

Tipo de artículo: Artículo original

Temática: Tecnologías de base de datos

Recibido: 04/10/2016 | Aceptado: 20/10/2016 | Publicado: 30/10/2016

SUBSISTEMA DE EVALUACIÓN DEL RIESGO DE SEGURIDAD DE LA INFORMACIÓN DEL SASGBD.

SUBSYSTEM RISK ASSESSMENT OF INFORMATION SECURITY SASGBD

Sailyn María Parra López^{1*}, Ubel Angel Fonseca Cedeño²

1 Centro de 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. smparra@uci.cu

2 Centro de 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. uafonseca@uci.cu.

¹e-mail: smparra@uci.cu

Resumen

El departamento de seguridad informática de ETECSA tiene entre sus principales responsabilidades garantizar y mantener la integridad de los datos en los sistemas gestores de bases de datos (SGBD). Por tal motivo, las auditorías de seguridad informática son parte esencial de las actividades llevadas a cabo por el departamento, siendo las evaluaciones del riesgo de seguridad de la información uno de los principales aspectos a tener en cuenta. Para realizar todo el proceso de auditoría se cuenta con el sistema para la realización de auditorías a los sistemas gestores de bases de datos (SASGBD), aunque el mismo no es capaz de ofrecer una evaluación de las auditorías tan acertada como la puede ofrecer un especialista del departamento. Con el propósito de contribuir con el proceso de toma de decisiones a la hora de evaluar las auditorías, se desarrolla un subsistema para la evaluación del riesgo de seguridad de la información; utilizando el razonamiento basado en casos y la lógica difusa. De esta manera se aprovecha la experiencia acumulada en las auditorías anteriores de este tipo y se evalúa el nivel de riesgo al cual está sometido el sistema auditado. Para el desarrollo de la propuesta de solución, se seleccionaron como principales tecnologías: los

marcos de trabajo Spring e Hibernate para la programación en Java, el SGBD PostgreSQL 9.1 y Visual Paradigm 8.0 como herramientas para el modelado. El proceso de desarrollo fue guiado por la metodología RUP.

PALABRAS CLAVES: Evaluación de riesgo; razonamiento basado en casos; lógica difusa; sistemas gestores de bases de datos.

Abstract

It security department of ETECSA has among its main responsibilities to ensure and maintain the integrity of the data in database management systems (DBMS). Therefore, it security audits are an essential part of the activities carried out by the department, being the risk assessments of information security one of the main aspects to consider. To make the entire audit process, it has the system for performing audits of databases management systems (SPADBMS), although it is not capable of offering a successful audit assessment as can offer an is specialist. For the purpose of contributing to the decision-making process at the time of evaluating audits, it's developed an information security risk assessment subsystem; by using case-based reasoning and fuzzy logic. Thereby the accumulated experience in previous audits of this type is used and the audited system risk level is evaluated. For the development of the proposed solution, they were selected as main technologies: spring and hibernate frameworks work for programming in java, the DBMS PostgreSQL 9.1 and 8.0 visual paradigm as tools for modeling. The development process was guided by the RUP.

KEY WORDS: Risk assessment; case-based reasoning; fuzzy logic; management systems databases.

Introducción

Los avances de los Sistemas de Información (SI) y las tecnologías han producido grandes resultados para las organizaciones, negocios y otras agencias en términos de productividad del trabajo, almacenamiento de la información, administración y oportunidad de ventajas competitivas. Sin embargo, a pesar de los disímiles beneficios brindados por los SI, estos también representan un indicador de riesgo bastante elevado para las organizaciones (Quigley, 2008).

Uno de los SI que con mucha frecuencia es objetivo de ataque, son los Sistemas Gestores de Bases de Datos (SGBD), por tanto, hoy día la seguridad de las bases de datos es de vital importancia como exponen (Ramakanth y 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 Internet Information Services (IIS) y el servidor de SQL, fueron atacadas entre abril y agosto del 2008 usando la inyección de SQL. En julio del 2008, el sitio web de malasia de karspersky fue atacado usando esta misma técnica.

El departamento de seguridad informática de la Empresa de Telecomunicaciones de Cuba S.A. (ETECSA) tiene entre sus principales responsabilidades garantizar y mantener la integridad de los datos en los SGBD. Para ello realiza las auditorías de seguridad informática, siendo las evaluaciones de riesgo de seguridad de la información (RSI) uno de los principales aspectos a tener en cuenta en dichas auditorías.

El departamento mencionado tiene estandarizado con listas de chequeo todo el proceso de revisión de los sistemas SGBD, las cuales no ofrecen la posibilidad de realizar la evaluación del riesgo; pero permiten detectar las vulnerabilidades y las amenazas en las que están expuestos los servidores de datos, premisa importante para determinar el nivel de RSI.

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 la 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 los casos. Además, el RSI se da en los términos: Alto, Medio o Bajo, por lo que se pueden generar ambigüedades en el resultado de la evaluación.

Aunque se cuenta actualmente con el Sistema para la realización de Auditorías a Sistemas Gestores de Bases de Datos (SASGBD), el mismo no es capaz de ofrecer una evaluación de las auditorías tan acertada como la puede ofrecer un especialista del DSI; además no le brinda a los especialistas elementos o criterios para tomar decisiones en el momento de realizar el diagnóstico del RSI en los SGBD. Este problema repercute en mayor tiempo y esfuerzo invertido por los especialistas para realizar las auditorías a los SGBD.

Por lo antes expuesto, se desarrolla un subsistema que contribuya con el proceso de toma de decisiones en las evaluaciones del RSI del SASGBD, haciendo uso de técnicas de la inteligencia artificial, como el razonamiento basado en casos y la lógica difusa para darle solución a los problemas presentes, de esta manera se aprovecha la experiencia acumulada en las auditorías anteriores de este tipo y se evalúa el nivel de riesgo al cual está sometido el sistema auditado.

Materiales y métodos o Metodología computacional

Estudio de sistemas similares

Se realizó un estudio de sistemas que evaluaran RSI. Encontrándose en su mayoría con sistemas inteligentes para el análisis del RSI. Estos sistemas, aunque están enfocados al riesgo de la seguridad en diferentes áreas, proponen a través de técnicas de Inteligencia Artificial una forma de evaluar, por lo que abren el camino para que puedan ser utilizadas las mismas en el área de la seguridad informática.

ALRAM y BDSS

El Sistema de Soporte de Decisión Bayesiana (con las siglas en inglés: BDSS) y el Automatizada Metodología (con las siglas en inglés: ALRAM) aplican algoritmos de evaluación del riesgo de la industria nuclear para la seguridad de la información, reemplazando el algoritmo FIPSPUB-65. BDSS es un sistema experto que proporciona al usuario una base de conocimiento amplia que dirige las vulnerabilidades y los resguardos, así como los factores de exposición de las amenazas y la frecuencia de los datos. Todos son totalmente mapeados y cruzados con algoritmos de modelación del riesgo e interfaces y presentación de lenguaje natural. ALRAM, sin embargo, requiere un experto para construir y mapear la base de conocimiento y entonces conducir a una evaluación del riesgo personalizada (Peltier, 2001).

MIDS

Es un Sistema de Detección de Intruso (con las siglas en inglés: MIDS). Tiene como principal ventaja una proporción baja de falsos positivos y las acciones preventivas y correctivas pueden ser realizadas

fácilmente. El sistema puede ser dividido en varias categorías, una de ellas es sistema experto. En esta categoría el sistema tiene un conjunto de reglas que describen los ataques empleados. Los eventos de auditorías son traducidos como hechos, llevando su significación semántica en el sistema experto y el motor de inferencia traza conclusiones usando ambas, reglas y hechos (Chou, 2011).

@RISK

Este sistema realiza análisis de riesgo utilizando la simulación para mostrar múltiples resultados posibles en un modelo de hoja de cálculo y le indica qué probabilidad hay de que se produzcan. Computa y controla matemática y objetivamente gran número de escenarios futuros y luego le indica las probabilidades y riesgos asociados con cada uno. Esto quiere decir que permite decidir qué riesgos tomar y cuáles evitar, tomando la mejor decisión en situaciones de incertidumbre (Palisade, 2014).

Luego de haber analizado las diferentes técnicas de la IA para la construcción de sistemas inteligentes, se consideró que las técnicas RBC y lógica difusa son adecuadas para evaluar el RSI en las auditorías de seguridad informática. Las mismas emplean las experiencias pasadas en forma de casos almacenados en una base de casos para apoyar la toma de decisiones en situaciones actuales similares. Se cuenta con una base de datos referencial de cientos de auditorías de seguridad informáticas. Estos datos no se han utilizado a no ser para mantenerlos archivados por seguridad, como vestigio comprobatorio de las auditorías realizadas. Por esta característica se emplea para determinar el diagnóstico en la evaluación del riesgo en los servidores de bases de datos auditados. Además se emplea la lógica difusa para manejar las ambigüedades existentes en la evaluación del RSI.

Método de acceso y recuperación de los casos

El algoritmo de acceso y recuperación tiene una secuencia lógica de pasos como se describe en (Martínez y Pérez , 2010), para la obtención de los casos similares que se utilizarán para obtener el diagnóstico de la evaluación del riesgo de seguridad de la información.

Una función de distancia o semejanza determina los casos más semejantes al nuevo problema y las soluciones de los casos recuperados se adaptan para obtener una nueva solución (Martínez Sánchez, Ferreira Lorenzo y García Lorenzo 2008) (Martínez y Pérez , 2010). La recuperación o selección de los ítems de la memoria permanente, semejantes al problema actual se realiza utilizando una función de semejanza.

Se selecciona la función distancia de Manhattan, apoyado en los resultados obtenidos en el modelo de los autores (Zhang y otros, 2011) donde se propone para la semejanza entre atributos, el uso de la misma.

$$d(X_{nr}(O_r), X_{ni}(O_t)) = |X_{nr} - X_{ni}| \quad (1)$$

(1)

La propuesta general de distancia entre los casos es:

$$f(O_r, O_t) = \sum_{i=1}^n \left(\frac{w_i * d(X_{nr}(O_r), X_{ni}(O_t))}{\sum_i^n w_i} \right)$$

(2)

Donde w_i es el peso de importancia de los rasgos, en este caso, el valor a utilizar es el del impacto especificado por los expertos para cada parámetro de la lista de chequeo.

Para aplicar la función de distancia, no se prevé que existan rasgos descriptores sin valores. La aplicación SASGBD, obliga al experto a corregir o adicionar los valores de los rasgos descriptores. Esto puede tener lugar debido a que existen parámetros en las listas de chequeo que obligatoriamente necesitan ser evaluadas por el auditor para llegar a contener un valor.

Los casos se almacenan en una base de datos referencial, posibilitando la recuperación de los casos a través de consultas SQL. Como variables de entrada, son importantes el nombre (GestorBD) y la versión del sistema gestor de bases de datos (VersionBD), premisas indispensables del algoritmo para distinguir los casos necesarios a recuperar según el servidor monitoreado y se requiere realizar el diagnóstico. De ahí la estructura de la Base de Casos que se muestra en la figura 1.

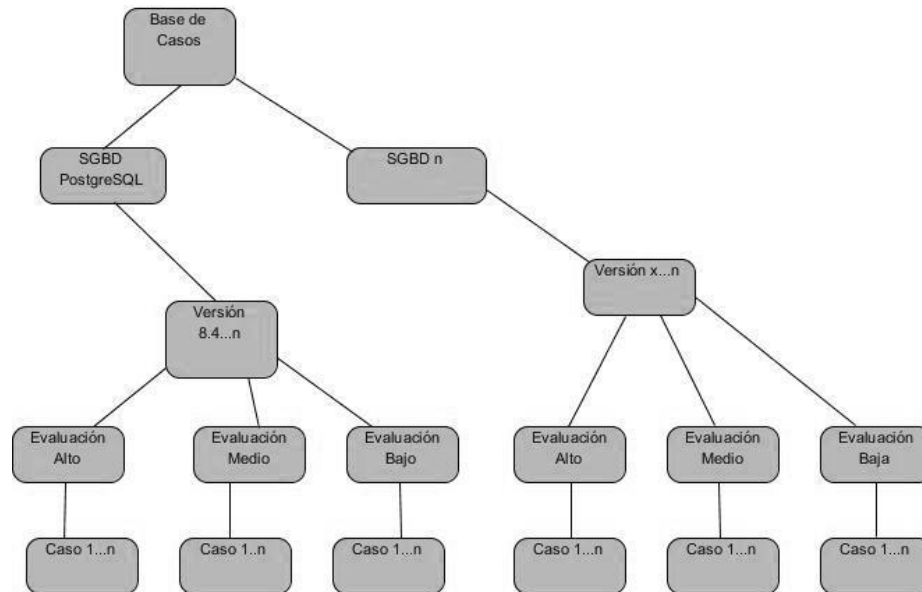


Figura 1: Representación de la estructura de la Base de Casos.

Umbral de semejanza

En el proceso de recuperación influyen otros elementos que cuentan con gran importancia. Uno de ellos es el umbral de semejanza que no es más que el valor mínimo que debe tener un caso para ser considerado a la hora de dar solución a nuevos problemas.

El valor del umbral está relacionado con el comportamiento de los casos en su semejanza, un umbral muy bajo puede conllevar a que se tengan muchos casos en la recuperación y un umbral muy elevado puede conllevar a que no se pueda dar una respuesta en la búsqueda. De ahí que el umbral sea dado por un experto. En caso de que el experto tenga problemas con la definición del umbral, el propio módulo de recuperación debe ser capaz de brindar la posibilidad de calcular el valor umbral de los rasgos.

Para ello, como se muestra en la tabla 1, se construye una matriz cuadrada donde la fila y las columnas están representadas por los casos que se encuentran almacenados en la BC y en la intersección está el valor de semejanza (β).

Tabla 1. Matriz de semejanza entre casos

	Caso 1	Caso 2	Caso 3	Caso n
Caso 1	$\beta(C1,C1)$	$\beta(C1,C2)$	$\beta(C1,C3)$	$\beta(C1,Cn)$
Caso 2	$\beta(C2,C1)$	$\beta(C2,C2)$	$\beta(C2,C3)$	$\beta(C2,Cn)$
Caso 3	$\beta(C3,C1)$	$\beta(C3,C2)$	$\beta(C3,C3)$	$\beta(C3,Cn)$
Caso n	$\beta(Cn,C1)$	$\beta(Cn,C2)$	$\beta(Cn,C3)$	$\beta(Cn,Cn)$

Para permitir que se tengan en cuenta los casos de la BC, la Dra. Natalia Martínez propone (Martínez Sánchez, Ferreira Lorenzo y García Lorenzo, 2008) un cálculo del umbral de semejanza, que viene dada por una media aritmética con los valores de semejanza entre los casos, mediante la siguiente expresión:

$$\beta = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta(O_i, O_j)$$

(3)

Dónde:

m: Número de casos

i: Casos en filas

j: Casos en columnas

$\beta(O_i, O_j)$: Función de semejanza entre los rasgos O_i y O_j .

Lógica difusa

La lógica difusa posibilita manipular las ambigüedades de valores e imprecisiones existentes en las evaluaciones del RSI. Además se selecciona por los resultados obtenidos en el modelo publicado en (Zhang, Lu y Zhang, 2011), donde se resuelve el cálculo del riesgo para la detección temprana de la influencia aviar e interviene el empleo de términos

ambiguos. Teniendo como diferencia que en esta investigación solo se utiliza en ocasiones donde no se puede emplear el RBC.

La función miembro $\mu(R_n)$ definida para los números difusos es la triangular, donde a es el límite inferior y b el límite superior de un valor del RSI en un parámetro Xn. La variable m denota el valor modal del número difuso.

$$\mu(R_n) = \begin{cases} 0, & \text{si } R_n \leq a \\ \frac{R_n - a}{m - a} & \text{si } R_n \in (a, m) \\ \frac{b - R_n}{b - m} & \text{si } R_n \in (m, b) \\ 0, & \text{si } R_n > b \end{cases} \quad (4)$$

Los criterios son evaluados en un conjunto lingüístico de términos lingüísticos de 3 términos, simétricamente y uniformemente distribuidos con la sintaxis representada en la Figura 2 según los criterios de los expertos.

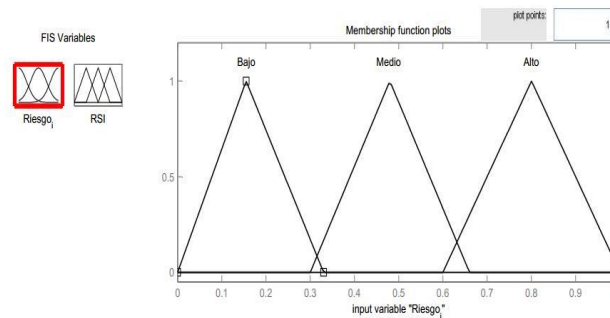


Figura 2. Conjunto de términos lingüísticos.

Proceso de experimentación

Tiempo de respuesta del subsistema

Las técnicas de indexación garantizan una recuperación eficiente de los casos, entendiéndose por eficiencia tanto los aspectos relacionados con el tiempo de recuperación como la garantía de que ningún caso relevante quede fuera de la búsqueda. Los índices son una estructura de datos que se mantiene en memoria y que puede ser recorrida de forma rápida para evitar la búsqueda directa en el contenido de la BC. Normalmente, los índices estarán constituidos por

combinaciones de los descriptores más importantes de los casos para permitir diferenciar unos casos de otros de manera poco costosa (Althoff, 2001).

Con el propósito de evaluar el tiempo de respuesta del sistema, se diseñó un experimento para probar las mejoras del comportamiento del RBC en el proceso de recuperación de los casos más semejantes con respecto al uso o no de técnicas de indexación para mejorar la eficiencia. Se tomó una muestra de 4 casos para realizar el proceso de recuperación con un total de 100 casos existentes en la BC.

Resultados y discusión

Se incorporan técnicas de Inteligencia Artificial, como el RBC y la Lógica difusa al sistema informático SASGBD, el cual evalúa el RSI de los SGBD: PostgreSQL, MySQL, SQL Server y Oracle ya que son los principales gestores hospedados en los servidores de ETECSA. 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 reutilizando los resultados de auditorías de seguridad informática anteriores como experiencias pasadas. Además, con la técnica lógica difusa se logró una mayor exactitud en la evaluación del impacto y del riesgo de seguridad de la información en los SGBD ya que posibilita manejar los términos imprecisos del resultado de la evaluación del RSI.

Resultados

Se implementaron las técnicas RBC y lógica difusa en la solución informática SASGBD la cual permite evaluar el RSI en los SGBD hospedados en los servidores de ETECSA.

Se realizó la evaluación del RSI alcanzado a través del RBC, mostrando el resultado del diagnóstico y a la vez, el caso más semejante encontrado para la presente auditoría, como se muestra en la figura 3. Además, se realizó la evaluación del RSI pero en este caso utilizando la lógica difusa, ya que existen ocasiones en que el RBC no es capaz de evaluar el riesgo porque no existan casos suficientes en la base de casos, como se muestra en la figura 4.

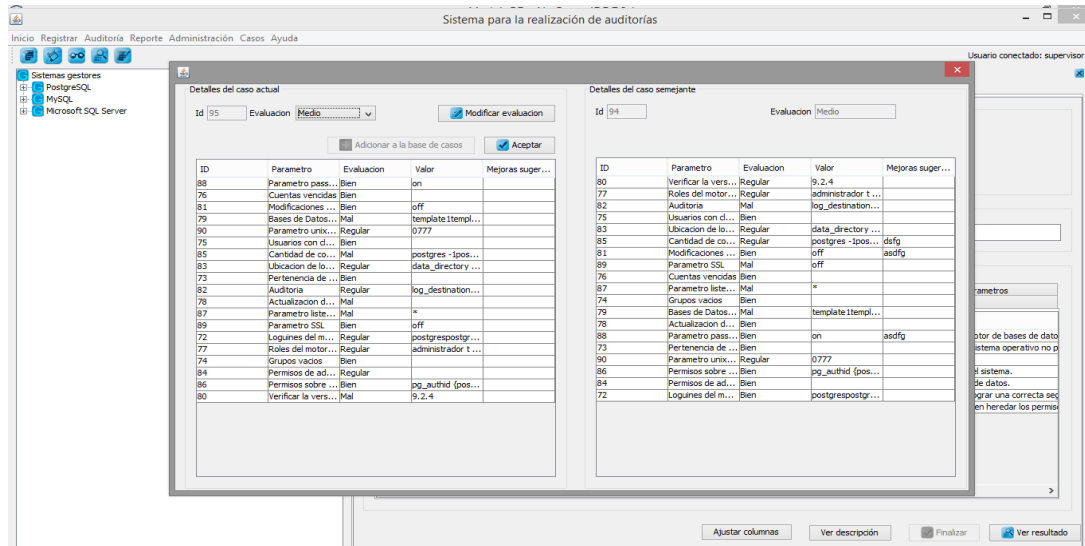


Figura 3. Evaluacion del riesgo usando razonamiento basado en casos.

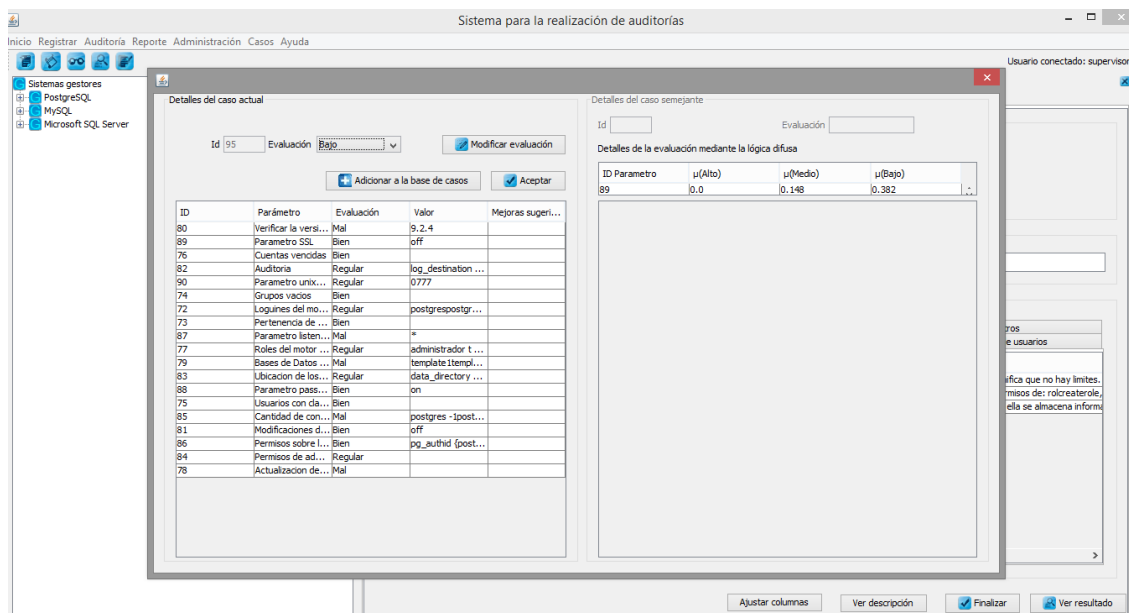


Figura 4. Evaluacion del riesgo usando lógica difusa.

Tabla 2. Resultados experimentales del tiempo de recuperación del algoritmo usando indexación.

Nr o.	Caso nuevo	Caso más cercano	Distancia más cercana	Evaluación del caso más cercano	Caso encontrado	Tiempo de recuperación	Tiempo de recuperación usando indexación	Evaluación final
1	178	160	0.126	Bajo	160	534700	4329	Bajo
2	175	172	0.053	Medio	172	492621	4016	Medio
3	171	172	0.052	Medio	172	442311	3950	Medio
4	162	161	0.063	Medio	161	450002	4063	Medio

Nota: En la Tabla 10 el tiempo de recuperación está expresado en milisegundos (ms).

Nota: El tiempo de respuesta se calculó hallando el promedio de recuperación de los casos.

El tiempo de respuesta del subsistema depende del tiempo de recuperación del SBC, como se pudo observar en la tabla anterior el uso de la indexación mejora en eficiencia el tiempo de respuesta con un promedio de 4758 ms. Con el experimento realizado se puede llegar a la conclusión que la solución desarrollada da cumplimiento al objetivo planteado en la investigación; apoyando la toma de decisiones en cuanto a agilidad del proceso y ayuda a los auditores en el análisis de los riesgos. Según la entrevista realizada a uno de los auditores de ETECSA, no se puede dar un tiempo exacto de cuanto demore un proceso de evaluación actualmente, ya que depende de varios factores. Sin embargo, se puede afirmar que el tiempo de respuesta de la solución desarrollada agiliza considerablemente la evaluación del riesgo, evitando el tiempo y esfuerzo de los auditores a la hora de evaluar las auditorías.

CONCLUSIONES

En el presente trabajo se ha presentado una solución para evaluar el riesgo de seguridad de la información en los SGBD hospedados en los servidores de ETECSA. Se incorporaron técnicas de Inteligencia Artificial como el RBC y la Lógica difusa que brindaron apoyo al proceso de toma de decisiones de manera satisfactoria. En la herramienta informática SASGBD se implementaron las técnicas RBC y Lógica difusa, demostrando la viabilidad de la propuesta.

REFERENCIAS

1. ALTHOFF, K.-D. Case-Based Reasoning. Handbook of Software Engineering and Knowledge Engineering. Institute for Experimental Software Engineering (IESE). [En línea]. 2001.s.n. Disponible en: <http://www.iis.uni-hildesheim.de/files/staff/althoff/Publications/althoff-CBR-update.pdf>.
2. CHOU, T. Information Assurance and Security Technologies for Risk Assessment and Threat Management: Advances. [En línea]. 2011. Disponible en: http://books.google.com.cu/books?id=2zIES3JYNpoC&pg=PA243&dq=cramm+risk+assessment+tool&hl=es&sa=X&ei=S3_oUrCuI8abrAGe9IG4Cw&ved=0CGwQ6AEwCA#v=onepage&q=cramm%20risk%20assessment%20tool&f=false.
3. RAMAKANTH, d. VINOD, k. SQL injection - database attack revolution and prevention. Journal of international commercial law and technology, 6 (4), 224 - 231. [En línea]. 2011. Disponible en: <http://www.jiclt.com/index.php/jiclt/article/view/141/139>.
4. MARTÍNEZ, I.G. y PÉREZ, R.E.B. Un Modelo para la Toma de Decisiones usando Razonamiento Basado en Casos en condiciones de Incertidumbre. Tesis Doctoral, Universidad Central Marta, Abreu Santa Clara, 2010.
MARTÍNEZ SÁNCHEZ, Ms.N., FERREIRA LORENZO, D.G. y GARCÍA LORENZO, D.M.M. El Razonamiento Basado en Casos en el ámbito de la Enseñanza/Aprendizaje. 25-32, s.l. : Revista de Informática Educativa y Medios Audiovisuales. 2008. Vol. V. 1667-8338.
5. PELTIER, T.R. Information Security Risk Analysis (pp. 296). [En línea]. 2001 S.l.: s.n. Disponible en: http://books.google.com.cu/books?id=O0_fO2Xvp98C&pg=PA231&dq=risk+assessment+security+%2B+expert+system&hl=es&sa=X&ei=Tr_qUq3tF6H7yAHc_ICQAg&ved=0CEQQ6AEwAg#v=onepage&q=risk%20assessment%20security%20%2B%20expert%20system&f=false.
6. PALISADE. El futuro en una hoja de trabajo. [En línea]. 2014. [Consulta: 20 octubre 2014]. Disponible en: <http://www.palisade-lta.com/risk/>.
7. QUIGLEY, m. Encyclopedia of information ethics and security Version details - Trove. [En línea]. 2008. [Consulta: 3 noviembre 2014]. Disponible en: <http://trove.nla.gov.au/work/26382737?Selectedversion=NBD42065776>.
8. ZHANG, J., Lu, J. Y 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>.