Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

Tipo de artículo: Artículo original
Temática: Inteligencia artificial
Recibido: 30/03/2013 | Aceptado: 20/02/2014

Solución basada en el Razonamiento Basado en Casos para el apoyo a las auditorías informáticas a bases de datos

Solution based on Case-Based Reasoning for supporting a computer auditing database

Yasser Azán-Basallo[1*], Leslye Bravo-García[1], Wilfredo Rosales-Romero[1], Dayán Trujillo-Márquez[1], Ernesto Arbois García-Romero[1], Annia Pimentel-Rivero[1]

[1] Centro Telemática. Facultad 2. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

[*]Autor para la correspondencia: yazan@uci.cu

Resumen

En el Departamento de Seguridad Informática de ETECSA a través de matrices de diagnóstico o listas de chequeo, se realiza el proceso de auditoría a los Sistemas Gestores de Bases de Datos. Después de terminado el monitoreo de los SGBD, los expertos determinan el nivel de riesgo de la seguridad de la información en los términos de: Alto, Medio y Bajo. Se propone la utilización de la técnica de inteligencia artificial Razonamiento Basado en Casos para emplearla en la etapa de evaluación del riesgo de seguridad de la información en los sistemas gestores de bases de datos para aprovechar la experiencia acumulada en las auditorías anteriores de este tipo. Se apoyó en los especialistas de ETECSA en la determinación de los rasgos que conforman el vector de los casos. La incorporación de la técnica RBC para el apoyo del análisis de las auditorías de seguridad informática a los gestores de bases de datos, agiliza el proceso y ayuda en el análisis de los riesgos de seguridad informática a los auditores.

Palabras clave: Auditoría de seguridad informática; matriz de diagnóstico; razonamiento basado en casos; razón de semejanza.

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

Abstract

In the Security Department Computing ETECSA through diagnostic matrices or checklists, the audit process is performed to Database Management Systems. After completing the monitoring of DBMS, experts determine the risk level of information security in terms of High, Medium and Low. The use of artificial intelligence technique Reasoning Case-Based, for use in the analysis phase of evaluation of the risk of security of the information to take advantage of the experience gained in previous audits of this type is proposed. He leaned on ETECSA specialists in determining the features that make the vector cases. The incorporation of Reasoning Case-Based technique to support the analysis of information security audits managers' database, streamlines the process and helps in the analysis of risks to information security auditors.

Keywords: *Case-based reasoning, computer security audit, diagnostic matrix, similarity ratio.*

Introducción

Los avances en los sistemas de información (SI) y la tecnología utilizada para soportar estos sistemas han producido grandes resultados para organizaciones, negocios y otras agencias en términos de productividad del trabajo, almacenamiento de la información, administración y oportunidad de ventajas competitivas. Mientras los SI prometen y ofrecen tremendos beneficios, estos también representan un significativo y sin precedentes, mayores niveles de riesgo para las operaciones organizacionales. Negocios, hospitales, escuelas, universidades, agencias gubernamentales y bancos dependen fuertemente de los SI. Esto incrementa la necesidad de la seguridad de la información (Quigley, 2008).

Uno de los SI a los cuales con mucha frecuencia son objetivos de atacantes son los gestores de bases de datos, como exponen (Ramakanth & Vinod, 2011): El 17 de agosto de 2009, el Departamento de Justicia de los Estados Unidos acusó a un ciudadano por el robo de 130 millones en tarjetas de crédito usando ataques de inyección de SQL. Aproximadamente 500.000 páginas web que usaban como servidor el Microsoft IIS y el servidor de SQL, fueron atacadas entre abril y agosto del 2008 usando la inyección de SQL.

La cantidad de vulnerabilidades reportadas a través de este ataque han ido aumentando en los últimos años según el Instituto Nacional de Vulnerabilidades de Estados Unidos de América como lo muestra la Tabla 1.

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

Tabla 1. Datos estadísticos de la vulnerabilidad: Inyección de SQL. (NIST, 2013).

Año	Número de Vulnerabilidades	% de Total
2007	252	3.87
2008	1,092	19.39
2009	948	16.54
2010	515	11.10
2011	289	6.96
2012	237	4.48
2013	116	2.59

Este mismo instituto demuestra como muchos otros tipos de vulnerabilidades de bases de datos han ido aumentado en años recientes. Uno de ellos es: Permisos, Privilegios, y Control de Acceso, cuyos datos están reflejados en la Tabla 2.

Tabla 2. Datos estadísticos de la vulnerabilidad: Permisos, Privilegios, y Control de Acceso. (NIST, 2013)

Año	Número de Vulnerabilidades	% de Total
2007	219	3.36
2008	449	7.97
2009	436	7.61
2010	356	7.67
2011	283	6.82
2012	604	11.42
2013	496	11.07

Estos datos demuestran la importancia que tiene hoy para la seguridad de la información, proteger los datos a las vulnerabilidades detectadas en los gestores de bases de datos.

Para la seguridad de la información, son importantes los controles de seguridad a los sistemas informáticos. Entre las razones está el impacto de los controles de seguridad de los sistemas informáticos a otros controles generales, la

vulnerabilidad de los sistemas de computadoras hacia la pérdida de recursos, el impacto del fracaso de la seguridad en datos confiables, la posibilidad de faltar a los cumplimientos con los requisitos legales, la posibilidad de pérdida de la contingencia si el riesgo del proceso de los datos son graves y no están asegurados y la vulnerabilidad de los sistemas de computadora a ser usados sin autorización.

Una auditoría de seguridad informática es un concepto que es relevante para la seguridad de la información, el cual según el autor (Ruiz, 2011): "es el estudio que comprende el análisis y gestión de sistemas para identificar y posteriormente corregir las diversas vulnerabilidades que pudieran presentarse en una revisión exhaustiva de las estaciones de trabajo, los servidores y las redes de comunicaciones". Durante una auditoría de seguridad informática se realizan las auditorías de la seguridad lógica y auditoría de las comunicaciones.

Uno de los pasos para garantizar la integridad de los datos, es a través de la realización de auditorías informáticas periódicas a las tecnologías de cómputo. Para las auditorías de seguridad informática se pueden utilizar las matrices de diagnósticos o listas de chequeo.

En el Departamento de Seguridad Informática (DSI) de la empresa cubana ETECSA tiene entre sus principales responsabilidades garantizar y mantener la integridad de los gestores de bases de datos (GBD), sistemas operativos y aplicaciones web que soportan todo el trabajo de las telecomunicaciones en Cuba. Uno de los pasos para garantizar la integridad de los datos, es a través de la realización de auditorías informáticas periódicas a las mencionadas tecnologías.

El departamento mencionado tiene estandarizado con matrices de diagnósticos todo el proceso de revisión de los sistemas GBD. De acuerdo a los autores (Broder & Tucker, 2011), las listas de chequeo son usadas para facilitar la recolección de información pertinente. Estas pueden tener muchas formas. Pueden ser simples listas de preguntas de si o no, o de preguntas de final abierto requiriendo respuestas en formas de redacción. Ellas pueden ser breves y enfoques limitados en la específica operación o actividad en cuestión, o pueden ser extensas en alcance y cubrir lo común en lo concerniente a la seguridad de todo las operaciones de la compañía. El propósito de las listas de chequeo es proporcional una grabación lógica de la información y asegurar que preguntas no importantes vayan sin ser solicitadas.

En la tabla 3 se muestra un fragmento debido a la extensión de la misma, de la lista de chequeo publicada por el Instituto de Seguridad en Internet conocida por sus siglas en inglés CIS de los Estados Unidos, que son usadas por los auditores, para el sistema gestor de base de datos (SGBD) MySQL.

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

Tabla 3. Fragmento de la lista de chequeo de parámetros a evaluar para el gestor de base de datos MySQL 4.1 (CIS, 2013).

Parámetro	Abreviatura
Configuración del Sistema Operativo	CSO
Máquina dedicada	MD
Ejecutar MySql en modo chroot.	ECR
Cuentas de servicio	CS
Permisos sobre el File System	PFS
Logs de errores	LE
Permisos sobre los directorios de datos	PDD
Permisos sobre los ficheros binarios	PFB
Ficheros de configuración	FC
Logs	L
Logs de errores	LE
Directorio de Logs	DL
Log Update	LU
General	G
Version del SGBD	VS
Cuenta del usuario root	CUR

Una vez recogida los datos en el monitoreo de los SGBD, se realiza la evaluación del riesgo según el monitoreo guiado por la lista de chequeo.

Los problemas en las auditorías de seguridad informática utilizando las listas de chequeo vienen dados muchas veces en el nivel de experiencia de un auditor para evaluar el nivel de riesgo de cada parámetro existente en la lista de chequeo. Por lo que el resultado de una evaluación de una auditoría de seguridad informática puede estar sustentado en buena medida de la experticia del auditor presente, por lo que existe una alta dependencia del mismo. Otro fenómeno presente en la evaluación del nivel del riesgo es que debido a las diferencias existentes en cuanto a la experticia de los expertos, los resultados pueden no tener la misma consistencia en todos casos.

Además en este departamento no cuentan con herramientas de computación para la realización de auditorías informática a los SGBD, porque son privativas y con precios muy altos. Las herramientas libres no ofrecen una solución unificada, no soportan los procesos de auditoría del DSI de ETECSA. Esta situación provoca que las

auditorías se realicen ejecutando scripts con consultas SQL a cada gestor y con comandos del sistema operativo donde se encuentre instalado el SGBD. Llegando a demorar una auditoría a una entidad de un día o dos.

Una vez recogida los datos que aportan la ejecución de los *script*, se realiza la comprobación de resultados. Este proceso se efectúa manualmente evaluando los datos con las matrices de diagnósticos. A su vez el informe general de auditoría, documento que valida el cumplimiento de las auditorías realizadas, es también confeccionado manualmente a partir de las deficiencias detectadas por los especialistas evaluadores de la seguridad informática con las matrices de diagnóstico de la auditoría, determinando el nivel de riesgo al cual están expuestas las tecnologías de la información evaluadas.

Además de lo expuesto anteriormente, a causa del mismo proceso, se han acumulado en una base de datos referencial cientos de auditorías de seguridad informáticas. Estos datos no se han utilizado a no ser para mantener archivado por seguridad, como vestigio comprobatorio de las auditorías realizadas.

Materiales y métodos

El cálculo del riesgo

El riesgo es la probabilidad de la causa de un problema cuando una amenaza es provocada por las vulnerabilidades (Arunabha, Chatterjee, Saha, Ambuj, & Sadhukhan, 2006). La probabilidad de sufrir daños o pérdidas. Se refiere a una acción, evento o un fenómeno natural que podría provocar un resultado indeseable, lo que resulta en un impacto negativo o una consecuencia (Center, 2010).

La fuente del problema es la vulnerabilidad y el problema en sí son las amenazas. Las amenazas están mucho más relacionadas con las características de los recursos y las vulnerabilidades son relevantes para los controles de seguridad (Arunabha, *et al*., 2006).

El cálculo del riesgo utilizando un método cuantitativo puede ser a través de la fórmula que publican (Bertolín, 2008) y (Burtescu, 2009):

$R_n = Impacto_n \times La\ probabulidad\ de\ ataque_n$ (1)

R_n: Es el riesgo perteneciente al parámetro n de la lista de chequeo.

El impacto es el peso del costo de la pérdida de un recurso del parámetro n de la lista de chequeo. Cada parámetro n de la lista de chequeo tiene un impacto cuyo valor es un término lingüístico de los vistos anteriormente para los niveles de riesgos. Los valores numéricos de los umbrales de los términos lingüísticos para el impacto se deben dejar a consideración del auditor, según cada caso, ya que no todas las entidades a auditar tienen las mismas características en cuanto a la seguridad de la información.

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

$La\ probabilidad\ de\ ataque_n$: Se puede efectuar a partir de lo planteado por el autor (Bertolín, 2008), partiendo de los hechos sucedidos con anterioridad en la organización o a través de la entrada de los valores por los expertos directamente al sistema. Se debe definir una escala por ejemplo del 0 al 1, donde los valores más altos implican mayor probabilidad.

Para proveer de una organización de los riesgos de seguridad de la información en los gestores de bases de datos se creó un organigrama de evaluación de los riesgos, de acuerdo con (Berger, 2003), a partir del fragmento de la lista de chequeo presentada en la figura 1, donde el primer nivel es el riesgo de la información global denotado como *RG;* el segundo nivel se encuentra los riesgos de los indicadores en los cuales están agrupados los parámetros; en el tercer nivel se ubican los riesgos de los parámetros de la lista de chequeo. A partir de la fórmula anterior y con la organización de los parámetros como se muestra en el organigrama se puede llegar al riesgo global, según (Berger, 2003).

$$\mathrm{RG} = \frac{\mathrm{CSO}}{4} + \frac{\mathrm{PFS}}{4} + \frac{\mathrm{L}}{4} + \frac{\mathrm{G}}{4} \quad (2)$$

Para determinar el riesgo de *CSO, PFS, L y G,* que son los parámetros generales o del segundo nivel existente en la figura 1, se debe utilizar los riesgos de los parámetros que ellos agrupan de la siguiente forma:

$$\mathrm{CSO} = \frac{\mathrm{MD}}{3} + \frac{\mathrm{ECR}}{3} + \frac{\mathrm{CS}}{3} \quad (3)$$

$$PFS = \frac{\mathrm{LE}}{3} + \frac{\mathrm{PFB}}{3} + \frac{\mathrm{FC}}{3} \quad (4)$$

$$L = \frac{LE}{3} + \frac{\mathrm{DL}}{3} + \frac{\mathrm{LU}}{3} \quad (5)$$

$$G = \frac{\mathrm{VS}}{2} + \frac{\mathrm{CUR}}{2} \quad (6)$$

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

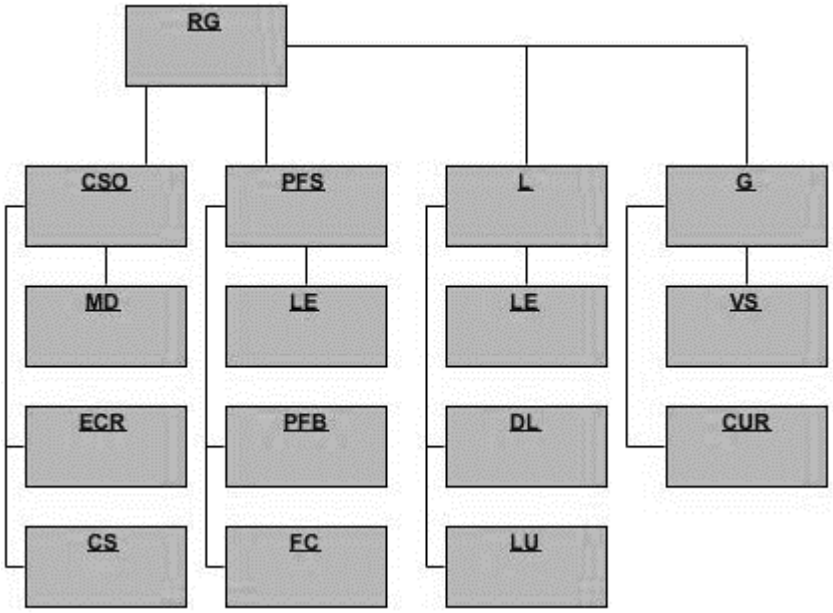

Figura 1. Organigrama de evaluación de los riesgos de seguridad de la información según la lista de chequeo.

Razonamiento basado en casos

El razonamiento basado en casos, señalan (Zhang, Lu, & Zhang, 2011), emplea las experiencias pasadas en forma de casos almacenados en una base de caso para apoyar la toma de decisiones en situaciones actuales similares. Una de las ventajas del razonamiento basado en casos, es la flexibilidad que ofrece respecto a la representación. Se puede elegir la implementación adecuada, aseguran (CIS, 2013), dependiendo del tipo de información a representar, variando desde un simple booleano, un número, datos dependientes del tiempo, relaciones entre datos, ficheros, *frames*, redes semánticas, etc.

Los sistemas expertos implementan con frecuencia esta técnica, los cuales se les conoce como Sistema Basados en el Conocimiento (SBC). En términos generales, un SBC puede ser definido como un sistema computarizado que usa conocimiento sobre un dominio para arribar a una solución de un problema de ese dominio. Esta solución es esencialmente la misma que la obtenida por una persona experimentada en el dominio del problema cuando se enfrenta al mismo problema Gálvez (Lio, 1998).

De acuerdo con los autores (Priti Srinivas Sajja & Akerker, 2010), algunas ventajas de los SBC son cuando existen las siguientes situaciones:

- El experto no está disponible.

- La experiencia se va a guardar para uso futuro o cuando la experiencia va a ser clonado o multiplicado.
- Se requiere la asistencia o entrenamiento inteligente para la toma de decisiones la solución del problema.
- El conocimiento de más de un experto tienen que ser agrupados en una sola plataforma.

Se escoge la técnica de razonamiento basado en casos (RBC) para emplearlo en la etapa de análisis de auditorías informáticas a gestores de bases de datos, para aprovechar la experiencia acumulada en las auditorías anteriores de este tipo almacenada en la base de datos y poder lograr informes de auditoría de seguridad más consistente entre los auditores noveles y los auditores de mayor experiencia.

Representaciones de los casos

Las formas de representar la solución, la justificación y el resultado varían según el dominio de aplicación del sistema RBC y la aproximación utilizada. En (Althoff, Aamodt, Magaldi, & Milne, 1995; Lozano & Fernández, 2008) distinguen dos grandes tipos de representaciones para los casos:

Representaciones planas: En este tipo de representación se define una serie de atributos, cada uno con un conjunto de posibles valores de tipos simples cadenas de caracteres, números, símbolos. En general no se define relación alguna entre los atributos ni entre sus valores, esto es, si utilizamos n atributos en la representación, un caso vendrá descrito por la n-tupla de valores de los atributos del caso en cuestión.

Representaciones estructuradas: También aquí se analizan listas de atributos y valores asociados, pero a diferencia de las representaciones planas, los valores pueden ser objetos que a su vez tienen atributos. De esta forma se consiguen expresiones más expresivas en las que se definen relaciones entre los atributos y/o entre los valores. Para implementar este tipo de descripciones se suelen utilizar cálculos de predicados, redes semánticas, lenguajes basados en marcos, lenguajes orientados a objetos y lógicas descriptivas, entre otros.

Características de la solución planteada

La representación de la matriz de diagnóstico se usó de forma plana, para mayor facilidad computacional ya que tiene la ventaja de que añadir nuevos casos resulta muy “barato” (rápido y fácil de implementar) (Lozano & Fernández, 2008). Va a existir variaciones según el gestor de base de datos que se vaya a auditar.

Los casos van a ser asignado por los siguientes rasgos:

- Rasgos descriptores:
    - Identificador: identificador único de la matriz de diagnóstico.

- IP: Es el número de IP otorgado al servidor que hospeda el gestor de base de datos.
- Entidad: Es el identificador de la entidad a la cual pertenece el servidor.
- Aplicación a la que pertenece el SGBD.
- El gestor de base de datos.
- La versión del SGBD.
- Parámetros: Son las propiedades de seguridad que se van a evaluar para verificar el estado de seguridad del servidor. Estos son los rasgos que van a determinar la base de casos. Los parámetros pueden variar de un gestor de base de datos a otro. En la tabla 3 se muestra un ejemplo de parámetros para el gestor de base de datos PostgreSQL. La cantidad de parámetros varía según el gestor de base de datos a analizar.
- Estado: Es el caso de que la matriz puede ocupar uno de los siguientes valores ordinales: cerrado, activo, en proceso.

– Rasgo objetivo: la evaluación de impacto riesgo según los valores encontrados en los parámetros, que pueden ser: ALTO, MEDIO Y BAJO.

Las representaciones planas en la recuperación de casos resultan muy lento cuando están hechas de forma secuencial, cuando el número de casos en la base es alto. Para contrarrestar este problema, los casos de almacenan en una base de datos referencial, posibilitando la recuperación de casos a través de consultas SQL (Lozano & Fernández, 2008). La recuperación de los casos a través de la base de datos tiene facilidades en la búsqueda de los casos semejantes, ya que está diseñada para hacer diferencias entre las auditorías según el tipo de gestor auditado. Dígase por ejemplo: PostgreSQL, MySQL, Oracle y otras. Incluso una clasificación que permite filtrar por la versión de gestor de base de datos. También es importante diferenciar en la recuperación de los casos, según la entidad auditada, ya que cada lugar tiene sus propias características en la seguridad de información, otra característica que el diseño de la base de datos permite. Todas estas maneras por las que se puede distinguir las bases de casos de interés, hacen que una base de datos para almacenar los casos sea conveniente.

Los rasgos pertenecientes a la matriz de diagnósticos que se muestra en la tabla 3, se obtuvo a partir de las mismas que utilizan los especialistas de ETECSA para realizar las auditorías informáticas a los gestores de bases de datos, por lo que se trabaja con los rasgos en acuerdo con los expertos.

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

Funciones de semejanza entre atributos

La selección de un caso semejante a un nuevo problema entre un grupo de casos almacenados, exige el uso de alguna herramienta que indique cuan semejantes son los casos analizados. Una medida que determine el grado de proximidad entre los casos, se conoce como "medida de semejanza".

El dominio de los rasgos descriptores a utilizar es numérico, debido que son los parámetros los utilizados para el cálculo del riesgo.

Función de semejanza para atributos numéricos

La distancia absoluta (llamada de *Manhattan* o de *City Block*) se representa como la diferencia absoluta sobre todas las dimensiones. Para cada parámetro existente en la lista de chequeo, se le debe aplicar la ecuación de distancia, por lo que para lograr que su resultado se ajuste al rango de valores existentes se debe ajustar de la siguiente forma:

$$d(\mathrm{z_{nr}},\ \mathrm{z_{ni}}) = \left|\frac{\mathrm{z_{nr}} - \mathrm{z_{ni}}}{r_{máx} + r_{mín}}\right| \quad (7)$$

Donde de un conjunto n de parámetros existente en la lista de chequeo. La variable z_{nr} corresponde al valor del resultado del cálculo del riesgo de seguridad de la información de un parámetro en un caso actual de la auditoría informática. La variable z_{ni} corresponde a un valor del riesgo de seguridad de la información i de un parámetro X_n, almacenada en la base de casos. Las variables $r_{máx}$ y $r_{mín}$ son los valores de rango máximo y mínimo respectivamente del conjunto de valores existentes.

Función de semejanza entre casos

La propuesta general de distancia entre los casos es:

$$\mathrm{f}\Big(\mathrm{W}, \mathrm{D}\big(\mathrm{X}(\mathrm{z_{nr}}), \mathrm{Y}(\mathrm{z_{ni}})\big)\Big) = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{w_i} \cdot \left|\frac{\mathrm{z_{nr}} - \mathrm{z_{ni}}}{r_{máx} + r_{mín}}\right|)}{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{w_i}} \quad (8)$$

Donde $D\big(X(z_{nr}), Y(z_{ni})\big)$ es la función de distancia vista en la fórmula 7. La variable w_i es el peso de importancia dada por los expertos a cada parámetro de la lista de chequeo. Cuando existan en los casos a analizar atributos ausentes, en estos casos se procede de la siguiente manera: cuando faltan los dos atributos no se lleva a cabo el cálculo de la distancia, pero cuando falta un atributo la distancia se hace 1, también llamada distancia trivial, de acuerdo con Díaz (Moreno, 1998).

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

El umbral de semejanza necesario para lograr llegar a determinar el nivel de semejanza entre los casos comparados se queda a criterio del experto. Esto se debe a que cada entidad a auditar tiene sus propias características de seguridad por lo que el valor del umbral de semejanza puede variar en cada auditoría.

Método de acceso y recuperación de los casos propuesto

El método de acceso a los casos más semejantes propuesto es a través de siguiente algoritmo:

1. El primer paso del algoritmo es buscar los casos pertenecientes a un gestor de base de datos y versión determinada. También se incluyen las propiedades para la búsqueda de los casos: la entidad, la aplicación e IP del servidor; para analizar primero los casos con estas 3 últimas propiedades.
2. Luego se inicia la comparación entre las distancias del Riesgo Global del valor del resultado del cálculo del riesgo de seguridad de la información *(RGr)* en un caso actual de la auditoría informática y el Riesgo Global de los casos *i (RGi)* existentes en la base de datos. La idea es buscar una cantidad *k* (número proporcionado por el experto) de casos similares. Los casos se guardan según la distancia de similitud y se agrupan en la base de datos en orden según esta distancia existente entre ellos; facilitando la agilidad de las consultas necesarias para su obtención.
3. Una vez obtenidos los *k* casos similares existentes en la base de datos al caso con valor *RGr,* se pasa a comparar las distancias con los parámetros más generales *N.* Como se muestra en la Figura 1, los parámetros *N* son aquellos parámetros pertenecientes al segundo nivel del organigrama de evaluación de los riesgos de seguridad de la información de la Figura 1. La comparación va a ser realizada a través de la fórmula de distancia (8). Como ya se tienen calculados los valores de los riesgos de cada parámetro general *N,* a través de las fórmulas (3, 4, 5 y 6), se obtiene la distancia entre los parámetros pertenecientes al segundo nivel del organigrama *Nr* y los parámetros *Ni* del caso *k* almacenado en la base de datos a través de la siguiente fórmula:

$$d(\mathrm{z_{Nr}},\ \mathrm{z_{Ni}}) = \sum_{i=1}^{i=k} \left| \frac{z_{\mathrm{Nr}} - z_{Ni}}{r_{máx} + r_{mín}} \right| \quad (9)$$

Con este cálculo de distancia entre los parámetros generales, se refina la selección de los *k* casos similares a otro número menor *p,* quienes se utilizarán para la adaptación de los resultados de cada parámetro del caso a resolver.

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

Método de adaptación de los casos para la respuesta del sistema

El método de adaptación utiliza la función de distancia (8) para discernir la similitud entre los parámetros del tercer nivel del organigrama en los p casos similares seleccionados con el método anterior.

1. El primer paso en la adaptación es comparar la distancia entre los riesgos de cada parámetro X_n del caso a resolver con los p casos similares.
2. Utilizar como solución a cada parámetro X_n del caso a resolver entre los p casos más similares el riesgo objetivo que más aparezca.
3. Para la solución del caso, tomar el valor de riesgo objetivo que más frecuente esté entre los p casos más similares.

Modelo de aprendizaje del sistema basado en casos

El autoaprendizaje se logra con la incorporación de los nuevos problemas que se van solucionando a través del mismo experto:

1. Para el autoaprendizaje se toma el valor obtenido de la distancia de la fórmula (8). Si la distancia es significativa se toman como un caso totalmente nuevo y se introduce en la base de datos.
2. En el caso que el valor de la función (8) esté en un rango de similitud a otros casos existentes, se utilizan los valores obtenidos por la función (9). Si existe una distancia entre un caso z_{Nr} y z_{Ni} cuya distancia d es menor que la distancia mínima (umbral aportado por el experto), entonces se desecha el nuevo caso de incorporarlo a la base de conocimientos. En caso contrario, el auditor experto debe decidir si el nuevo caso debe ser incorporado a la base de conocimiento.

Resultados y discusión

Se construyó una solución informática capaz de monitorear los SGBD: PostgreSQL, MySQL, SQL Server y Oracle ya que son los principales gestores hospedados en los servidores de ETECSA. Esta aplicación posibilita configurar los parámetros que quieren ser evaluados y adaptarse a la versión del SGBD. Además añadir nuevos *scripts* para monitorear adicionales parámetro e incorporarlos a los matrices de diagnósticos. Generar automáticamente un informe y proponer un nivel de riesgo. En la Figura 2, se muestra un ejemplo de datos de una matriz de diagnóstico.

Esta aplicación automatiza el proceso de monitoreo. Por lo que posibilita la disminución de tiempo en la ejecución de la auditoría en este sentido. Anteriormente a esta aplicación, una auditoría necesitaba para su culminación un día o a

Revista Cubana de Ciencias Informáticas
Vol. 8, No. 2, 2014
ISSN: 1994-1536 | RNPS: 0547
Pág. 52-68
http://rcci.uci.cu

varios, pero a partir de esta se logra generar el informe en pocos minutos con una evaluación del riesgo de seguridad de la información.

Además se logró mejorar la capacidad de trabajo de los auditores en cuanto a la cantidad de auditorías logradas a realizar y en la estandarización de los datos extraídos del monitoreo a las SGBD.

En la evaluación de los riesgos, los expertos toman en cuenta los resultados de anteriores auditorías, permitiendo mejor consistencia en los resultados y se aprovecha la experticia acumulada en las anteriores auditorías.

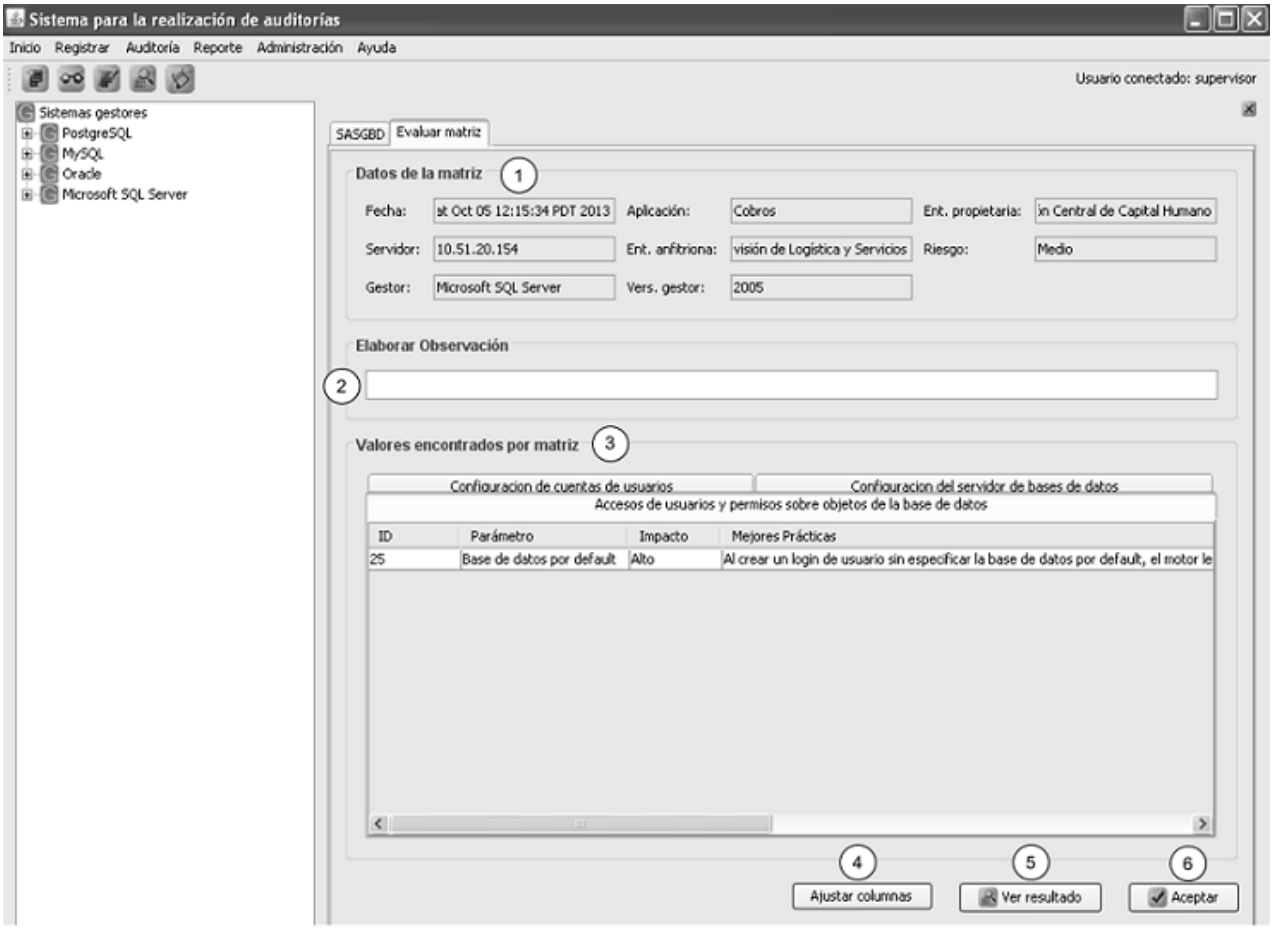

Figura 2. Pestaña del Sistema SASGBD que muestra los datos de una matriz de diagnóstico.

Antes del modelo, la aplicación arroja un resultado del cálculo del riesgo en números, pero los valores para los informes determinados en Alto, Medio y Bajo quedan en manos de auditor. Con el nuevo modelo se le propone al auditor un valor como resultado de la determinación del rasgo objetivo a partir de casos similares.

Se realizaron pruebas además del gestor de base de datos PostgreSQL, al MySQL y al Microsoft SQL Server. Arrojando que el método utilizado para el primer SGBD también opera para los otros, comprobando la generosidad de la solución.

En las pruebas realizadas quedó percibido que todavía el auditor necesita para realizar la auditoría, información que le tiene que proveer el administrador de gestor de la base de datos auditada en la entidad donde esté ocurriendo el hecho, pero ahora es menor.

Conclusiones

En este trabajo se ha presentado una solución para la evaluación del riesgo de la seguridad de la información a los gestores de bases de datos. Con la incorporación de la técnica RBC al análisis del riesgo de seguridad de la información, se propone un mecanismo que permita una mayor exactitud en la evaluación. Lograr una menor dependencia del auditor experto en la auditoría de bases de datos. Al disponer y utilizar las auditorías pasadas como conocimiento para ser reutilizado por los auditores noveles, permite mejorar la consistencia en los resultados.

Con la herramienta informática SASGBD se automatiza el proceso de monitoreo y análisis del riesgo de seguridad de la información, posibilitando una mayor rapidez en las auditorías de este tipo, pasando de varios días a un tiempo menor de varios minutos.

Referencias

- ALTHOFF, K., AAMODT, A., MAGALDI, R., & MILNE, R. Evaluating case-based reasoning systems. *Unicom Seminars & AI Intellingence*. 1995.
- ARUNABHA, M., CHATTERJEE, S., SAHA, D., AMBUJ, M., & SADHUKHAN, S. K. 2006, 04-07 January). *e-Risk Management with Insurance: A Framework Using Copula Aided Bayesian Belief Networks.* Paper presented at the System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on. 2006.
- BERGER, B. Data-centric quantitative computer security risk assessment. *Information Security Reading Room, SANS*. 2003. Disponible en: [http://www.sans.org/reading_room/whitepapers/auditing/data-centric-quantitative-computer-security-risk-assessment_1209#page=2&zoom=auto,0,725].
- BERTOLÍN, J. A. Seguridad de la información: redes, informática y sistemas de información. 2008. Disponible en:

[http://www.google.com.cu/books?hl=es&lr=&id=_z2GcBD3deYC&oi=fnd&pg=PP1&dq=gesti%C3%B3n+de+riesgo+de+seguridad+inform%C3%A1tica+ISO/IEC&ots=wqpnvGI0Ph&sig=Vyoy_6v-T-IReKYYOAHmjwi2N58&redir_esc=y#v=onepage&q&f=false].

- BRODER, J. F., & TUCKER, G. Risk Analysis and the Security Survey. 2011. Disponible en: [http://www.google.com.cu/books?hl=es&lr=&id=fLmgIGT18jIC&oi=fnd&pg=PP1&dq=risk+assessment%2Bformula%2Bsecurity&ots=q1KpmlGAY&sig=lwkAVW1G2jr7wkBjlR4dQftaQ2k&redir_esc=y#v=onepage&q=risk%20assessment%2Bformula%2Bsecurity&f=false].
- BURTESCU, E. DATABASE SECURITY - ATTACKS AND CONTROL METHODS *Journal of Applied Quantitative Methods*, *4*(4), 449-454 2009. Disponible en: [http://www.ijest.info/docs/IJEST12-04-04-014.pdf].
- Center, M.-S. I. S. a. A. Cyber Security: Risk Management A Non-Technical Guide Essential for Business Managers Office Managers Operations Managers 2010 [Consultado el: 12 de octubre de 2013]. Disponible en: [http://msisac.cisecurity.org/resources/guides/documents/Risk-Management-Guide.pdf#page=3&zoom=auto,179,0].
- CIS, C. f. I. S. CIS Benchmarks 2013. [Consultado el: 12 de noviembre de 2013]. Disponible en: [http://benchmarks.cisecurity.org/downloads/multiform/].
- Lio, D. G. *Sistemas Basados en el Conocimiento*. Departamento de Ciencia de la Computación, Facultad de Matemática, Física y Computación, Universidad Central "Martha Abreu" de Las Villas. Santa Clara, Cuba. 1998.
- LOZANO, L., & FERNÁNDEZ, J. Razonamiento Basado en Casos:"Una Visión General". 2008. Disponible en: [http://www.infor.uva.es/~calonso/IAI/TrabajoAlumnos/Razonamiento%20basado%20en%20casos.pdf].
- MORENO, J. M. D. Introducción a la topología de los espacios métricos. 1998. Disponible en: [http://books.google.com.cu/books?id=MrwJ35HBRDEC&printsec=frontcover&dq=Introducci%C3%B3n+a+la+Topolog%C3%ADa+de+Los+Espacios+M%C3%A9tricos&hl=en&sa=X&ei=CFgiUYzfO8-10AHss4GgAg&ved=0CCoQ6AEwAA].
- NIST, N. I. o. S. a. T. Datos estadísiticos de Inyección de SQL 2013. [Consultado el: 10 de noviembre de 2013]. Disponible en: [http://web.nvd.nist.gov/view/vuln/statistics-results?cves=on&query=&cwe_id=CWE-89&pub_date_start_month=-1&pub_date_start_year=2007&pub_date_end_month=-1&pub_date_end_year=-1&mod_date_start_month=-1&mod_date_start_year=-1&mod_date_end_month=-1&mod_date_end_year=-1&cvss_sev_base=&cvss_av=&cvss_ac=&cvss_au=&cvss_c=&cvss_i=&cvss_a=].

- PRITI SRINIVAS SAJJA, & AKERKER, R. Advanced Knowledge Based Systems: Model, Applications & Research Vol. 1. P. S. Sajja & R. Akerker (Eds.). 2010. Disponible en: [http://books.google.com.cu/books?id=F8pHNwrytFgC&printsec=frontcover&dq=Advanced+Knowledge+Based+Systems:+Model,+Applications+%26+Research&hl=es&sa=X&ei=I_TvUqv4BsXskQfQ74DgCQ&ved=0CCsQ6AEwAA#v=onepage&q=Advanced%20Knowledge%20Based%20Systems%3A%20Model%2C%20Applications%20%26%20Research&f=false].
- QUIGLEY, M. Encyclopedia of Information Ethics and Security. 2008. Disponible en: [http://books.google.com.cu/books?id=H2VuBddvMLAC&pg=PT432&dq=importance+of+the+audit+of+security+of+systems&hl=en&sa=X&ei=Z5NEUdzANeaV7AaXmoDgCA&ved=0CEcQ6AEwBQ].
- RAMAKANTH, D., & VINOD, K. SQL Injection - Database Attack Revolution And Prevention. *Journal of International Commercial Law and Technology*, *6*(4), 224-231. 2011. Disponible en: [http://www.jiclt.com/index.php/jiclt/article/view/141/139].
- RUIZ, A. J. *myEchelon: Un sistema de Auditoría de Seguridad Informática Avanzado bajo GNU/Linux*. Titulación de Ingeniero en Informática, Universidad de Almería, Almería, España. 2011. Disponible en: [http://www.adminso.es/images/9/9c/Alberto_PFC.pdf].
- ZHANG, J., LU, J., & ZHANG, G. A Hybrid Knowledge-based Risk Prediction Method Using Fuzzy Logic and CBR for Avian Influenza Early Warning. *Journal of Multiple-Valued Logic & Soft Computing*, *17*(4), 363-386. 2011. Disponible en: [http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=83c1714d-fa11-46ea-a0f3-c955d04587e0%40sessionmgr14&vid=1&hid=19].