Artículo de Investigación

Citación: Carvajal, Carlos. y Bayona, Diego. "Extensión de taxonomía y tratamiento de valores faltantes sobre un repositorio de incidentes de seguridad informática". En: *Ingeniería, Vol. 18, Num. 1, págs. 24-49*

Extensión de taxonomía y tratamiento de valores faltantes sobre un repositorio de incidentes de seguridad informática

Taxonomy extension and missing-values treatment over an informatics-security incident repository

Carlos Javier Carvajal Montealegre
Ingeniero de Sistemas
Universidad Distrital
Francisco José de Caldas
Ing.carlosj@gmail.com

Diego Nicolás Bayona
Ingeniero de Sistemas
Universidad Distrital
Francisco José de Caldas
Nicolas.bayona@gmail.com

Zulima Ortiz Bayona
Msc. en Matemáticas
Grupo de Investigación Arquisoft, rama de Seguridad Informática,
Universidad Distrital
Francisco José de Caldas
zortiz@udistrital.edu.co

Resumen

En este artículo se detalla el proceso de estimación de datos faltantes mediante el teorema de Bayes, sobre un repositorio de incidentes de seguridad informática compuesto por datos de tipo categórico. Así mismo, se hace uso de una taxonomía, ampliada y redefinida para acoplarse a los incidentes encontrados.

Palabras clave: datos categóricos, incidente de seguridad, minería de datos, pre-procesamiento de datos, repositorio, taxonomía.

Abstract

This paper describes the missing-values estimation process through the Bayes theorem acting over an information security incident repository composed by categorical data. Additionally, an augmented taxonomy is defined to account for the identified incidents.

Key words: categorical data, data mining, data pre-processing, security incident, repository, taxonomy.

1. Introducción

Un incidente de seguridad se define como la agrupación de una o más acciones llevadas a cabo por un atacante para lograr un resultado no autorizado en un sistema informático. Pueden ser distinguidas de otro grupo de acciones por las características de quien ataca, como ataca, qué objetivos tiene, quien fue el blanco del ataque y cuando se ejecutaron dichas acciones.

Fecha recibido: ene. 18/2013
Fecha modificado: may. 27/2013
Fecha aceptado: may. 31/2013

Este tipo de incidentes son reportados y dichos reportes son comúnmente accesibles desde Internet; en algunas ocasiones los reportes mantienen una estructura diferenciando algunas características previamente definidas, en otros casos son descripciones vagas y subjetivas. El conjunto de incidentes debe asegurar que sus elementos puedan ser descritos mediante las mismas características, para lo cual es necesario contar con un estándar descriptivo o taxonomía. Lo anterior con el fin de permitir análisis de minería de datos.

La taxonomía propuesta por el instituto Sandia en el artículo "A Common Language for Computer Security Incidents" [1], permite identificar y caracterizar estos incidentes, los campos de la taxonomía se observan en la figura 1.

Durante la recopilación de incidentes de seguridad realizada como parte de la investigación en el proyecto de grado "Análisis de Incidentes de Seguridad Informática Mediante Minería de Datos, para Modelado de Comportamiento y Reconocimiento de Patrones" [2] se observó la necesidad de ajustar la taxonomía antes descrita. De la misma forma, fue

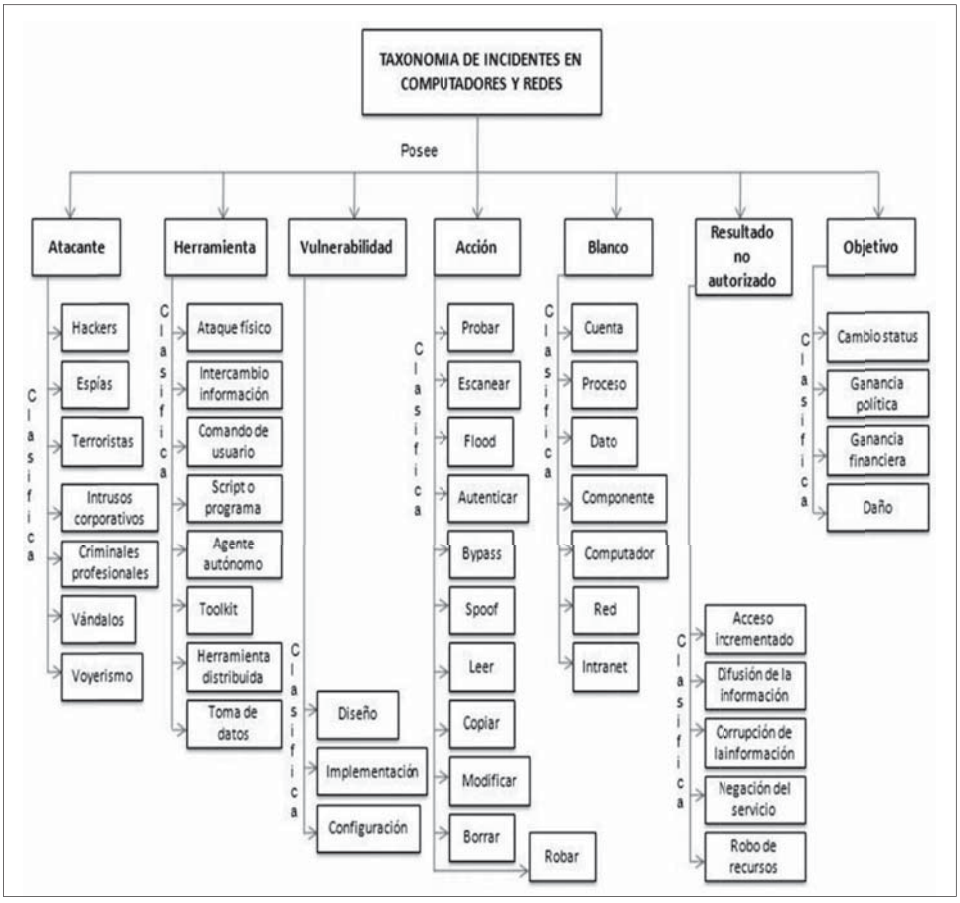

Figura 1. Taxonomía de incidentes de redes y computadores.

necesario considerar el problema de valores faltantes (missing values), en la construcción del repositorio, debido a que algunos valores de los incidentes no pueden ser capturados. Los resultados obtenidos en dicho proyecto se reportan en este artículo.

La organización del artículo es la siguiente: en la sección 1 se describe la obtención de los datos, para la Sección 2 se explica cómo y porque se extendió la taxonomía base, la Sección 3 muestra el procedimiento para completar los valores faltantes del repositorio, la sección 4 expone las pruebas y la sección 5 presenta los resultados obtenidos y las conclusiones.

2. Obtención de Datos

Se consultaron fuentes de reportes de incidentes, que fueran de acceso libre en Internet; es normal que las empresas que han sido blancos de ataques informáticos deseen mantener cierto nivel de reserva sobre las características del ataque, muchas veces mantienen en secreto la ocurrencia del incidente. Por otra parte los incidentes reportados en un sitio Web suelen ser los mismos citados en otros sitios, estos factores minimizan el número de ataques que pueden reunirse, no obstante se identificaron tres fuentes de fácil consulta y que proporcionaban gran número de datos sobre los incidentes:

Web Hacking Incident DataBase: Es un proyecto del Consorcio de Seguridad de Aplicaciones Web (Web Application Security Consortium) [3], dedicado a mantener una lista de incidentes de seguridad informática en aplicaciones Web. En esta se incluyen solo ataques dirigidos, aquellos que violaron la seguridad de las aplicaciones web. El sitio cuenta con filtros de búsqueda capaces de consultar los incidentes registrados, como: método de ataque, debilidad de la aplicación Web, resultado y entidad atacada. De esta fuente se adquirieron 569 incidentes con fecha de ocurrencia fuera mayor al 2007. Las características de los incidentes de esta base de datos se resumen en la Tabla I.

Chronology of Data Breaches: *Open Security Foundation,* es fuente de información acerca de violaciones de seguridad. La base de datos incluye datos como: el tipo de violación de seguridad, el tipo de entidad atacada, la ubicación geográfica de la entidad y la descripción del ataque [4]. Se tomaron 579 registros ocurridos entre el 2005 y el 2010, teniendo como condición que la violación de seguridad fuera HACK (Hacking o Malware) o DISC (Divulgación no intencional). Las características de los incidentes de esta base de datos se resumen en la Tabla II.

COL-CSIRT: Es un proyecto de investigación de la Universidad Distrital [5] con el fin de crear un equipo de respuesta a incidentes de seguridad informática. El grupo de investigación cuenta con una base de datos del centro de respuesta a incidentes y ataques con código malicioso; esta información proviene de las siguientes fuentes: Centro de estudios de respuesta y tratamiento de incidentes de seguridad de Brasil [6], y Centro de respuesta a incidentes de seguridad de Inteco [7]. Se obtuvieron 169 incidentes desde febrero de 2008 a octubre de 2010.

Tabla I. WEB HACKING INCIDENT DATABASE

Campo	Descripción
WHID-ID	Hace referencia al consecutivo registrado para cada incidente de seguridad informática almacenado en el repositorio. El formato definido para el campo es el siguiente WHID AAAA-Consecutivo.
Entry Title	Corresponde al Título con el que fue registrado el incidente de seguridad informática en la base de datos: WHID AAAA-Consecutivo: Sumario del incidente.
Incident Description	Corresponde a la descripción del incidente de seguridad informática, donde registra los detalles de la incidencia.
Reference	Enlaces a sitios web donde se encuentra almacenada información referente al incidente de seguridad informática en mención.
Date Ocurred	Registra la fecha en la que ocurrió el incidente de seguridad informática. El formato de la fecha es el siguiente: MM-DD-AA.
Attack Method	Corresponde al método utilizado para realizar el ataque.
Application Weakness	Referencia a la debilidad vulnerada en la aplicación que permitió el éxito del ataque.
Outcome	Registra la finalidad del ataque realizado, es decir, lo que el ataque obtuvo como resultado para el atacante.
Attacked Entity Field	Registra el tipo de organización de la entidad atacada.
Attacked Entity Geography	Corresponde al sitio geográfico donde sucedió el incidente. El campo contiene el nombre de la ciudad atacada o en su defecto el nombre del país atacado.
Mass Attack	Indica si el ataque fue masivo. El campo registra la palabra Yes o No, informando si el ataque descrito en el Incidente de Seguridad Informática fue de tipo masivo.
Mass Attack Name	Registra el nombre del ataque masivo realizado.
Number of Site Affected	Número de sitios afectados en los ataques realizados en el incidente de seguridad. El formato del campo es un número entero que informa la cantidad de sitios atacados.
Attack Source Geography	Referencia al sitio geográfico de donde se realizó el ataque. El formato del campo es la dirección URL de la ubicación en el sitio web Google Maps.
Attacked System Technology	Indica el tipo de tecnología atacada.
Cost	Indica el costo monetario para la(s) entidad(es) atacada(s).
Items Leaked	Tipo de información robada en el ataque descrito en el Incidente de Seguridad informática.
Numbers of records	Indica el número de ataques realizados en el incidente registrado. El formato para el campo indica el número de ataques realizados en el incidente de seguridad informática. Para los casos que no se conoce el número de ataques, registra el valor “UNKNOWN”.
Additional link	Almacena enlaces adicionales donde se registra información del incidente. El campo registra las direcciones URL adicionales que brindan información del incidente.

Tabla II. CHRONOLOGY OF DATA BREACHES	
Campo	**Descripción**
Date	Fecha de suceso del incidente de seguridad informática. El formato del campo corresponde al siguiente tipo de información "Mes Día de Año".
Name Entity	Nombre de la entidad atacada en el incidente.
Type Entity	Tipo de organización de la entidad víctima del incidente (Institución Educativa, Financiera, etc).
Type of Breach	Tipo de violación de seguridad el cual fue ejecutado en el incidente de seguridad informática (Hack,Disc).
Total Records	Cantidad de ataques realizados para el incidente en cuestión.
Description	Descripción de los detalles de la ocurrencia del incidente.
Geography Ubication	Ubicación geográfica donde el incidente de seguridad fue efectuado. El valor del campo indica la ciudad del suceso.

Los datos de estas tres fuentes se estandarizaron según la taxonomía Sandia mencionada anteriormente. Para ello se recorrieron de manera manual y se llenaron los campos definidos en la taxonomía.

3. Extensión de la Taxonomía

Luego de leer, analizar y clasificar los incidentes, se encontró dificultad en completar campos como la Herramienta, el Objetivo, el Atacante y el Resultado no autorizado, pues esta información se extrajo de la descripción proporcionada, que en más del 50% de los casos no era lo suficientemente explicativa. Para esos incidentes fue necesario una consulta adicional en Internet, en distintos artículos para tener mayor conocimiento sobre los ataques y completar los datos de clasificación [8][9][10][11][12][13].

Por otra parte cabe resaltar que los campos extras de la WHID almacenan información de gran interés para identificar los incidentes, entre los que se cuentan:

Debilidad de la Aplicación: Especifica el fallo que permitió la ocurrencia del incidente, con los siguientes posibles valores: Autenticación Insuficiente, Autorización insuficiente, Entropía insuficiente, Falta de Configuración del Antivirus, Falta de configuración de Infraestructura, Falta de configuración en el Servidor Web, Falta de configuración en la base de datos, Falta de configuración de la aplicación, Funcionalidad vulnerable, Indexado Inseguro, Insuficiente anti-automatización, Insuficiente expira-

ción de sesión, Insuficiente recuperación de contraseña, Manejo inapropiado de entradas, Manejo inapropiado de salidas, Permisos inapropiados sobre archivos del sistema, Protección insuficiente en la capa de transporte y Proceso de validación insuficiente.

Método de Ataque: Proporciona información detallada acerca de cómo se vulneró la seguridad y se obtuvo el acceso no autorizado, con los siguientes posibles valores: Abuso de funcionalidad, Automatización de procesos, Acceso no autorizado, ARP *Spoofing*, *Bot*, *Clickjacking*, Comando del sistema operativo, Cross Site Request Forgery (CSRF), Cross Site Scripting (XSS), Denegación de servicio, Desconocido, Divulgación no intencional, Fuerza bruta, Gusano, Hijacking DNS, Inclusión de archivo local, Inclusión de archivo remoto, Localización predecible de recursos, Navegación forzada, *Phishing*, Secuestro de dominio, Secuestro de sesión, *Spyware*, Software de rastreo, SQL *Injection*, Suplantación de contenido, Troyano y Virus.

Estos dos campos, por los valores que pueden tomar, dan un panorama más exacto del incidente y por tal motivo fueron incluidos en la taxonomía, generando una primer mejora.

A medida que se fueron analizando más incidentes se hicieron latentes nuevas necesidades, por una parte muchos de los valores de la taxonomía base (inclusive de los dos nuevos campos tomados de WHID) no abarcaban características esenciales de los incidentes y por otra parte, debido a la información que se suele reportar y encontrar sobre los incidentes, el campo Acción no podría ser diligenciado. Por las razones anteriores, en primer lugar se eliminó el campo Acción de la taxonomía y en segundo lugar se crearon nuevos valores para la clasificación, en las figuras 2 y 3 se muestra la taxonomía final extendida. En la Tabla III se resumen los nuevos valores creados para la taxonomía con su justificación. Estos a su vez, permitieron clasificar finalmente todos los incidentes encontrados cubriendo un rango más amplio que el de la clasificación existente, además de proporcionar mayor conocimiento del vector de ataque.

Tabla III. NUEVOS VALORES PARA TAXONOMÍA

Valores	Campo	Justificación
Usuario Interno	Atacante	Incidentes causados por errores de empleados con acceso autorizado al sistema tal como el envío de correos a destinatarios no autorizados, negligencia al evitar el uso de contraseñas o instalación de software para fines de uso personal en computadores que albergan información privada de la entidad.
Políticas de Seguridad	Vulnerabilidad	
Sin Intencionalidad	Objetivo	
Usuario Externo	Atacante	Un usuario ajeno a la entidad, pero con un cierto nivel de acceso descubre por accidente errores o agujeros de seguridad que le permiten obtener privilegios más allá de los que tiene permitidos.

Bot	Método de Ataque	Incidentes donde el causante de la vulneración es un programa malicioso que busca infectar máquinas y convertirlas en miembros de redes zombie.
Spyware	Método de Ataque	Incidentes donde el software malicioso responsable de la vulneración está diseñado para recopilar anónimamente la información del equipo infectado y reenviar esta información a destinatarios específicos.
Software de Rastreo	Método de Ataque	Incidentes donde el ataque es realizado a través de software tipo escáner de Redes LAN; donde el programa analiza los puertos NetBios, las direcciones IP, las direcciones MAC, y puertos lógicos TCP/IP/UDP. El análisis es con el fin de encontrar vulnerabilidades que puedan será aprovechadas por el atacante.
No Aplica	Herramienta	Indica que el incidente fue ocasionado por: un error, una falta de configuración, o la incorrecta validación de las vulnerabilidades de seguridad en los distintos componentes de Software o Hardware del sistema afectado.
No Aplica	Método de Ataque	Incidentes donde un usuario no tuvo que realizar ninguna acción especial fuera de las habituales para interactuar con el sistema, son ocasionados por que el sistema o componente es factible de atacar, debido a alguna vulnerabilidad en la configuración o en el incumplimiento de las políticas de seguridad informática.
No Aplica	Debilidad de la Aplicación	No existe debilidad en la aplicación, sujeto totalmente a incidentes donde existe un error humano. Publicaciones de información confidencial son el ejemplo más claro.
Falta de Configuración del Antivirus	Debilidad de la Aplicación	Incidentes donde un software de antivirus desactualizado permitió que un programa malicioso pudiera vulnerar la seguridad.
Falta de Configuración de la Infraestructura	Debilidad de la Aplicación	Se pasan por alto ciertas directivas sobre configuraciones de red, como evitar que las peticiones de internet lleguen a ciertas maquinas o tener un ancho de banda insuficiente para un servicio de alto flujo de datos.
Falta de configuración en el servidor WEB	Debilidad de la Aplicación	Por negligencia, los servidores Web se dejan con una configuración por omisión para contraseñas, estructura de directorios y nombres de carpetas entre otros, permitiéndole al atacante conocer exactamente donde están los archivos que desee acceder y como acceder a estos.
Falta de Configuración en Base de Datos	Debilidad de la Aplicación	Aunque las aplicaciones que accedan a información almacenada en bases de datos estén correctamente configuradas y sigan las directivas de seguridad, la base de datos carece de una contraseña por lo que la información puede ser accedida por otros medios diferentes a las aplicaciones determinadas para este fin.
Funcionalidad Vulnerable	Debilidad de la Aplicación	Un usuario descubre como evadir los controles de una aplicación mediante medios permitidos por la misma aplicación. [17].

Figura 2. Taxonomía extendida de incidentes seguridad informática. Parte 1

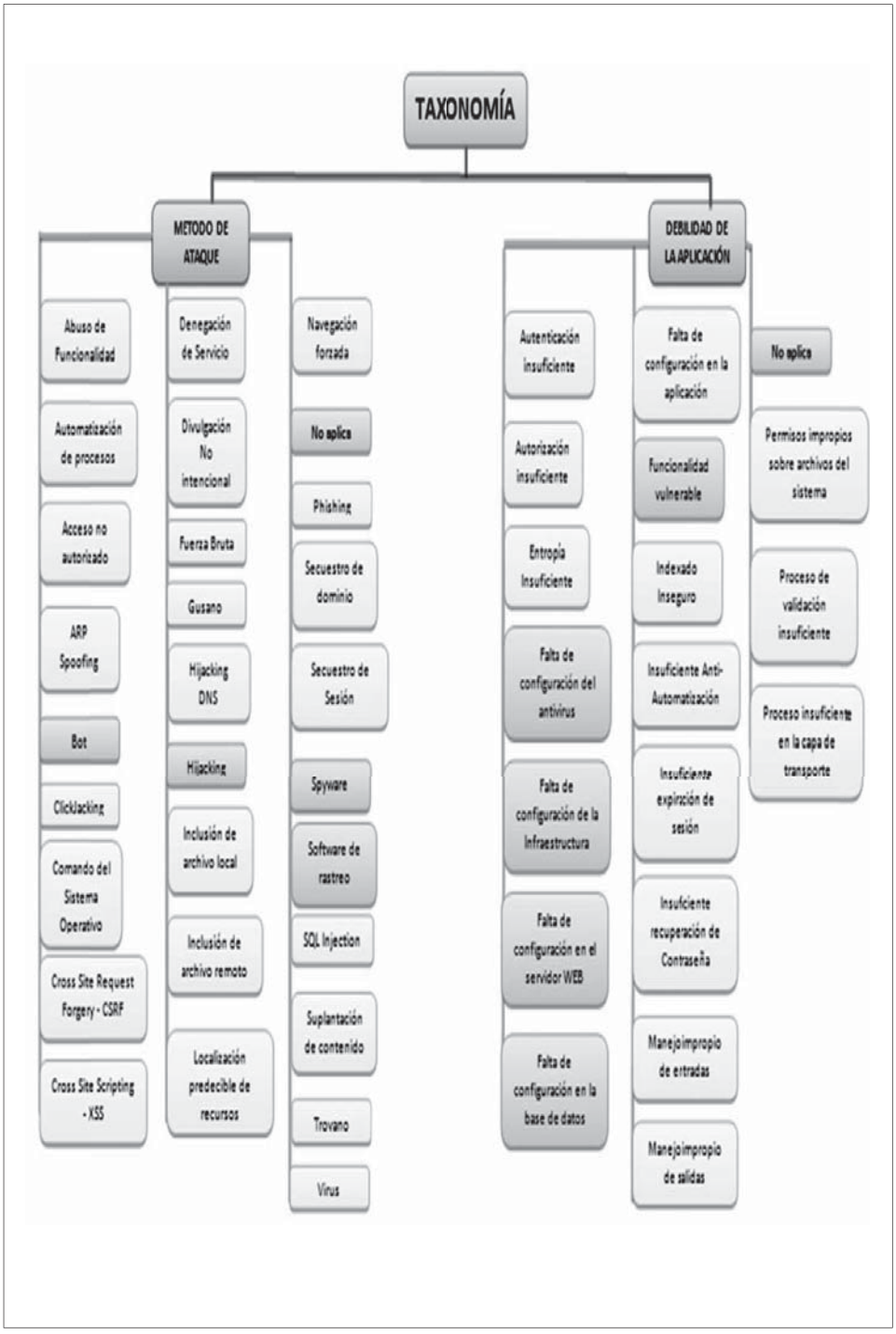

Figura 3. Taxonomía extendida de incidentes seguridad informática. Parte 2

4. Tratamiento de valores faltantes

Una vez finalizado el proceso de estandarización de datos, el número total de registros recopilados y descritos de acuerdo a la taxonomía mejorada fue de 1313 incidentes de seguridad, no obstante, debido a la precaria información relacionada para el 25% de los incidentes, no fue posible completar todos los campos descriptivos, este problema es más conocido en la literatura como *missing-values*, valores faltantes o en blanco. De los 1313 registros, 982 tuvieron todos los campos completos, mientras que 331 presentaron uno o más valores faltantes, los casos de mayor impacto se presentaron en los campos Método de Ataque 145 registros, Debilidad de la Aplicación 123 registros, Vulnerabilidad 144 registros y Herramienta 262 registros con valores faltantes. En las figuras 4 y 5 se puede observar la distribución de valores para los campos Método de Ataque y Debilidad de la Aplicación y se visualiza el alto número de registros con estos campos en blanco.

Tal como lo sugieren autores consultados [14][15][16], las opciones para tratamiento de *missing-values* son:

- Ignorar el registro cuando tenga gran cantidad de atributos faltantes.
- Llenar el valor faltante de manera manual.
- Usar una constante global para llenar el valor faltante, por ejemplo el texto "desconocido".
- Usar el valor promedio que muestra el atributo en los registros.
- Usar el valor más probable determinado mediante herramientas basadas en inferencia, tales como el formalismo Bayesiano y árboles de decisión.
- Métodos de aprendizaje supervisado y no supervisado tales como el clustering, o las redes neuronales.
- Estas opciones se analizaron de manera individual:
- Ignorar los registros con valores faltantes conllevaría a perder el 25% de los incidentes recopilados.
- Llenar el valor faltante manualmente no es una opción pues el repositorio se construyó de esta manera.

Definir un valor por omisión de llenado causaría que estos algoritmos consideraran dicha constante como un valor de "interés"; calcular un valor promedio es una buena opción para repositorios cuyos campos sean numéricos, sin embargo en este caso son categóricos con lo cual no es posible hallar un promedio [17] [18]. Los métodos de aprendizaje supervisado y no supervisado serán usados a futuro en tareas propias de la minería de datos [19], tales como la detección de *outliers* y el análisis de cluster en el momento en que el repositorio este estandarizado totalmente. Dado este análisis, la estimación del valor más probable se escogió como la opción más conveniente para el caso de estudio. La técnica utilizada se define con detalle a continuación.

4.1 Estimación de valores faltantes mediante un enfoque Bayesiano.

Se utilizó una técnica con base en la propuesta de estimación de valores categóricos faltantes de Xiao-Bai Li [20], donde los valores faltantes de tipo categóricos se estiman como un problema de clasificación mediante cálculos probabilísticos, es decir se calcula el valor faltante del atributo de interés, como la clase (categoría) más probable dados los valores para los restantes atributos presentes. Esta probabilidad se obtiene mediante una aproximación Bayesiana:

Sea $c_1,...,c_L$ particiones del espacio muestral. Entonces, para cada evento X en el espacio muestral,

$$P(c_k \mid X) = \frac{P(c_k)P(X \mid c_k)}{\sum_{r=1}^{L} P(c_r)P(X \mid c_r)}, k = 1..., L, \quad (1)$$

Donde $P(c_k)$ es llamada la probabilidad a priori y $P(c_k \mid X)$ es la probabilidad a posteriori.

Se considera un conjunto de datos con un atributo de clase de dos tipos posibles, c_1 y c_2, y M-1 atributos que no son de clase, $X_1,...,X_{M-1}$. Para un nuevo registro $x=(x_1,..., x_{M-1})$ que deba ser clasificado; el clasificador Bayesiano asigna al valor de clase a c_1 si $P(c_1 \mid x) > P(c_2 \mid x)$, de lo contrario asigna a c_2. La probabilidad a posteriori $P(c_k \mid x)$ puede ser derivada del Teorema de Bayes (1). El proceso involucra la estimación de $P(c_k)$ y $P(x \mid c_k)$ a partir de los datos. Mientras que $P(c_k)$ es fácil de obtener, evaluar a $P(x \mid c_k)$ es muy costoso en cuanto a recursos computacionales para conjuntos de datos con una alta dimensionalidad. Para sortear este inconveniente, se asume que los atributos son condicionalmente independientes entre sí. Bajo esta suposición, $P(x \mid c_k)$ puede ser fácilmente calculado por:

$$P(x|c_k)=\prod_{j=1}^{M-1} P(x_j|c_k) \quad (2)$$

Lo que se conoce como un clasificador ingenuo de Bayes (Naive Bayes classifier).

El método permite estimar las probabilidades de múltiples atributos faltantes con base en los atributos existentes. Dado un conjunto de datos con N registros y M atributos categóricos, $X_1,...,X_M$, sea L_i el número de categorías en X_i, y sea N_i el número de registros con X_i valores conocidos, y N_{ij} el número de registros donde X_i es igual a la k-esima categoría c_{ij}. Adicionalmente, sea $N_{jr} \mid i_k$ el número de registros donde X_j es igual a la r-ésima categoría c_{jr}, dado $Xi=c_{ik}$ con $j \neq i$. El proceso para completar los valores faltantes a partir de este método se describe en la Tabla IV.

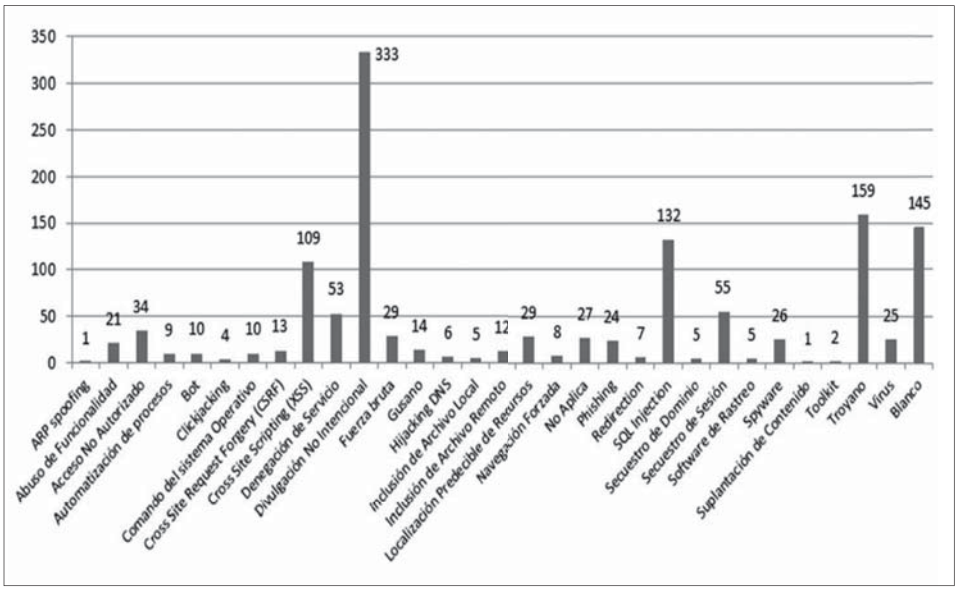

Figura 4. Distribución de valores Iniciales para Método de Ataque.

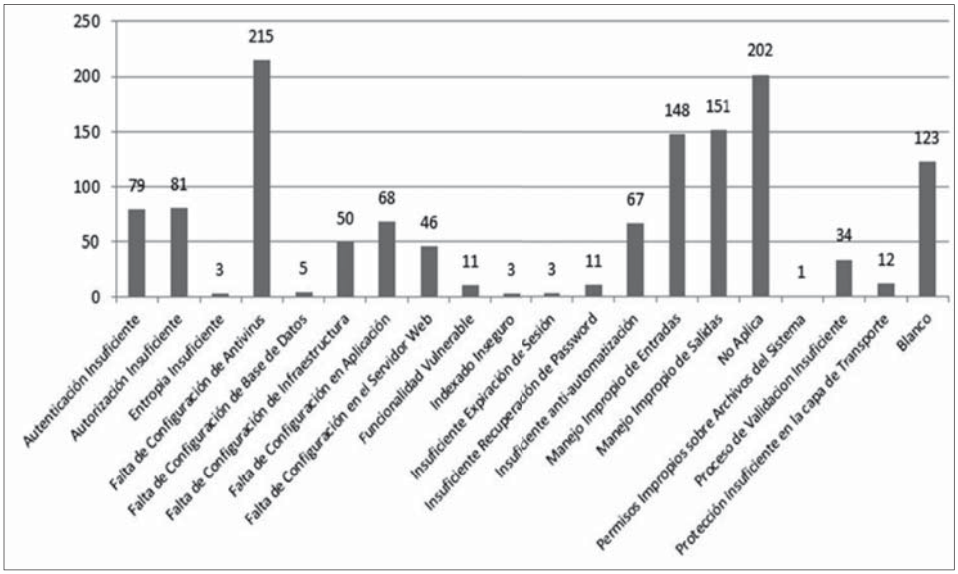

Figura 5. Distribución de valores Iniciales para Debilidad de la Aplicación.

4.2 Ejemplos de estimación

Consideremos una versión simplificado de nuestro repositorio, compuesto por 4 atributos categóricos: ATACANTE, HERRAMIENTA, VULNERABILIDAD y OBJETIVO.

Tabla IV. ESTIMACION DE VALORES FALTANTES

Procedimiento

1. Calcular las probabilidades a priori para cada atributo:

$$P(X=C_{ik}) = N_{ik}/N_i,\ i = 1,\dots, m;\ k = 1,\dots, L_i.$$

2. Calcular las probabilidades condicionales de Xj, dado Xi=cik:

$$P(X_j=C_{jr}|X_i = C_{ik}) = N_{jr}|_{ik} / N_{ik},\ j = 1,\dots, M;\ j \neq i;\ r = 1,\dots, L_i.$$

3. Para un registro x con un valor faltante en X_l, sea J el índice para el conjunto de atributos sin valores faltantes en x, x_j l parte correspondiente de x. Calcular las probabilidades a posteriori, según las ecuaciones (1) y (2):

$$P(X_i = c_{ik}|x_J) = \frac{1}{P(x_J)} P(X_i = c_{ik}) \prod_{j \in J} P(X_j = c_{jr}|X_i = c_{ik}), k = 1, \dots, L_i$$

4. Reemplazar el valor faltante Xi de x basándose en las probabilidades calculadas en el paso 3 usando el valor con la máxima probabilidad a posteriori. (MaxPost).

Suponemos que el repositorio se compone de 20 registros, donde cuatro de estos registros presentan valores faltantes como se observa en la Tabla V, para los registros 17 y 18 se observa que el campo faltante es VULNERABILIDAD, en este caso iniciamos calculado las probabilidades a priori para el mismo:

P(Diseño) = 4/18

P(Configuración) = 7/18

P(Políticas de Seguridad) = 7/18

A continuación se calculan las probabilidades condicionales para el atributo ATACANTE cuando sea igual a Hackers dado cierto valor de VULNERABILIDAD:

P(Hackers | Diseño) = 4/4

P(Hackers | Configuración) = 4/7

P(Hackers | Políticas de Seguridad) = 2/7

De la misma forma calculamos las probabilidades condicionales para {Herramienta = Agente Autónomo} y {Objetivo = Daño}

P(Agente Autónomo | Diseño) = 1/4

P(Agente Autónomo | Configuración) = 6/7

P(Agente Autónomo | Políticas de Seguridad) = 3/7

P(Daño| Diseño) = 4/4

P(Daño | Configuración) = 3/7

P(Daño | Políticas de Seguridad) = 2/7

Luego se calculan las probabilidades a posteriori:

$P(DI|H,AA,D) = 1/P\ (4/18)(4/4)(1/4)(4/4) = 0{,}05556/P$

$P(C|H,AA,D)=1/P\ (7/18)(4/7)\ (6/7)\ (3/7) = 0{,}08163/P$

$P(PS|H,\ AA,D) =1/P\ (7/18)\ (2/7)\ (3/7)\ (2/7) = 0{,}013605/P$

DI=Diseño, C= Configuración, PS= Políticas de Seguridad, H=Hackers, AA = Agente Autónomo D = Daño

Finalmente calculamos:

$P(DI|H,AA,D) = 0{,}05556P/0{,}05556P + 0{,}08163P + 0{,}013605P = 0{,}36844$

$P(C|H,AA,D) = 0{,}08163P/0{,}05556P + 0{,}08163P + 0{,}013605P = 0{,}54133$

$P(PS|H,AA,D) = 0{,}013605P\ /0{,}05556P + 0{,}08163P + 0{,}013605P = 0{,}09022$

Así que el valor más probable y el que reemplazara al valor faltante en los registros 17 y 18 es Configuración, este mismo cálculo se podría llevar a cabo para reemplazar los faltantes de los registros 19 y 20.

El mismo proceso se llevó a cabo con el repositorio real para completar los valores faltantes.

Tabla V. Repositorio Simplificado para Ejemplo de Estimación

No	Atacante	Herramienta	Vulnerabilidad	Objetivo
1	Hackers	Agente Autónomo	Diseño	Daño
2	Criminales Profesionales	Agente Autónomo	Configuración	Ganancia Financiera
3	Hackers	Agente Autónomo	Políticas de Seguridad	Daño
4	Hackers	Agente Autónomo	Configuración	Daño
5	Hackers	Agente Autónomo	Configuración	Cambio de Status
6	Hackers	Agente Autónomo	Configuración	Daño
7	Criminales Profesionales	Agente Autónomo	Configuración	Ganancia Financiera
8	Criminales Profesionales	Agente Autónomo	Políticas de Seguridad	Ganancia Financiera
9	Hackers	Comando de Usuario	Diseño	Daño
10	Hackers	Agente Autónomo	Configuración	Daño
11	Hackers	Agente Autónomo	Políticas de Seguridad	Daño
12	Hackers	Comando de Usuario	Diseño	Daño
13	Hackers	Comando de Usuario	Diseño	Daño
14	Usuario Interno	Intercambio de Información	Políticas de Seguridad	Sin Intencionalidad
15	Usuario Interno	Intercambio de Información	Políticas de Seguridad	Sin Intencionalidad
16	Usuario Interno	Intercambio de Información	Políticas de Seguridad	Sin Intencionalidad
17	Hackers	Agente Autónomo	?	Daño
18	Hackers	Agente Autónomo	?	Daño
19	?	?	Configuración	Ganancia Financiera
20	Usuario Interno	?	Políticas de Seguridad	?

4.3 Exclusión de Atributos con Bajo Contenido de Información

Para la estimación Bayesiana realizada, la dimensionalidad del incidente fue reducida, es decir se retiraron las siguientes columnas:

- **ENTIDAD**, 993 valores distintos.
- **FECHA**, 842 valores distintos.
- **PAIS**, el 70% de los incidentes tiene como país de origen a Estados Unidos, el 11% a Colombia, el 1,83% al Reino Unido, 1,29% en India, 1,14% en Australia y 12,34% en otros países.

El motivo de esta reducción se debe a que las columnas **ENTIDAD** y **FECHA**, para fines prácticos son identificadores del registro, pues en los incidentes recopilados pocos son reincidentes para una misma **ENTIDAD** y pocos ocurrieron en una misma fecha, ya que estas columnas manejan un alto rango de valores, no son aptas para agrupar por subconjunto.

En el otro extremo tenemos el campo **PAÍS**, ya que la mayoría de los incidentes ocurrieron en Estados Unidos, por lo tanto si se considerara en análisis Bayesiano de los valores faltantes sesgaría la decisión hacia el subconjunto definido para ese país. Por las razones anteriores los cálculos de probabilidad de las columnas mencionadas no se tomaron en cuenta para la estimación Bayesiana aunque siguen haciendo parte del repositorio.

5. Validación del método

Con el fin de comprobar la efectividad en la estimación de valores faltantes mediante el enfoque Bayesiano y probar el grado de confianza en el repositorio; se realizó una validación contra los registros del repositorio que tenían los datos completos. La distribución final del repositorio se muestra en la Tabla VI:

Tabla VI. Repositorio de datos

Tipo de Registro	Número de Registros	Porcentaje
Completo	982	74,79%
Con valores faltantes	331	25,21%

Tabla VII. Pruebas estimación valores faltantes

	Cantidad	Porcentaje
Aciertos	203	69%
Errores	91	31%

De la porción del repositorio que no tenía valores faltantes, se eligió un subconjunto equivalente al 30% (294 registros). Para cada registro del subconjunto de validación, se eliminó una columna elegida al azar y los valores eliminados fueron calculados con el enfoque Bayesiano descrito, luego se compararon los resultados obtenidos con el valor que tenía el campo inicialmente. Los resultados globales se presentan en la Tabla VII, los resultados discriminados por campo en la Tabla VIII y Figura 6.

Tabla VIII. Pruebas por campo de estimación de valores faltantes

Campo	Aciertos	Errores	Porcentaje Aciertos	Porcentaje de error
Tipo de organización	14	11	56,00%	44,00%
Método de ataque	28	15	65,12%	34,88%
Debilidad de la aplicación	23	11	67,65%	32,35%
Atacante	32	11	74,42%	25,58%
Herramienta	28	11	71,79%	28,21%
Vulnerabilidad	20	0	100%	0%
Blanco	20	9	68,97%	31,03%
Resultado no autorizado	21	13	61,76%	38,24%
Objetivo	21	6	77,78%	22,22%

Como se evidencia el porcentaje más bajo de aciertos corresponde al Tipo de Organización, demostrando que no existe una relación estrecha entre este campo y el resto del vector del ataque. Por el contrario la Vulnerabilidad depende totalmente del resto de campos del registro logrando un porcentaje de aciertos del 100%.

El clasificador Bayesiano muestra alta dependencia de la distribución de valores propia del conjunto de datos y obtiene valores acertados para registros que tengan un comportamiento común y conocido para el clasificador, esto se deduce de la inspección de

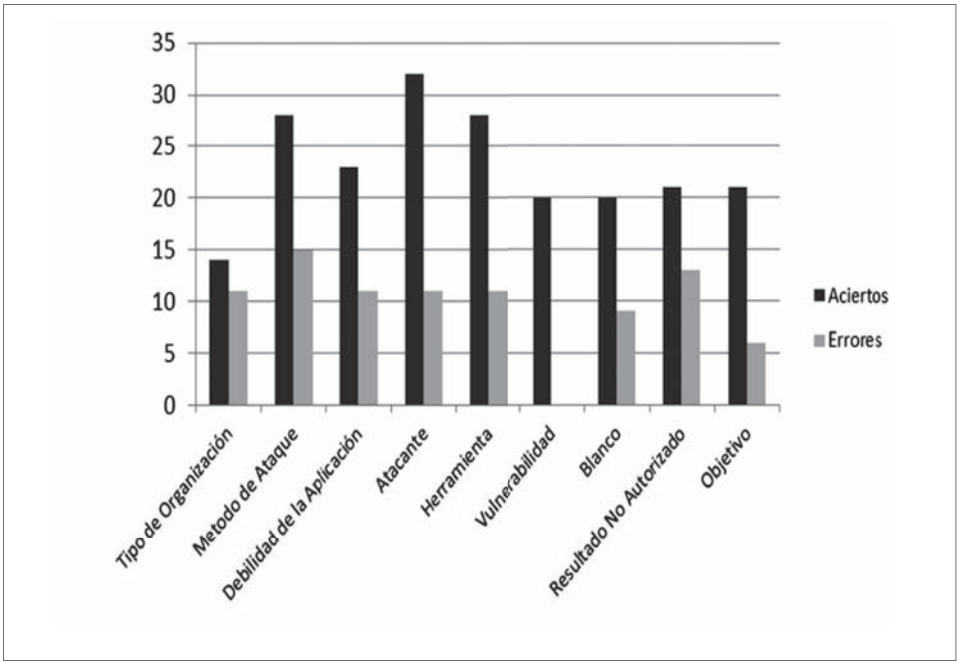

Figura 6. Cantidad de aciertos y errores mediante estimación Bayesiana

Tabla IX. Pruebas por campo de estimación de valores faltantes

Campo eliminado	Valor esperado	Valor obtenido	Motivo del error
Método de Ataque	Secuestro de Sesión	No Aplica	Único incidente en el conjunto de datos de Secuestro de Sesión teniendo como blanco a Computador, ya que no existían más incidentes de este tipo la probabilidad tiende a cero.
Atacante	Usuario Interno	Hackers	Incidente donde un Usuario Interno uso un Troyano, siendo esto un comportamiento poco común.

algunos de los registros que no fueron clasificados correctamente como los que se presentan en la Tabla IX.

5.1 Comparación contra el valor promedio

El clasificador fue enfrentado con uno de los métodos más comunes en el tratamiento de los valores faltantes el cual consiste en elegir el valor promedio del campo para completar el registro. Los resultados totales del método del valor promedio se muestran en la tabla X y los resultados discriminados por campo se presentan en la Tabla XI y Figura 7.

Tabla X. Pruebas estimación valores faltantes

	Cantidad	Porcentaje
Aciertos	101	34,35%
Errores	193	65,64%

Tabla XI. Pruebas por campo de estimación de valores faltantes

Campo	Aciertos	Errores	Porcentaje Aciertos	Porcentaje de error
Tipo de organización	9	16	36,00%	64,00%
Método de ataque	8	35	18,60%	81,40%
Debilidad de la aplicación	6	28	17,65%	82,35%
Atacante	14	29	32,56%	67,44%
Herramienta	10	29	25,64%	74,36%
Vulnerabilidad	3	17	15,00%	85,00%
Blanco	20	9	68,97%	31,03%
Resultado no autorizado	18	16	52,94%	47,06%
Objetivo	13	14	48,15%	51,85%

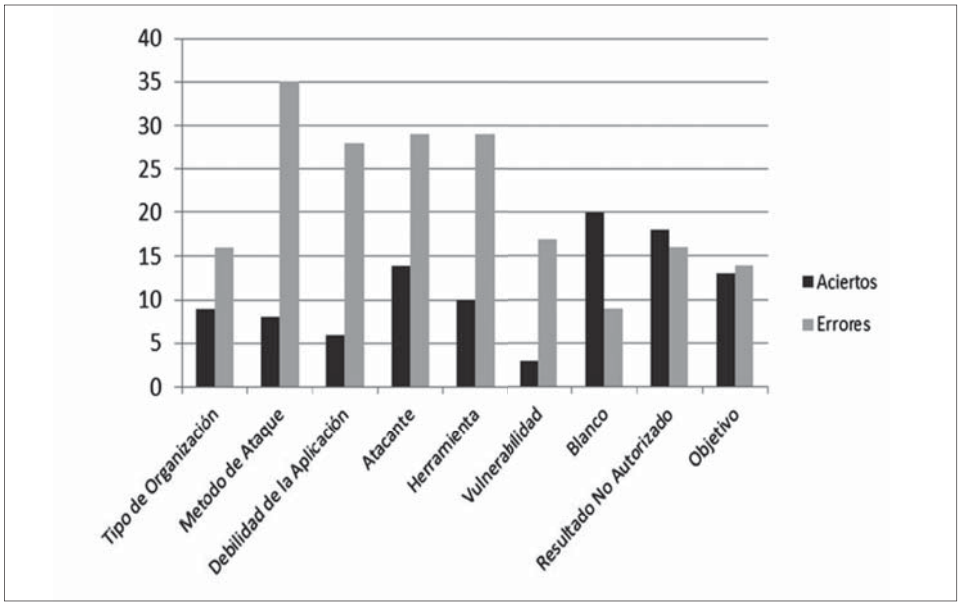

Figura 7. Cantidad de aciertos y errores mediante valor promedio

El método del valor por promedio muestra resultados menos efectivos en comparación a los del clasificador Bayesiano, el porcentaje de aciertos del clasificador Bayesiano (69%) es más del doble del porcentaje del método del promedio (34%) y además el valor promedio no logró sobrepasar en ninguno de los campos al competidor.

6. Resultados

Finalmente se obtuvo un repositorio de 1234 registros con todos los atributos diligenciados (Tipo de Organización, Entidad, Método de Ataque, Debilidad de la Aplicación, País, Fecha, Atacante, Herramienta, Vulnerabilidad, Blanco, Resultado No Autorizado y Objetivo), la técnica usada permitió completar el 76 % de los registros con valores faltantes, a pesar de la estimación Bayesiana, fue necesario eliminar 79 registros de los 1313 originales, pues éstos presentaban valores faltantes en las columnas **Método de Ataque**, **Debilidad de la Aplicación**, **Herramienta** y **Vulnerabilidad**, es decir una pérdida de información del 44% de la totalidad de atributos usados para hallar las probabilidades *a-posteriori*.

El repositorio final muestra (Figura 8) que los métodos de ataque predilectos son la Divulgación No Intencional (333 incidentes), los troyanos (161 incidentes), la inyección de SQL (134 incidentes) y el Cross Site Scripting (122 incidentes). El uso de Virus comunes se ve rezagado (34 incidentes), al igual que los gusanos (14 incidentes). Los métodos menos populares son Clickjacking y Toolkit (ambos con 4 incidentes), ARP spoofing y Suplantación de Contenido (1 incidente cada uno).

La Divulgación no Intencional fue causada en un 83% de los casos por Usuarios Internos, esto demuestra la insuficiente implantación de Políticas de Seguridad y el énfasis que debe aplicarse a la capacitación de los empleados sobre la importancia de la seguridad de la información. El alto uso de troyanos lleva a la misma conclusión, ya que en el 93% de los casos la debilidad de la aplicación fue Falta de Configuración en el Antivirus.

Para los incidentes de SQL Injection y Cross Site Scripting, la Vulnerabilidad fue Implementación en un 87% y 84% respectivamente, la responsabilidad de estos incidentes recae en los desarrolladores del software. La deficiencia en el desarrollo del software se centra en la codificación de la interfaz de usuario como se ve en la figura 9, con 152 incidentes causados por un Manejo Impropio de Salidas y 150 por Manejo Impropio de Entradas. El número de incidentes de este tipo debería ser tomado en cuenta en las empresas con el fin de analizar la capacidad de los desarrolladores contratados y tomar medidas de tipo gerencial, como capacitar o cambiar el equipo de desarrolladores.

En contra de una suposición popular, las entidades de tipo financiero obtuvieron el cuarto lugar de tipo de entidades atacadas, siendo las más vulneradas las que clasifican en la categoría BSO - Negocios - Otros, conformado por empresas de tecnología, medios de comunicación y hoteles entre otros. Por otra parte, las instituciones dedicadas a la salud son las menos atacadas. En la figura 10 se observa la distribución de incidentes por categoría.

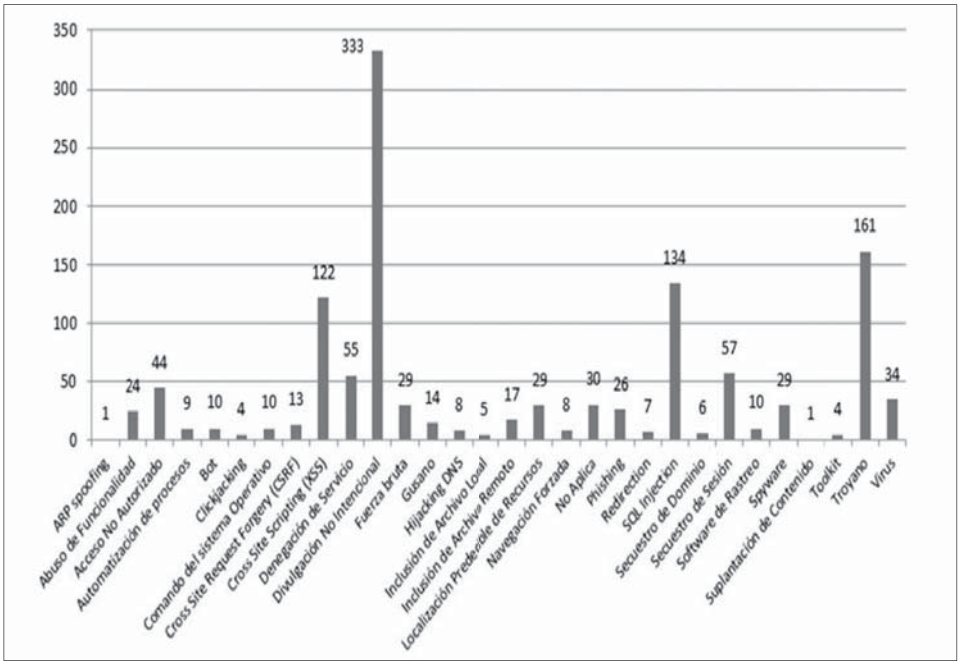

Figura 8. Distribución final de Valores Método de Ataque

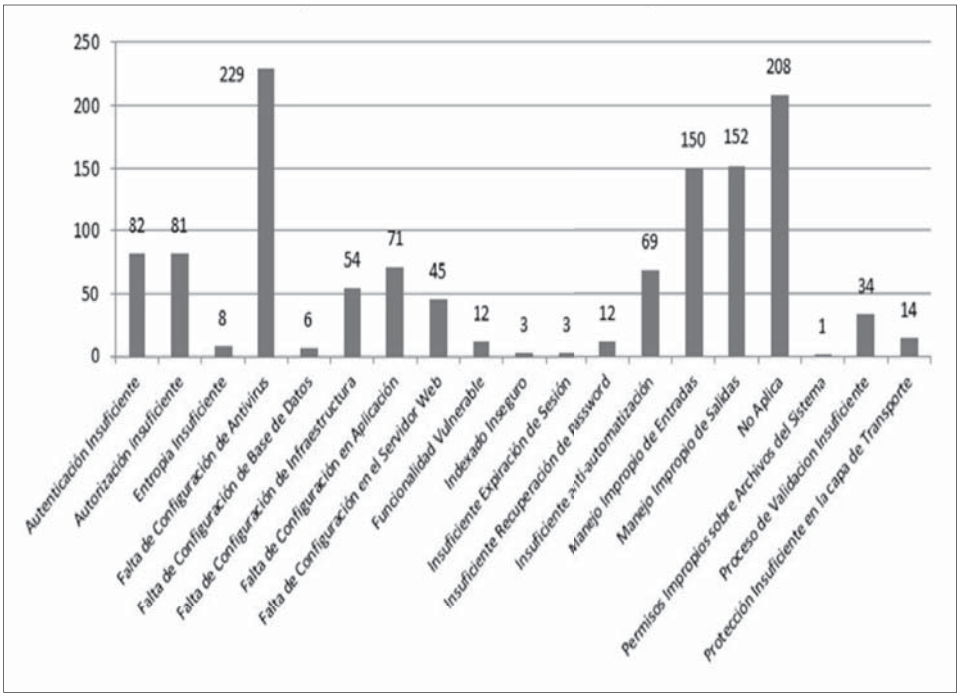

Figura 9. Distribución final de Valores Debilidad de la Aplicación

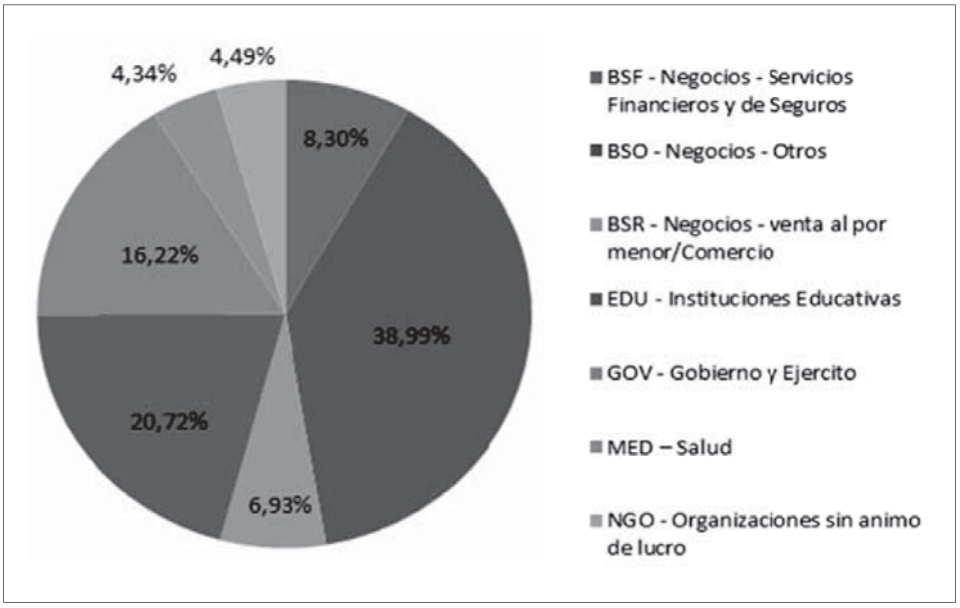

Figura 10. Distribución de incidentes por Tipo de Organización

7. Conclusiones

El reporte de incidentes de seguridad es una práctica que cada empresa debe tener en cuenta, esto con el fin de permitir rastrear y analizar los ataques y sus posibles relaciones, teniendo en cuenta que en Colombia las entidades financieras están reglamentadas por la circular 052 del 2007 de la Superintendencia Financiera [21] que fijó un estándar para la seguridad y calidad en el manejo de la información, enfocada a los datos de los clientes asegurando confidencialidad y disponibilidad. La taxonomía propuesta en este artículo se propone como solución para el reporte de los incidentes, pues permite determinar los perfiles de los mismos, con campos fáciles de diligenciar, evitando que los reportes sean descripciones ambiguas.

Según la norma ISO 27001[22] un sistema de gestión de seguridad (SGSI) debería implementarse en instituciones de cualquier naturaleza; en este sentido es de gran utilidad contar con un registro de los incidentes informáticos y un análisis estadístico de los mismos que permita encontrar las causas de un incidente que tengan como consecuencias inconformidades respecto a un conjunto de requerimientos.

De acuerdo a lo anterior, como trabajo futuro fruto de esta investigación, se propone la publicación del repositorio final obtenido en un sitio Web, además del desarrollo de una página que permita diligenciar los campos de la taxonomía propuesta para reportar nuevos incidentes de seguridad informática, con el fin de alimentar el repositorio con nuevos registros. Una vez esto se logre, el repositorio puede ser utilizado para investigaciones similares y/o enfocadas a la minería de datos, que permitan obtener conclusiones de gran valor para el área de seguridad informática, un tema que en la actualidad tiene una relevancia trascendental en el mundo.

Glosario

Concepto	Definición
Hacker	Un individuo el cual realiza acciones impulsado por reto, estatus, o la emoción de obtener acceso.
Criminales Profesionales	Ente el cual realiza acciones impulsado por obtener ganancias financieras para su propio bien.
Espías	Atacante el cual realiza acciones impulsado en obtener información la cual será utilizada para generar ganancia política.
Terroristas	Atacante el cual realiza acciones impulsado en generar miedo para obtener ganancia política.

Ladrón Corporativo	Individuo el cual realiza acciones impulsado en obtener información de la competencia corporativa para generar ganancia financiera.
Vándalos	Delincuente el cual realiza acciones impulsado en causar daño a una o varias entidades.
Agente Autónomo	Programa o fragmentos de programas los cuales operan independientemente del usuario. Algunos ejemplos son los virus o gusanos.
Ataque Físico	Robo físico o daño de un computador, red sus componentes o los sistemas soportados (tal como aire acondicionado, energía eléctrica, etc.)
Comando de usuario	Interacciones con una interfaz de entrada del sistema. Un ejemplo es ingresar comando de Unix a través de una conexión Telnet, o comandos un puerto SMTP.
Intercambio de Información	Un medio para obtener información desde otros atacantes o de personas siendo atacadas (llamada ingeniería Social).
Script o Programa	Indica el uso de programas diseñados específicamente para explotar vulnerabilidades o scripts que automatizan la ejecución de comando de usuario. Un ejemplo es un Shell Script para explotar un bug de un software, o un software de cracking para obtener Password.
Toolkit	Un paquete que contiene scripts, programas y/o agentes autónomos diseñado para facilitar la labor de atacantes inexpertos. Un ejemplo es el Toolkit llamado Rootkit.
Diseño	Fallo en la especificación de los requerimientos funcionales y/o no funcionales.
Implementación	Error en la codificación del software y/o falta de buenas prácticas de programación
Configuración	Falta de ajuste en los parámetros del software y/o hardware, tal como, credenciales con una contraseña por defecto, permisos de escritura globales para archivos, y/o contar con servicios vulnerables habilitados. También incluye la falta de actualización automática de software.

Flood	Acceder a un objetivo repetidamente con el propósito de sobrecargar la capacidad del objetivo.
Bypass	Evadir un proceso usando un método alternativo para acceder a un objetivo.
Spoof	Asumir la apariencia de una entidad diferente en una red de comunicaciones.
Cuenta	Entidad lógica compuesta por el acceso al domino del usuario sobre un computador, red, o sistema; el cual es controlado acorde a un registro de información el cual contiene el nombre del usuario, contraseña y restricciones de uso.
Proceso	Conjunto formado por un programa ejecutable, los datos del programa, contador del programa, puntero del programa, y toda la otra información necesaria para ejecutar dicho programa.
Dato	Unidad lógica compuesta por representación de hechos, conceptos, o instrucciones, de una manera adecuada para comunicación, interpretación o procesamiento por humanos o por medios automáticos. Los datos puede encontrarse en forma de archivos en la memoria volátil o no volátil de un computador, o un dispositivo de almacenamiento, o en la forman de un dato en tránsito a través de un medio de transmisión.
Componente	Elemento físico que conforman un computador o red.
Computador	Entidad física compuesta por un dispositivo que consiste de uno o más componentes asociados, incluyendo unidades de procesamiento y unidades periféricas, que son controladas por programas almacenados y que pueden realizar procesos computacionales, que incluye operaciones aritméticas, u operaciones lógicas, sin intervención humana durante la ejecución.
Red	Entidad física y/o lógica compuesta por un grupo de host (computadores) interconectados o interrelacionados, elementos de switching, y ramas interconectadas.
Acceso incrementado	Obtención de privilegios de lectura, escritura y/o borrado no contemplados.
Difusión de la información	Divulgación de información privada, donde dicha información no es autorizada para la visualización a los usuarios que fue divulgada.

Corrupción de la información	Resultado de la alteración no autorizada de datos en un computador, sistema, o red.
Negación del Servicio	Desenlace de la degradación no intencional o bloqueo de un computador, recursos de red o un sistema.
Robo de recursos	Producto del uso no autorizado de un computador, sistema, o recursos de red.
Cambio de Estatus	Objetivo asociado a la necesidad de ser visto como un atacante habilidoso para violar la seguridad de cualquier sistema.
Daño	Finalidad asociada a la necesidad del atacante para ocasionar daño a cualquier sistema.
Ganancia Financiera	Intención asociada a la necesidad del atacante para obtener beneficios financieros.
Ganancia política	Búsqueda para obtener un mejor posicionamiento político, debido a la degradación verbal o visual de algún partido político, o grupos con pensamientos políticos.

Referencias

[1] John D. Howard, Thomas A. Longstaff. "A Common Language for Computer Security Incidents". Sandia National Laboratories, Octubre 1998.

[2] Carlos Carvajal, Diego Bayona, Análisis de Incidentes de Seguridad Informática Mediante Minería de Datos, para Modelado de Comportamiento y Reconocimiento de Patrones, Tesis para acceder al título de Ingeniería de Sistemas de la Universidad Francisco José de Caldas, Bogotá, Colombia, 2012

[3] Web Application Security Consortium. "Web Hacking Incident DataBase", http://projects.webappsec.org/w/page/13246995/Web-Hacking-Incident-Database

[4] Privacy Rights ClearingHouse, "Chronology of Data Breaches". http://www.privacyrights.org/data-breach

[5] COL-CSIRT, Grupo de Investigación Universidad Distrital. http://gemini.udistrital.edu.co/comunidad/grupos/arquisoft/colcsirt/

[6] CERT Brasil. "Centro de estudios de respuesta y tratamiento de incidentes de Seguridad de Brasil". http://www.cert.br/

[7] INTECO España – Instituto Nacional del Tecnologías de la Comunicación. Centro de respuesta a incidentes de seguridad TIC. http://cert.inteco.es

[8] "Hackers hit Tunisian websites" ALJAZEERA, Red de Noticias, 3 de Junio 2011, http://www.aljazeera.com/news/africa/2011/01/201113111059792596.html

[9] "Secunia recovers from DNS redirection hack" The Register, Publicación Online de Tecnología, 26 de Noviembre de 2010. http://www.theregister.co.uk/2010/11/26/secunia_back_from_dns_hack/

[10] "Cops: Hacker Posted Stolen X-rated Pics on Facebook" PCWorld, Magazín de Computación, 2 de Noviembre de 2010. http://www.pcworld.com/businesscenter/article/209584/cops_hacker_posted_stolen_xrated_pics_on_facebook.html

[11] COMPUTERWORLD, Magazín de Computación. http://www.computerworld.com/s/article/9149218/Bank_sues_victim_of_800_000_cybertheft

[12] "Bank sues victim of $800,000 cybertheft" The Washington Post, Periódico de la ciudad de Washington, 26 de Enero de 2010. http://www.washingtonpost.com/wp-dyn/content/article/2010/03/18/AR2010031805464.html

[13] "Hackers Take Over BP Twitter Feed" FOX NEWS, Red de noticias, 27 de Mayo de 2010. http://www.foxnews.com/tech/2010/05/27/hackers-bp-twitter-feed/

[14] Jiawei Han, Micheline Kamber. "Data Mining: Concepts and Techniques". Morgan Kaufmann Publishers 2000, Capitulo 3 "Data Preprocesing".

[15] Roderick J. A. little, Donald B. Rubin. "Statistical Analysis with Missing Data". Capitulo 6 "Theory of Inference Based on the Likelihood Function".

[16] Dorian Pyle. "Data Preparation for Data Mining". Capitulo 8 "Replacing Missing and Empty Values". http://www.temida.si/~bojan/MPS/materials/Data_preparation_for_data_mining.pdf

[17] Jaisheel Mistry. "Estimating Missing Data and Determining the Confidence of the Estimate Data"Seventh International Conference on Machine Learning and Applications, 2008.

[18] Ludmila Himmelspach. "Clustering Approaches for Data with Missing Values: Comparison and Evaluation",Institute of Computer Science Heinrich-Heine-Universität Düsseldorf, 2010.

[19] Jiawei Han, Micheline Kamber. "Data Mining: Concepts and Techniques". Morgan Kaufmann Publishers 2000, Capitulo 8 "Cluster Analysis".

[20] Xiao-Bai Li . "A Bayesian Approach for Estimating and Replacing Missing Categorical Data". ACM Journal of Data and Information Quality, Vol. 1, No. 1, Article 3, Junio 2009.

[21] "ABC de la Circular Externa 052 DE 2007", Comunicado de Prensa, Superintendencia Financiera de Colombia, Noviembre 7 de 2007. http://www.superfinanciera.gov.co/

[22] "ISO/IEC 27001 : 2005 Tecnología de la Información- Técnicas de seguridad- Sistemas de gestión de seguridad- Requerimientos ". Estándar Internacional, Octubre 2005. http://www.iso.org/iso/home.html

Carlos Javier Carvajal Montealegre

Es Ingeniero de Sistemas de la Universidad Distrital Francisco José de Caldas de Bogotá, Colombia. Participó en el grupo de trabajo ACM UD-GIIA (capitulo estudiantil de la Association for Computing Machinery de la Universidad Distrital Grupo de Interés en Inteligencia Artificial) como líder del tema de algoritmos genéticos para prototipo de agente autónomo (2008). Participó en el grupo de investigación GIIRA-UD de la Universidad Distrital en el tema Factores Humanos y Multimedia (2007-2008). Actualmente se desempeña como Analista Programador y Líder Técnico de Proyectos de Software. e-mail: Ing.carlosj@gmail.com

Diego Nicolás Bayona Sastoque

Es Ingeniero de Sistemas de la Universidad Distrital Francisco José de Caldas de Bogotá, Colombia. Participó en el grupo de trabajo ACM UD-GIIA (capitulo estudiantil de la Association for Computing Machinery de la Universidad Distrital Grupo de Interés en Inteligencia Artificial) como líder del tema de algoritmos genéticos para prototipo de agente autónomo. 2008. Actualmente se desempeña como Services Manager en Spring Mobile Solutions S.A.S. e-mail: nicolas.bayona@gmail.com

Zulima Ortiz Bayona

Es Matemática de la Universidad Nacional, Bogotá, Colombia. Obtuvo su título de Maestría en Matemáticas en la Universidad Nacional de Colombia. Es especialista en Teleinformática de la Universidad Distrital Francisco Jose de Caldas. Es integrante del Grupo de Investigación Arquisoft Rama de Seguridad Informática de la Universidad Distrital Francisco José de Caldas. Profesora Universidad Nacional. Actualmente se desempeña como profesora en el área de Matemáticas en la Universidad Distrital Francisco Jose de Caldas en Bogotá, Colombia. e-mail: zortiz@udistrital.edu.co