Pendeteksi Phishing Menggunakan Metode Rule Based Attribute Checking

Raja Ryan Pradana[1], Parman Sukarno[2], Erwid Musthofa Jadied[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]rajaryanpradana@students.telkomuniversity.ac.id, [2]parmansukarno@telkomuniversity.ac.id,
[3]jadied@telkomuniversity.ac.id

Abstrak

Internet merupakan bagian penting dari kehidupan saat ini. Namun, seiring dengan berjalannya waktu, internet juga menjadi sarana penyerangan oleh oknum-oknum tidak bertanggung jawab dan salah satu tindak kejahatan ini adalah *phishing*. Dengan *phishing* orang akan dapat mencuri informasi pribadi dari korban. Untuk penanganannya, ada beberapa cara yang bisa dilakukan seperti mendeteksi URL yang akan dimasuki apakah merupakan URL *phishing* atau tidak.

Dari beberapa penelitian terkait pendeteksi *phishing* seperti metode PRISM: An algorithm for inducing modular rules dan Phishing Websites Detection Using Data Mining Classification Model, akurasi yang dihasilkan masih rendah yaitu 84% dan 87% untuk metode tersebut. Oleh karena itu tujuan dari metode penelitian yang diajukan ini adalah untuk meningkatkan akurasi pendeteksian *phishing* tersebut. Metode yang diajukan dalam penelitian ini menggunakan metode Rule Based Attribute Checking dimana sistem akan mengambil sembilan atribut yang paling berpengaruh dalam pendeteksian *phishing* dan setiap atribut diberikan bobot sesuai tingkatannya. Sistem kemudian menjumlahkan total dari bobot yang didapat untuk menentukan status URL yang diperiksa apakah URL tersebut *trustworthy*, *fairly legitimate*, *unsolved*, *suspicious*, atau *phishy*. Dengan menggunakan 200 dataset yang diambil secara acak(100 *phishing* dan 100 *legitimate*) untuk pengujian, metode ini berhasil mencapai akurasi pendeteksian sebesar 85.5%.

Kata kunci : *Phishing*, pendeteksi, atribut, *rule based*

Abstract

Internet is an important part of life today. However, over time, the internet is also a means to attack by unscrupulous individuals and one of these crimes is phishing. With phishing people will be able to steal personal information from victims. To handle this, there are several ways that can be done and one of them is detecting the URL whether it is a phishing URL or not before entering the URL.

There have been several studies related to phishing detection such as PRISM: An algorithm for inducing modular rules dan Phishing Websites Detection Using Data Mining Classification Model. But, the accuracy of these methods can only reach 84% and 87%. Therefore with this new proposed method we aim for higher accuracy in detecting phishing. This proposed method use Rule Based Attribute Checking where the system will check nine attributes that have the most impact in detecting phishing and the system will give certain amount of score for each attribute. After accumulating the total score that the URL gets, the system will determine it's status ranging from trustworthy, fairly legitimate, unsolved, suspicious, or phishy. Using 200 random datasets(100 phishing and 100 legitimate) for testing purpose, we managed to achieve a total accuracy of 85.5% with this method

Keywords: Phishing, detector, attribute, rule based

1. Pendahuluan

Latar Belakang

Internet pada masa ini merupakan hal yang tidak bisa lepas dari kehidupan sehari-hari. Hal ini pun membuat orang saat ini bergantung pada internet dan sudah sangat banyak kegiatan yang mewajibkan terhubung internet sehingga jika internet di dunia ini berhenti untuk satu tahun maka akan merusak alur kegiatan penting seperti perbankan, penerbangan, dan juga akan menyebabkan kerugian yang sangat besar. Menurut situs *brooking*, matinya internet dapat menyebabkan kerugian negara sebesar 2.4 miliyar USD [6].

Sayangnya, pengguna internet tidak paham bahwa banyak pencuri data di internet. Ini yang menjadi pemicu orang-orang sering memasukkan data dan informasi pribadi penting yang harusnya tidak boleh diketahui oleh siapapun. Melihat peluang ini banyak bermunculan orang-orang yang mencoba untuk bisa mencuri data penting

orang lain, yang salah satunya adalah *phishing*. Ada banyak metode yang dilakukan oleh oknum untuk melakukan *phising* ini dan yang paling marak seperti meniru *website* resmi yang cukup terkenal atau menggunakan iklan dan memberikan *link* yang yang nantinya membuat korban masuk ke *website* dimana oknum dapat mencuri data yang diisi oleh korban didalamnya.

Alih-alih menghentikan pemakaian internet untuk menghindari *phishing*, justru hal ini bisa dihindari dengan pendeteksian dini sebelum oknum tersebut berhasil mencuri data pribadi. Penelitian tentang pendeteksian phising ini sudah banyak dilakukan seperti yang dilakukan Mohammad menggunakan metode *Rule-Based* yang menggunakan beberapa atribut yang ada di *website* sebagai parameter untuk digunakan dan memiliki frekuensi kemunculan 1.6%-100% yang mengindikasikan *phishing* tergantung parameter yang digunakan [3]. Lalu Jabri dan Ibrahim mengusulkan metode algoritma PRISM yang sudah di modifikasi dan berhasil mengungguli pendahulu nya dengan akurasi 87% [4]. Selain itu terdapat juga metode PRISM [1] yang memiliki akurasi 84%.

Pada penelitian ini, aplikasi pendeteksi situs *phishing* akan dibangun dalam basis *desktop* dalam bahasa pemrograman Java. Metode yang digunakan adalah *rule-based* dan *attribute checking*. Aturan yang diterapkan dalam metode ini ada sembilan atribut yang dijadikan parameter dengan perbedaan bobot nilai masing-masing parameter untuk mendapatkan nilai akurasi yang lebih baik. Hasil dari penelitian ini adalah nilai akurasi sebesar 85.5%.

Topik dan Batasannya

Rumusan masalah yang diangkat adalah rendahnya tingkat akurasi pendeteksian URL *phishing* menggunakan metode yang ada saat ini. Penelitian ini memiliki batasan masalah sebagai berikut:

1. Sistem harus terkoneksi internet;
2. Serangan *phishing* yang dapat dideteksi berbentuk URL.

Tujuan

Tujuan yang ingin dicapai pada penelitian ini adalah :

1. Meningkatkan tingkat akurasi pendeteksian URL *phishing* dengan menggunakan metode *Rule Based Attribute Checking*.

Organisasi Tulisan

Paper ini terdiri dari lima buah bagian: Pendahuluan, Studi Terkait, Sistem yang Dibangun, Evaluasi, dan diakhiri dengan bagian Kesimpulan.

2. Studi Terkait

Penelitian terkait pendeteksian *phishing* adalah sebagai berikut :

1. *Phishing*
   Phishing merupakan salah satu jenis serangan siber dimana penyerang memcancing target untuk mengakses situs palsu yang menyerupai situs aslinya. Cara yang paling populer adalah mengirimkan URL tersebut melalui surel, meskipun seiring berjalannya waktu penggunaan aplikasi pesan instan mulai dilirik oleh penyerang.
   Ada beberapa ciri yang paling mudah dilihat pada situs *phishing* seperti URL yang terlalu panjang, penggunaan IP pada URL dan beberapa ciri lain yang bisa langsung terlihat secara langsung. Meskipun begitu, ada juga URL *phishing* yang sangat menyerupai URL aslinya (penggunaan karakter spesial) dimana untuk mendeteksi keaslian situs tersebut perlu memeriksa hal-hal krusial yang tidak dapat dimanipulasi oleh si penyerang seperti sertifikat, usia domain, hingga peringkat situs tersebut.

2. Pendeteksian *phishing* dengan metode PRISM yang sudah dimodifikasi. [4]
   PRISM yang dimodifikasi adalah metodologi yang berorientasi *attribute-value*. Metodologi ini pertama mempelajari data dari *dataset* untuk menentukan parameter yang berpengaruh dengan menggunakan program yang mereka kembangkan. Akurasi dari metode ini mencapai 87%.

3. Pendeteksian *phishing* dengan metode *Intelligent Rule Based*[3]
   Intelligent Rule Based adalah metode pendeteksian *phishing* yang diusulkan oleh Rami. Metode ini mengambil atribut dari URL dan konten nya yang kemudian dijadikan parameter. Menurut penelitian ini, terdapat sembilan parameter yang paling berpengaruh dalam pendeteksian *phishing*.
   Keunggulan yang dimiliki *paper* ini adalah parameter pendeteksian *phishing* yang digunakan sudah diuji dan sudah ditentukan parameter mana yang paling bagus digunakan dalam pendeteksian.
   Kekurangan nya adalah mereka melakukan pengujian terhadaap masing-masing parameter sehingga tingkat keakuratan keseluruhan nya tidak dapat diketahui.