Sistema Multi-agente para la Detección de Fraudes en el Correo Electrónico

Christian J. Lucero[a], Perla J. Castro[a], María de Guadalupe Cota[a], Pedro Flores Pérez[a], Juan P. Soto Barrera[a]

[a]*Universidad de Sonora*
Blvd. Luis Encinas y Rosales S/N, Colonia Centro, CP 83000 Hermosillo, Sonora, México.
[crizlucero,janethkasztro]@gmail.com, [lcota,pflores,jpsoto]@gauss.mat.uson.mx

2013 Published by $\mathcal{DIFU}_{100}ci@$ *http://www2.uaz.edu.mx/web/www/publicaciones*
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En este trabajo se presenta un sistema multi-agente para la detección de amenazas y fraudes a través del correo electrónico. Esta propuesta utiliza la tecnología de agentes software para identificar aquellos correos considerados como amenazas de fraude para los usuarios. Para esto, los agentes hacen uso de reglas tipo Prolog con el fin de detectar patrones que permitan identificar correos electrónicos maliciosos, y cuya intención sea obtener información confidencial de los usuarios mediante el uso de phishing y bots.

Palabras clave: Análisis de Contenido, Bots, Correo electrónico, Multi-agente, Phishing..

1. Introducción

En la actualidad el uso del correo electrónico es uno de los medios de intercambio de información más utilizados por los usuarios en internet. Por lo tanto, también es uno de los medios más utilizados por los crackers o maleantes cibernéticos para engañar a las personas y vulnerar sistemas. Existen varios métodos de robo de información, siendo uno de ellos el phishing, cuyo objetivo es obtener información confidencial de las personas a través de engaños [1]. Otro método comúnmente utilizado son los bots (abreviatura de robots), el cual es un programa informático que simula a un ser humano para interactuar con las personas y estafarlas [2]. Dicho programa envía archivos infectados a través del correo electrónico. Con el fin de buscar una alternativa que contribuya a disminuir dicha problemática, en este trabajo se propone un sistema multi-agente para la detección de amenazas en el correo electrónico. En la Sección 2, se describen dos de los métodos más utilizados en el correo electrónico para engañar y estafar a las personas: phishing y bots. En la Sección 3, se modela el sistema multi-agente, para esto, se utiliza la metodología INGENIAS. Por otro lado, en la Sección 4 se describe el desarrollo e implementación del sistema propuesto. Por último, en la Sección 5 las conclusiones son presentadas.

2. Amenazas de estudio

En la actualidad podemos encontrar diversos tipos de programas utilizados para estafar a las personas a través de sistemas de gestión de correos electrónicos, por ejemplo: Adware, Phishing, Spyware, Riskware, Bots etc., así como métodos basados en ingeniería social, tales como el Phishing. Este trabajo se centra en identificar amenazas basadas en el phishing con ayuda de los bots. Es por esto que en los siguientes apartados se aborda más a detalle cada uno de estos temas.

2.1. Phishing

El termino phishing se empezó a definir alrededor de 1995, y consiste en enviar correos electrónicos con el fin de obtener información personal del usuario, para posteriormente utilizarla en algún tipo de fraude [3]. Por ejemplo, el contenido de un correo suelen incluir un enlace a un sitio conocido pero re-direccionando a una página web falsa. De esta manera, el usuario se confía y proporciona información personal a un sitio web falso en donde se realiza la estafa [4]. Los phishing más conocidos son las “estafas nigerianas”. Este tipo de estafas tienen al menos 5 variantes [5]:

- Estafa basada en el depósito de una fuerte cantidad de dinero.
- Estafa por lotería.
- Estafa por herencia de familiar desconocido.
- Estafa de un prisionero de guerra.
- Estafa de venta de celulares o recargas de tiempo aire.

2.2. Bots

Un bot (abreviatura de robot) es un programa o script que realiza tareas automáticamente. Una de estas tareas consiste en imitar el comportamiento del ser humano al intercambiar información con otras personas. Por lo general, los bots se utilizan para enviar correos electrónicos a diversos usuarios, creando continuidad en el intercambio de mensajes, y haciendo creer al usuario que está interactuando con una persona real [2]. Este tipo de fraude es conocido como “Fraud bots”, lo cuales envían correos de manera insistente hasta lograr engañar a los usuarios y así obtener información personal y bancaria, ofreciéndoles falsos premios, herencias, etc., [6]. Debido a que este tipo de fraudes representa

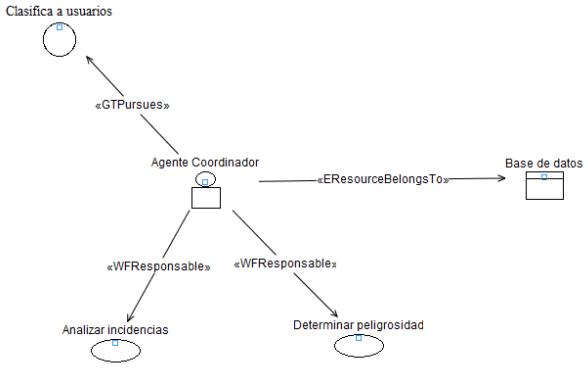

Figura 1. Agente Coordinador.

uno de los métodos menos costosos para las asociaciones delictivas, han logrado convertirse en los ataques más utilizados para engañar a las personas a través del correo electrónico.

3. Modelo del Sistema Muti-agente

Como una alternativa para proteger a los usuarios de las amenazas descritas en la sección anterior, en este apartado se presenta el modelo del sistema multiagente propuesto siguiendo la metodología de INGENIAS [7]. El sistema se compone de un conjunto de agentes los cuales interactúan y realizan tareas enfocadas en proteger al usuario del correo electrónico de posibles ataques. Para esto, se cuenta con un agente encargado de analizar correos electrónicos entrantes. Otro agente que de acuerdo a las incidencias del remitente decide mover aquellos correos considerados como posibles amenazas a la carpeta de “No deseados”. Y por último, un agente que proporcione seguridad en el envío de mensajes entre los agentes. A continuación se describe a detalle cada uno de estos agentes:

- Agente coordinador. Es el agente encargado de coordinar a los agentes que intervienen en el sistema. Además determina si el remitente del correo electrónico entrante es una amenaza para el usuario. Para esto, el agente coordinador analiza las incidencias presentadas y determina la peligrosidad de las mismas (ver Figura 1).

- Agente local. Es el agente encargado de detectar aquellos correos identificados como phishing (Figura 2).

Para esto, el agente obtiene los correos entrantes y analiza el contenido del mensaje recibido. Dicho análisis

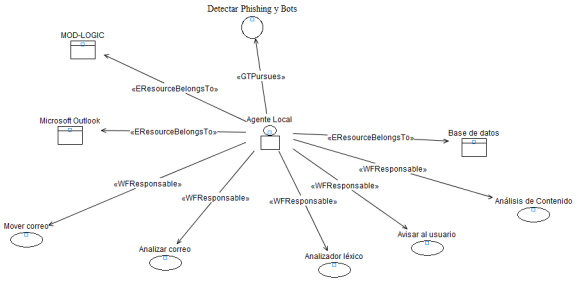

Figura 2. Agente Local.

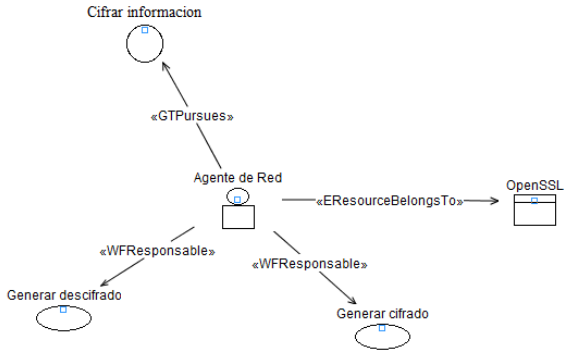

Figura 3. Agente de Red.

permite al agente enviar alertas al usuario en caso de que detecte una amenaza. Para realizar el análisis de las palabras, el agente local utiliza la aplicación MOD-LOGIC.

- Agente de red. Es el agente encargado de proveer la seguridad durante el intercambio de información entre el agente local y el agente coordinador (ver Figura 3).

Una vez descritos los agentes que componen el sistema, a continuación se ilustra el modelo de la organización los agentes (ver Figura 4).

Por cuestiones de espacio el modelo del entorno y el de objetivos y tareas del sistema, se explican en [8].

4. Prototipo

El prototipo propuesto fue implementado en .NET, con el uso de dll´s para acelerar el proceso de análisis.

Detalles de implementación:

Como parte de la programación del conocimiento de los agentes, el sistema utiliza reglas tipo Prolog y técnicas de análisis de contenido (apartado 4.1). Por otra parte, y con el fin de proteger la información que se

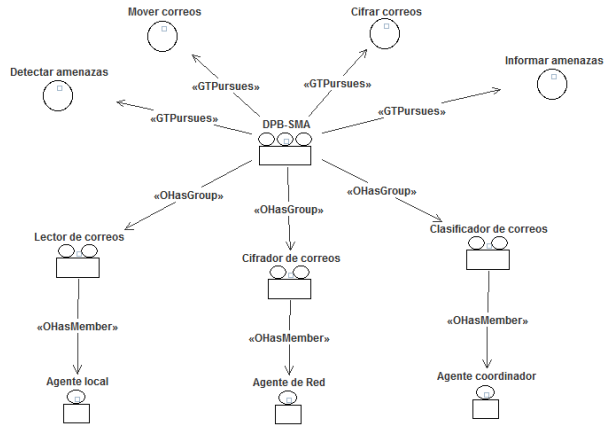

Figura 4. Modelo de la Organización.

intercambia a través de la red se utilizarán las librerías OpenSSH [9] y OpenSSL [10]. Además se diseñó un analizador léxico (tokenizador), el cual descompone la frase en tokens para un manejo más sencillo en el análisis de contenido (apartado 4.2), una base de datos para el control de sesiones de agentes, un glosario con una clasificación ontológica de las palabras que son consideradas con cierto grado de peligrosidad para el usuario, y una pizarra para el control de ocurrencias de aquellos eventos considerados como amenazas para la seguridad del sistema. La base de datos almacena las palabras claves consideradas “peligrosas”, así como el registro de las amenazas en un historial.

4.1. Análisis de Contenido

El análisis de contenido es un conjunto de técnicas utilizadas por los agentes para distinguir el significado simbólico de los mensajes. Estos mensajes, por lo regular, no tienen un único significado, ya que en ocasiones cambia su semántica según el contexto en que se presenta cierta información. Por lo tanto, el objetivo del análisis de contenido consiste en realizar una inspección en las frases para identificar aquellas que representen un peligro para el sistema o para el usuario. Para esto, el sistema considera aquellos casos y frases que por lo general son utilizadas para cometer fraudes. A continuación se describen algunos de los casos considerados, así como las frases asociadas a cada uno de ellos (ver Tabla 1).

Existe una infinidad de casos que se pueden abordar, sin embargo, en este prototipo sólo se consideraron los casos descritos anteriormente, ya que es importante reconocer que nuestro idioma es muy extenso y por lo tanto su análisis muy complejo. Como resultado del estudio realizado de las frases consideradas como ame-

Tabla 1. Casos y frases fraudulentas utilizadas en los contenidos de los mensajes de correo electrónicos.

Caso	Frases fraudulentas
Solicitud de información personal	¿Cómo se llama la calle dónde vives? Te mandaré un pequeño presente, a dónde te lo envío? Creo que me quitaron mi cuenta del Facebook, préstame la tuya. Te has ganado un premio ingresa tu número de cuenta.
Solicitud de información bancaria	Has sido uno de los 5 seleccionados del Sorteo Tec 2012, mándanos tu número de cuenta para que puedas disfrutar tu regalo. Necesitamos tu NIP para saber si ganaste la rifa.
Avisos falsos sobre premios y ofertas	Usted ha sido especialmente seleccionado para escuchar esta oferta. Si compra nuestro producto, obtendrá un maravilloso premio o bono gratis. Usted ha ganado uno de cinco valioso premios. Usted ha ganado una importante suma de dinero en un sorteo de lotería extranjera.

Tabla 2. Palabras claves.

Adorable	Crédito	Fabuloso	Mandar
Agradable	Creer	Facebook	Metroflog
Alta	Crei	Fantástico	Mirar
Ardiente	Cuenta	Feo	Mostrar
Asombroso	Débito	Foto	Muestro
Banamex	Decir	Fraude	MySpace
Banco	Depositar	Ganar	Necesitar
Banearan	Desagradable	Gracioso	NIP
Banorte	Dice	Gratis	Nómina
Bello	Dinero	Grotesco	Número
Bonito	Elevado	Hermoso	Oferta
Buena	Eliminar	Horrible	Pagar
Cargamos	Encantador	Increíble	Pensar
Caro	Enlace	Ingresar	Perfil
Checar	Enviar	Lindo	Piensa
Comprar	Error	Lotería	Premio
Correcto	Especial	Lujoso	Santander
Costoso	Etiquetar	Mal	Scotiabank
Twitter	Suscripción	Suba	Seleccionar
Valioso	Taggear	Sube	Sorprendente
Ver	Tarjeta	Subir	Sorteo
Vivir	Tuenti	Súper	

nazas de fraude, se hizo una selección de palabras que proporcionan información significativa en la interpretación del contenido del mensaje, con el fin de detectar patrones que permitan clasificar los mensajes considerados como amenazas o riesgos para los usuarios (Tabla 2).

Debido a que una oración se compone de sustantivos, verbos y adjetivos, las palabras fueron clasificadas siguiendo dicho esquema, para después obtener su núcleo y cambiar cada letra por un símbolo, de esta manera, se evitan problemas de repetición de palabras. En la Tabla 3 se enlistan los verbos, sustantivos y adjetivos seleccionados. La lista puede variar dependiendo del tiempo de conjugación presentado en las frases.

Además se agregaron las palabras con mala ortografía, ya que es común que este problema se presente en la mensajería instantánea.

4.2. Tokenizador

Este módulo es el encargado de separar las palabras, para esto, el tokenizador selecciona cada palabra para codificarla utilizando una simbología previamente establecida (Tabla 4) y la compara con las contenidas en la base de datos. Por medio de consultas determina si la palabra pertenece al conjunto de verbos, sustantivos o adjetivos. Después de esto, las palabras son ordenadas para después ser evaluadas con reglas tipo Prolog, y así analizar los mensajes para determinar la peligrosidad del mismo. Se debe de tener en cuenta el orden en el cambio de los caracteres, por ejemplo: en la palabra **quiere**, su codificación debe ser la siguiente **6e-e**, pero al tomar el orden de la tabla sería de la forma **qu4e-e**, lo que daría una codificación incorrecta y ocuparía más espacio del necesario.

La búsqueda de las coincidencias con las tablas de la base de datos es considerando la inserción de las palabras codificadas pero en su raíz y no como palabra completa, esto es para evitar la existencia de los mismos datos en donde sólo cambian unos pocos caracteres. Por ejemplo, en la tabla 5, se muestran las palabras que contendrá la base de datos (como son originales, en su raíz léxica y codificada a la simbología propuesta) para algunos de los adjetivos.

En el siguiente algoritmo describe la acción a realizar por el tokenizador.

1. Función token(frase)
2. //Comentario: Función que separa las palabras y las busca en el diccionario.//
3. Obtiene la primera palabra;
4. Hacer
5. Convertir a minúsculas y omitir acentos de la
6. palabra;
7. Codificar la palabra a la simbología propuesta;
8. Buscar la palabra codificada en el diccionario;
9. Si encuentra palabra entonces
10. Aumenta contador;
11. Fin si
12. Mientras (obtiene la siguiente palabra)

Tabla 3. Palabras claves clasificadas.

Verbo	Sustantivo	Adjetivo
Banearan	Banamex	Adorable
Cargamos	Banorte	Agradable
Checar	Crédito	Alta
Comprar	Cuenta	Ardiente
Creer	Débito	Asombroso
Creí	Dinero	Bello
Decir	Facebook	Bonito
Depositar	Foto	Buena
Dice	Lotería	Caro
Enviar	Metroflog	Correcto
Etiquetar	MySpace	Costoso
Eliminar	NIP	Desagradable
Ganar	Nómina	Elevado
Ingresar	Suscripción	Encantador
Mandar	Premio	Especial
Mirar	Santander	Fabuloso
Mostrar	Scotiabank	Fantástico
Muestro	Sorteo	Feo
Piensa	Tarjeta	Gracioso
Seleccionar	Tuenti	Gratis
Subir	Twitter	Grotesco
Sube	Banco	Hermoso
Suba	Fraude	Horrible
Taggear	Número	Increíble
Ver	Enlace	Lindo
Error	Oferta	Lujoso

Tabla 4. Simbología establecida.

Sustituto	Letras o Sílabas
0	b, v
1	ce, ci, si, se, zi, ze
2	d, de
3	g, j, gu
4	i, y, ll
5	ca, co, cu, ka, ko, ku
6	qui, qi, ki, que, qe, ke, q, k
7	t, te
8	u, w
9	x, cs, xh, cc, sh, zh, ch
*	sa, so, su, za, zo, zu
^	s, es, z, ez
+	pe, p
?	N
-	r, rr
Se omite	H

Tabla 5. Tabla de adjetivos

Idadjetivo	palabra	original
0	a2o-a0le	Adorable
1	a3-a2a0le	Agradable
2	al7a	Alta
3	a-24en7	Ardiente
4	a*m0-o	Asombroso
5	0e4	Bello
6	0on47	Bonito
7	08en	Buena
8	5-	Caro
9	5-ec7	Correcto
10	5^7o	Costoso

4.3. Programación del Conocimiento

Para realizar la programación del conocimiento se utilizó *ProLog*, ya que por su forma de abstracción de objetos, permite evaluar diferentes premisas para tener una solución del problema rápido y exacto. Además se utiliza la librería *MOD-LOGIC*. Dicha librería es la que permite analizar los patrones para saber si en realidad el correo es amenazante. A continuación se muestra un ejemplo de las reglas ProLog en el sistema utilizando la librería MOD-LOGIC.

1. SEMÁNTICA DE PREDICADOS:
2. palabra(Word, número).
3. REGLAS:
4.
5. DATOS
6. palabra(ado-a0le,0).
7. palabra(al7a,2).
8. palabra(c-e247,100).
9. palabra(24ne-,101).
10. palabra(9e5,202).
11. palabra(2+o17,207).

Como se muestra en el código anterior, la búsqueda de las palabras se realiza a base de una plantilla, la cual cuenta con tres secciones:

- **Semántica de Predicados**: En esta sección se definen los significados de los argumentos de cada predicado incluido.
- **Reglas**: En este apartado se define un conjunto de reglas lógicas para determinar la peligrosidad del correo.
- **Datos**: Permite registrar datos relacionados con los hechos que forman parte de la base de conocimiento. Estos datos están conformados por dos argumentos: el primero es la palabra “neutra”, y el segundo es el código de identificación dentro del sistema.

Al finalizar el análisis, se puede identificar si el correo electrónico en realidad es dañino. Esto es para identificar al remitente y tener alertado al usuario sobre lo ocurrido. Cabe mencionar que en esta versión del sistema, la sección de reglas no está descrita en la plantilla para MOD-LOGIC, debido a que la librería no proporciona todavía las estructuras condicionales, las cuales equivaldrían a usar "if-then". Por lo tanto, para el desarrollo de esta versión, la sección de reglas fueron programadas en C++. Por ejemplo, el análisis de la palabra **costosísimo**, primero se debe de tener codificada de la siguiente forma: **5ˆ7o11mo**. Una vez encontrado el tipo de la palabra, se busca en su estado neutral, que sería **5ˆ7o***, se obtendría el código de identificación, el cual será **10**. Una vez con el código de identificación, se analizarán en las reglas donde se harán las conjugaciones para determinar si en realidad el uso de esa palabra es amenazante.

4.4. Manejador de Pizarra

El manejador de pizarra es el módulo encargado de llevar un registro de los eventos generados en el sistema con el fin de:

- Tener un seguimiento de los mensajes que se consideren peligrosos
- Registrar a los usuarios que envían mensajes considerados peligrosos

5. Conclusiones

Una de las razones principales por las que consideramos muy importante nuestra propuesta fue porque aun existiendo muchos tipos de aplicaciones para la detección de malware, *bots* y *phishing*, éstas no son por completo seguras ya que sólo registra las actividades de las que tiene conocimiento y no de patrones que tienen algunas aplicaciones dañinas para los sistemas o los usuarios. El sistema propuesto se enfoca a las amenazas llamadas *bots* y *phishing*, ya que son las amenazas que más impactan a los usuarios, al robar su identidad y generar fraudes sin que los usuarios se percaten de lo que pasa. Por tal motivo en este artículo se describe el modelo y desarrollo de un prototipo que analiza los correos que reciben los usuarios, con el objetivo de encontrar la verdadera intención del mensaje recibido en un correo electrónico. Dicho prototipo constituye una propuesta de solución para la seguridad en los correos electrónicos, en el ambiente Windows-NT, con la herramienta de Microsoft Outlook 2010.

Referencias

[1] Panda Security, 2012, Phishing, 20 de Enero 2012, http://www.pandasecurity.com/spain/homeusers/security-info/cybercrime/phishing/.

[2] ALEGSA, 2011, Definición de Bot - ¿Qué es un Bot?, 20 de Enero 2012, http://www.alegsa.com.ar/Dic/bot.php

[3] Phishing.org, 2012,URL: http://www.phishing.org/history-of-phishing/.

[4] Asociación de Internautas, 2011, Guía rápida para reconocer páginas web falsas que simulan entidades bancarias, Disponible: http://seguridad.internautas.org/html/863.html. Último acceso: 06 Septiembre 2011.

[5] Phishing activity trends report, 2005, Phishing activity trends report Anti-phishing working group, technical report. Disponible en: http://www.Antiphishing.org/reports/apwgreportjan2006.pdf

[6] J. Rodriguez, 2012, Tipos y funcionamiento de bots maliciosos usados por los ciberdelincuentes, 21 Mayo 2012., http://www.xatakaon.com/seguridad-en-redes/\tipos-y-funcionamiento-de-bots-maliciosos-\usados-por-los-ciberdelincuentes.

[7] J. Gómez-Sanz, 2002, Metodología de Desarrollo de Sistemas Multiagente (Ph.D. Tesis). Universidad Complutense de Madrid.

[8] C. Lucero, 2013, Modelo de seguridad multi-agente para la detección de fraudes en correos electrónicos, Tesis de licenciatura, Universidad de Sonora.

[9] OpenSSH. 2004, OpenSSH, 20 de Enero 2012, http://www.openssh.com/es/index.html

[10] OpenSSL, 2009, OpenSSL: The Open Source toolkit for SSL/TLS, 20 de Enero 2012, http://www.openssl.org