Makalah Tugas Kuliah EL6115

Operasi Keamanan dan *Incident Handling*

REVIEW TEKNIK DETEKSI DAN KLASIFIKASI *ANTI-PHISHING*

Oleh:

AZZAHRA RATU KAMILA

23214302

Dosen:

Dr. Ir. Budi Rahardjo

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

PROGRAM MAGISTER TEKNIK ELEKTRO

INSTITUT TEKNOLOGI BANDUNG

2016

ABSTRAK

Phishing adalah upaya penipuan yang dilakukan pelaku untuk bisa mendapat data pribadi dari korban seperti informasi akun bank, nomor kartu kredit, jaminan sosial, dan lain-lain. Serangan phishing biasanya menggunakan e-mail dan website yang didesain untuk menjebak user memberikan data personalnya dengan mengatasnamakan nama bank atau perusahaa ternama, e-commerce, dan perusahaan kartu kredit. Ada beragam cara untuk melakukan penanggulangan terhadap serangan phishing. Review pada makalah ini bertujuan untuk menggali lebih dalam cara melakukan deteksi dan klasifikasi anti-phishing serta kelebihan dan kekurangannya.

Kata kunci: *phishing*; *user*; *cyber*; *anti-phishing*; *e-mail*; *website*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I PENDAHULUAN

Dalam kurun waktu satu dekade, perkembangan transaksi *online* berkembang dan meningkat tajam. Namun seiring pertumbuhan *e-commerce,* jumlah pelaku kejahatan *cyber* juga meningkat drastis. Para pelaku kejahatan *cyber* ini sebagian besar menggunakan *phishing* sebagai cara untuk melakukan ekspansi di World Wide Web.

Phishing adalah upaya penipuan untuk mendapatkan data pribadi dari korban seperti informasi bank, nomor kartu kredit, jaminan sosial, rincian pekerjaan, akun dan *password online shopping*, dan lain sebagainya. Serangan *phishing* menggunakan *e-mail* atau *website* yang didesain sedemikian rupa untuk mengelabui user untuk memberikan data personal finansialnya dengan cara menggunakan nama bank atau perusahaan ternama, *e-commerce*, dan perusahaan kartu kredit [1].

Menurut laporan dari Gartner Inc., selama setahun terakhir sebelum April 2004, terdapat 1,8 juta korban serangan phishing, dan kerugian dari penipuan tersebut mencapai $ 1,2 milyar [2]. APWG (Anti-Phishing Working Group) menyediakan *solution directory* yang memuat informasi tentang *anti-phishing* yang digunakan di perusahaan-perusahaan kelas dunia.

Hingga saat ini metode *blacklisting* merupakan salah satu anti-*phishing* yang sering dipakai untuk melakukan pencegahan, tetapi seiring berkembangnya teknologi maka kejahatan *cyber*, khususnya *phishing*, dibutuhkan cara-cara lain untuk bisa mendeteksi *phishing* sedini mungkin. Dalam makalah ini akan dibahas mengenai skenario kasus phishing, atribut yang biasa digunakan untuk mengidentifikasi website asli dengan website phishing, serta teknik *anti-phishing.* Makalah ini akan memberikan gambaran atau perspektif mengenai *phishing*, solusi beserta kelebihan dan kekurangannya, dan dapat membantu untuk prediksi terhadap arahan penelitian selanjutnya mengenai *phishing.* Bab I berisi tentang pendahuluan, Bab II berisi tinjauan pustaka mengenai penelitian yang berkaitan dengan topik makalah, Bab III berisi rincian analisa, dan Bab IV berisi kesimpulan dari makalah ini.

BAB II TINJAUAN PUSTAKA

Pada bab ini diberikan tinjauan pustaka mengenai penelitian yang berhubungan dengan topik makalah, antara lain: klasifikasi serangan phishing, skenario kasus phishing, metode deteksi dan klasifikasi menggunakan machine learning based.

2.1 Klasifikasi Serangan Phishing

Serangan phishing memiliki beberapa kategori berdasarkan cara pelaku mencuri informasi. Beberapa peneliti menyebut dengan nama berbeda seperti *deceptive*, *malware based*, *content-injection phishing,* dan lain-lain.

a) Deceptive Phishing

Webpage yang dimasuki oleh korban akan meminta korban untuk memasukkan informasi pribadi untuk melakukan verifikasi akun, penagihan akun fiktif, sistem malfungsi yang mensyaratkan korban untuk memasukkan informasi, layanan gratis yang membutuhkan keaktifan korban untuk mendapatkan layanan tersebut, dan berbagai tawaran menarik lain untuk menarik perhatian para korban mengklik *link phishing.*

b) Web Spoofing

Web spoofing merupakan serangan keamanan untuk meneliti dan mengubah semua halaman yang dikirim ke *user machine*, kemudian meneliti semua informasi yang dimasukkan oleh korban yang menggunakan *user machine* tersebut untuk mengambil data serta informasi. Cara kerja *web spoofing* cukup berbahaya karena meskipun *browser* yang digunakan menjamin keamanan koneksi, masih bisa ditransformasi oleh pelaku untuk keuntungannya sendiri.

c) Email Spoofing

Merupakan *spoofing* dalam bentuk *e-mail,* yaitu ketika nama pengirim pada *e-mail* hingga *header e-mail* diubah sedemikian rupa agar serupa atau mirip dengan *e-mail* dari sumber yang sah.

d) Malware Based Phishing

Teknik ini menggunakan kode *malicious* pada *user machine* yang bisa mendeteksi informasi atau data rahasia yang dimasukkan oleh korban saat menggunakan akun tertentu di komputernya. *Malware* bisa dalam berbentuk *attachment* yang disusupkan setelah mencari celah lemah keamanan, menjadi suatu *file* yang terlihat aman diunduh.

e) Tabnabbing

Banyak pelaku kejahatan *cyber* yang memanfaatkan kebiasaan korban yang suka membuka banyak *tab* di halaman *browser.* Pelaku kemudian menggunakan kesempatan inin untuk mengakses informasi melalui *website* yang sering dikunjungi user melalui *cookies.* Kemudian pelaku akan membuat suatu *website* yang diserupakan dengan *website* yang sering dikunjungi korban untuk meminta data akun dan *password.*

f) Session Hijacking

Serangan ini merupakan aktivitas saat pelaku menunggu korban untuk masuk ke akun, misalkan akun *e-banking*. Pada saat itu suatu *software* yang sudah ditanam di komputer korban akan mengambil alih kontrol dengan melakukan aktivitas tidak sah seperti melakukan transfer rekening tanpa sepengetahuan korban.

g) Search Engine Phishing

Pelaku akan membuat suatu *website* semenarik mungkin yang memberikan penawaran menggiurkan. *Website* tersebut ditaruh di *index* mesin pencari agar korban menganggap bahwa penawaran tersebut bukan penawaran yang berbahaya. Pelaku mendapatkan data korban saat korban memasukkan datanya secara sukarela ke *website* ini.

h) DNS-based phishing

Phishing berbasis *Domain Name System* (DNS) atau *host* biasa disebut sebagai *pharming.* Permintaan untuk *Uniform Resource Locators* atau *name service* menyambungkan alamat pada alamat yang sudah diganggu oleh pelaku dengan memutus komunikasi antara *browser* korban dengan *host files* atau nama domain perusahaan yang sebenarnya.

2.2. Skenario *Phishing*

Gambar 2.1 Proses Skenario Website/URL *Phishing*

Gambar 2.1 menunjukkan skenario umum yang biasa digunakan oleh pelaku saat melakukan kejahatan *phishing.* Ratusan *e-mail* akan dikirim berikut dengan URL palsu yang mengarah pada *website* buatan pelaku. Selanjutnya jika korban percaya bahwa *e-mail* tersebut sah, berasal dari bank, perusahaan kartu kredit, organisasi, atau perusahaan asuransi tertentu, maka korban akan mengirimkan data pribadi pada pelaku. Akun yang dimiliki pelaku bisa digunakan untuk berbagai aktivitas tidak sah, antara lain transfer rekening ilegal, belanja menggunakan kartu kredit korban, menjual akun tersebut ke satu pihak, dan lain-lain.

2.3. Atribut Untuk Membedakan Antara *Website* Asli dan *Website* Phishing

Banyak penelitian yang sudah dilakukan untuk bisa menciptakan tool yang lebih baik dan lebih efisien, tetapi hingga saat ini sangat sulit untuk benar-benar menghilangkan sumber masalahnya atau untuk mengestimasi jumlah user yang berhasil terpancing dan menjadi korban *phishing.*

Salah satu cara untuk bisa membedakan antara *website* asli dengan *website phishing* bisa dengan mengidentifikasi perbedaan di antara keduanya. Parameter tersebut terbagi ke dalam enam kriteria yang ditunjukkan pada Tabel 2.1.

Tabel 2.1 Parameter untuk Mengidentifikasi Website Asli dan *Phishing* [3]

Jenis-Jenis Identifikasi	No.	Parameter Identifikasi
Identitas berbasis web link	1	Menggunakan IP address
	2	Permintaan link yang abnormal
	3	Link abnormal
	4	Rekaman DNS abnormal
	5	URL berbahaya
Enkripsi dan keamanan	1	Menggunakan sertifikasi SSL
	2	CCIA
	3	Detail cookie yang anomali
	4	Nama sertifikasi yang unik
	1	Pengarahan berulang ke page
erifikasi berbasis source code dan client sic	2	Link tersembunyi
	3	Serangan straddling
	4	Server form handler
	5	Serangan pharming
Page layout dan konten	1	Penyamaan lafal
	2	Website kopian
	3	Meniadakan klik kanan
	4	Kotak dialog pop-up
	5	Mengisi informasi dengan menggunakan tombol submit
Bar web address	1	Web address yang panjang
	2	Mengganti huruf yang serupa di URL
	3	Terdapat sufiks dan prefiks tambahan pada URL
	4	Menggunakan simbol @
Faktor sosial	1	Membutuhkan waktu cukup lama untuk mengakses akun
	2	Security sensitive
	3	Solusi publik umum

BAB III PEMBAHASAN

3.1 Teknik *Anti-Phishing*

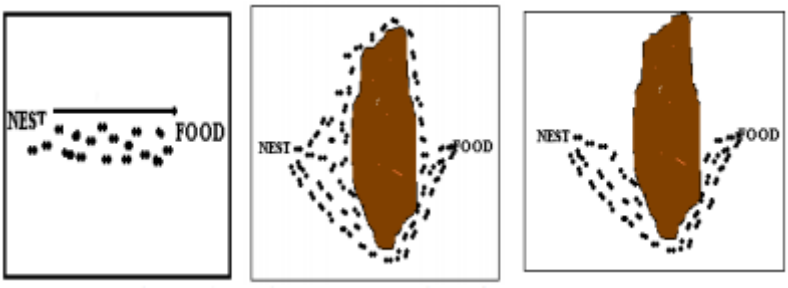
Hingga saat ini penelitian mengenai cara mendeteksi phishing masih terus dikembangkan. Solusi yang tepat untuk mencegah atau mengatasi *phishing* adalah dengan *anti-phishing*. Pada bab ini akan diterangkan mengenai kelebihan dan kekurangan teknik *anti-phishing* yang umum dipakai.

3.1.1 Pendekatan Berbasis *List*

Anti-phishing yang paling sering digunakan dengan pendekatan berbasis *list.* Terdapat dua jenis yaitu *white list* yang merupakan sekumpulan basis data URL yang dianggap aman dan *black list*, URL yang memiliki kemungkinan membahayakan user. Ada beberapa cara untuk melakukan sistem berbasis list ini yaitu Phish Tank Site Checker, Google Safe Browsing, Fire Phish, dan Calling ID Link Advisor. Hasil dari *anti-phishing* akan berbentuk semacam *toolbar* atau ekstensi yang mengingatkan pada user jika mereka ingin mengakses *website* tertentu. Diperlukan *update* untuk mengetahui URL *white list* dan URL *black list.*

3.1.2 Optimisasi Koloni Semut

Ilustrasi dari sistem koloni semut terdapat pada Gambar 3.1.



Gambar 3.1 Perilaku Pergerakan Koloni Semut

Optimisasi digambarkan serupa dengan semut. Pada gambar pertama diperlihatkan semut yang berbaris saat sedang pergi menuju obyek. Ketika sudah sampai, maka koloni semut akan

terbagi ke dalam dua kelompok yang akan bergerak ke arah ke kanan dan ke kiri untuk menghindari obyek tersebut. Feromon yang disebar semut akan berkumpul sangat cepat di jalur yang lebih pendek. Semut memiliki insting untuk mengikuti alur yang memiliki jumlah feromon lebih banyak, sehingga semut akan menilai jalan yang menurutnya lebih pendek untuk mencapai tujuannya. Hal ini juga diadopsi oleh ACO (*Ant Colony Optimization*).

3.1.3 PhishZoo

Cara ini bisa langsung mendeteksi website *phishing* yang dibuat serupa dengan *website* asli dengan membandingkan antar konten *webpage*. Agar tidak terdeteksi, biasanya pelaku akan membuat *website phishing* tersebut harus dibuat semirip mungkin[4].

3.1.4 K-Nearest Neighbor (k-NN)

k-NN merupakan *classifier* yang dibuat dengan cara melakukan *filtering* terhadap *e-mail phishing.* Berdasarkan input data k-terdekat, maka sampel akan dipilih menggunakan fungsi yang sudah didefinisikan pada k-terdekat sebelumnya. Setelah itu *e-mail* tersebut akan diberi label sesuai kelas yang sama dengan *bulk* di antara k-set yang ada[5].

3.1.5 Pendekatan Berbasis *Webpage Source Code*

Strategi ini menggunakan deteksi dengan melakukan pengecekan pada *webpage source code.* Karakter *website phishing* diambil melalui standar W3C untuk mengevaluasi keamanan pada *website.* Setiap karakter pada *webpage source code* akan dicek satu persatu, jika terdapat atribut atau karakter dari *website phishing*, maka tingkat massa keamanan akan diturunkan. Setelah itu akan dilakukan perhitungan tingkat persentasi keamanan berdasarkan massa terakhir dari *website* tersebut dengan mengecek *source code* dari *website* asli dan *website phishing*[1]. Contoh dari atribut terdapat pada Gambar 3.2.

phishing characteristics
1. Different domain
2. External images
3. Suspicious URLs
4. Domain tag
5. Iframe
6. Email
7. Suspicious Script
8. Popup window

Gambar 3.2 Karakter *Phishing* [1]

3.1.6 *Fuzzy Logic*

Fuzzy logic bisa dipakai untuk melakukan pengukuran risiko operasional[6]. Teknik ini bisa memberikan informasi lebih pada manajer risiko untuk bisa melakukan *assessment* dan membuat *ranking* risiko *website phishing*. Risiko ini sudah diperhitungkan secara kuantitatif berbasis pada amalgamasi dari data histori dan input.

Hasil dari pembahasan pada subbab 3.1 terdapat pada Tabel 3.1.

Tabel 3.1 Kelebihan dan Kekurangan Teknik *Anti-Phishing*

Teknik	Kelebihan	Kekurangan
Pendekatan Berbasis List	- Akurasi 100% berdasarkan keputusan untuk melakukan blacklisting website - False positive rate yang dihasilkan rendah - Biaya lebih murah dan mudah digunakan	- Memproduksi memory overhead yang banyak - Jika tidak di-update maka tingkat akurasi blacklisting berkurang drastis
Optimisasi Koloni Semut	- Pendekatan ini menghasilkan tingkat akurasi yang cukup dengan menentukan aturan terbaik untuk mencapai tujuan - Bisa dipakai untuk lingkungan dinamis - Semua koloni memiliki memori	False negative rate bisa jauh meningkat jika dibandingkan dengan pendekatan yang lain
PhishZoo	- Bisa mengklasifikasi zero-day phishing dan serangan target	- Memerlukan image site yang sepadan

	- Pendekatan ini bisa mendeteksi serangan baru - Mengurangi false positive rate	- Kurang cocok digunakan untuk mendeteksi website phishing
K-Nearest Neighbor	- Lebih bisa diandalkan untuk mendapatkan rating positif yang tinggi - Bisa mencapai tingkat akurasi yang tinggi	- Membutuhkan banyak angka - Biaya tinggi - Membutuhkan memori yang banyak
Pendekatan berbasis webpage source code	- Tingkat persentasi keamanan suatu website bisa didapatkan dengan membandingkan antar source code - Cara pengecekan lebih simpel	- Masih terbatas pada source code PHP
Fuzzy Logic	- Memori yang dipakai jauh lebih sedikit - Kecepatan dalam melakukan analisis tinggi	- Tidak 100% efektif - Desain kompleks

BAB IV KESIMPULAN

Dalam makalah ini berbagai teknik *anti-phishing* yang sudah dikemukakan pada penelitian sebelumnya dibahas untuk mengetahui cara kerja, kelebihan dan kekurangannya. Sebagian besar dari teknik *anti-phishing* fokus pada konten *webpage*, URL, dan *e-mail*. Sebagian memiliki tingkat keakuratan yang tinggi dibanding dengan yang lain, tetapi rata-rata hanya bisa mendeteksi jenis serangan yang sudah masuk ke dalam list atau sudah diketahui dan memiliki biaya yang tinggi.

Untuk penelitian yang akan datang, makalah ini bisa dipakai sebagai rujukan untuk mengembangkan teknik baru yang bisa mendeteksi *threat* secara akurat dan mengurangi *memory overhead*, serta mengurangi *false positive rate.*

DAFTAR PUSTAKA

[1] O.A. Batarfi, & M.G. Alkhozae, "Phishing Websites Detection based on Phishing Characterisics in the Webpage Source Code," *International Journal of Information and Communication Technology Research,* Vol. 1, No. 6, October 2011.

[2] David Geer, "Security Technologies Go Phishing", *IEEE Computer* 38(6):18-21, 2005.

[3] Gaurav, Madhuresh Mishra, Anurag Jain, "Anti-Phishing Techniques: A review," *International Journal of Engineering Research and Applications*, Vol.2, Issue 2, pp. 350 -355, March-April 2012.

[4] Sadia Afroz, Rachel Greenstadt, "PhishZoo: Detecting Phishing Websites By Looking at Them," *Fifth IEEE Interational Conference On Semantic Computing (ICSC) 2011*, pp. 368-375, Sept 2011.

[5] Jagruti Patel, Sheetal Mehta, "A Literature Review On Phishing E-mail Detection Using Data Mining," *International Journal of Engineering Sciences & Research Technology*, Vol.4, Issue 3, March 2015.

[6] S. Shah, "Measuring Operational Risks using Fuzzy Logic Modeling," Article, Towers Perrin, July 2003.