2024.

[4] Liarna La Porta. 2018. "What is Punycode? Fake domains that deceive the human eye",https://www.jamf.com/blog/punycode-attacks/. Diakses pada 12 Juni 2024.

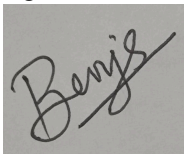
[5] McAfee. 2024. "What is malware and why do cybercriminals use malware? | McAfee?". What is malware and why do cybercriminals use malware? | McAfeen. Diakses pada 12 Juni 2024.

[6] AWS. 2024. "Infoblox Inc. built a patent-pending homograph attack detection model for DNS with Amazon SageMaker", https://aws.amazon.com/blogs/machine-learning/infoblox-inc-built-a-patent-pending-homograph-attack-detection-model-for-dns-with-amazon-sagemaker/ Diakses pada 12 Juni 2024.

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 Juni 2024



Benjamin Sihombing - 13522054