Pep Lluís Ferrer Gomila · M. Francisca Hinarejos Campos
(editores)

Actas de la XIV Reunión Española sobre Criptología y Seguridad de la Información

Actas de la XIV Reunión Española sobre Criptología y Seguridad de la Información

RECSI XIV

Maó, Menorca, Illes Balears, 26-28 Octubre de 2016

Publicado por:
Departamento de Ciencias Matemáticas e Informática
Universidad de les Illes Balears
Ctra. de Valldemossa, km 7.5. Palma (Illes Balears)
http://recsi16.uib.es

ISBN: 978-84-608-9470-4

Créditos:

Primera edición – Octubre 2016

Organizadores

Colaboradores

Prefacio

Puede afirmarse, sin ningún género de dudas, que las Tecnologías de la Información y de las Comunicaciones (TIC) ya son una realidad de uso cotidiano en casi todos los ámbitos de nuestra sociedad. Solo dos ejemplos de este hecho. Según la Comisión Nacional de los Mercados y la Competencia (CNMC) la facturación del comercio electrónico en España tuvo un incremento en el tercer trimestre del año 2015 del 29,2 %, alcanzando los 5.303 millones de euros. Por otra parte, el 96 % de la población adulta de España posee algún tipo de teléfono móvil, y un 80 % cuenta con algún tipo de smartphone.

Sin embargo, no está tan claro que el despliegue de las TIC venga acompañado siempre del adecuado grado de seguridad. Cada vez es más habitual encontrar noticias en los medios de comunicación sobre problemas de seguridad relacionados con las TIC (fraudes, virus, intrusiones, ataques de denegación de servicio, etc.). Por ello es necesario que nuestra comunidad de ingenieros y científicos, que nos dedicamos al campo de la seguridad, sigamos desarrollando y difundiendo sistemas que permitan alcanzar un desarrollo de las TIC con las protecciones pertinentes, para las múltiples facetas de nuestra vida que quedan afectadas por esas tecnologías.

La Reunión Española de Criptología y Seguridad de la Información (RECSI) es el congreso científico español de referencia en el ámbito de la Seguridad de las TIC. Esta reunión se inició, hace ahora 25 años, en Palma y en octubre de 2016 se celebrará la decimocuarta edición, de nuevo en las Illes Balears, esta vez en Menorca. Las pasadas ediciones tuvieron lugar en Madrid (1992), Barcelona (1994), Valladolid (1996), Torremolinos (1998), Santa Cruz de Tenerife (2000), Oviedo (2002), Leganés (2004), Barcelona (2006), Salamanca (2008), Tarragona (2010), San Sebastián (2012) y Alicante (2014).

Sirva este prefacio para transmitir nuestro más sincero agradecimiento a todos los que han hecho posible que un año más celebremos este encuentro. En primer lugar debemos destacar a los investigadores en temas de seguridad, sin los que no tendría sentido esta reunión. Finalmente, nuestro reconocimiento a los patrocinadores, colaboradores y compañeros que con su esfuerzo han hecho posible este evento.

29 de Junio, 2016
Palma de Mallorca

M. Francisca Hinarejos
Pep Lluís Ferrer

Organización RECSI 2016

Comité Organizador

Ferrer Gomila, Pep Lluís (Universitat de les Illes Balears)
Hinarejos Campos, María Francisca (Universitat de les Illes Balears)
Huguet Rotger, Llorenç (Universitat de les Illes Balears, **Presidente**)
Isern Deyà, Andreu Pere (Universitat de les Illes Balears)
Mut Puigserver, Macià (Universitat de les Illes Balears)
Payeras Capellà, Maria Magdalena (Universitat de les Illes Balears, **Vicepresidenta**)

Comité Científico

Abascal Fuentes, Policarpo (Universidad de Oviedo)
Almenárez mendoza, Florina (Universidad Carlos III de Madrid)
Areitio Bertolín, Javier (Universidad de Deusto)
Borrell Viader, Joan (Universidad Autónoma de Barcelona)
Bras Amorós, Maria (Universidad Rovira i Virgili)
Caballero Gil, Pino (Universidad de La Laguna)
Castellà Roca, Jordi (Universidad Rovira i Virgili)
Climent Coloma, Joan Josep (Universidad de Alicante)
Domingo Ferrer, Josep (Universidad Rovira i Virgili)
Durán Díaz, Raúl (Universidad de Alcalá)
Fernández Arrieta, Miquel (Mondragón Unibertsitatea)
Fernández-Medina Patón, Eduardo (Universidad de Castilla-La Mancha)
Ferrer Gomila, Pep Lluís (Universitat de les Illes Balears)
Fúster Sabater, Amparo (C.S.I.C.)
García Bringas, Pablo (Universidad de Deusto)
García Teodoro, Pedro (Universidad de Granada)
González Vasco, Mª Isabel (Universidad Rey Juan Carlos)
Gutiérrez Gutiérrez, Jaime (Universidad de Cantabria)
Hernández Encinas, Luis (C.S.I.C.)
Hernández Goya, Candelaria (Universidad de La Laguna)
Herrera Joancomartí, Jordi (Universidad Autónoma de Barcelona)
Hinarejos Campos, María Francisca (Universitat de les Illes Balears)
Huguet Rotger, Llorenç (Universitat de les Illes Balears)
Jacob Taquet, Eduardo (Universidad del País Vasco/Euskal Herriko Unibertsitatea)
López Muñoz, Javier (Universidad de Málaga)
Martín del Rey, Ángel (Universidad de Salamanca)
Martínez López,Consuelo (Universidad de Oviedo)
Megías Jiménez, David (Universitat Oberta de Catalunya)
Miret Biosca, Josep Maria (Universidad de Lleida)
Morillo Bosch, Paz (Universidad Politécnica de Catalunya)
Mut Puigserver, Macià (Universitat de les Illes Balears)
Padró Laimon, Carles (Universidad Politécnica de Catalunya)
Payeras Capellà, Maria Magdalena (Universitat de les Illes Balears)
Peinado Domínguez, Alberto (Universidad de Málaga)
Ramió Aguirre, Jorge (Universidad Politécnica de Madrid)
Ribagorda Garnacho, Arturo (Universidad Carlos III de Madrid)
Rifà Coma, Josep (Universidad Autónoma de Barcelona)
Sáez Moreno, Germán (Universidad Politécnica de Catalunya)
Salazar Riaño, José Luis (Universidad de Zaragoza)
Sánchez Ávila, Carmen (Universidad Politécnica de Madrid)
Sebé Feixa, Francesc (Universidad de Lleida)
Soriano Ibáñez, Miguel (Universidad Politécnica de Catalunya)
Villar Santos, Jorge (Universidad Politécnica de Catalunya)
Zamora Gómez, Antonio (Universidad de Alicante)
Zurutuza, Urko (Mondragón Unibertsitatea)

Lista de Contribuciones

Seguridad en Redes de Nueva Generación: Arquitectura para la Gestión de Incidencias 1

Lorena Isabel Barona Lopez, Jorge Maestre Vidal, Angel Leonardo Valdivieso Caraguay, Marco Antonio Sotelo Monge y Luis Javier García Villalba

Autenticación implícita eficiente y con privacidad 7

Blanco-Justicia y Josep Domingo-Ferrer

MCSEC: Una plataforma para mobile crowd sensing seguro con protocolos oportunistas 13

Carlos Borrego, Joan Borrell, Marc Dalmau, Sergi Delgado-Segura, Angela Fabregues, Gerard Garcia-Vandellos, Carlos Lacambra, Ramon Marti, Guillermo Navarro-Arribas, Cristina Pérez-Solà, Remei Ridorsa y Sergi Robles

A Publicly Verifiable Counter-Based Loyalty System 18

Núria Busom Figueres, Francesc Sebe y Magda Valls

Modelo de madurez de cultura organizacional de seguridad de la información. Una visión desde el pensamiento sistémico-cibernético 24

Jeimy Cano

Modelos lineales basados en CA para las secuencias auto-shrinking 30

Sara D. Cardell y Amparo Fúster-Sabater

Control Seguro y Anónimo para el Acceso de Vehículos a Zonas Urbanas de Tráfico Restringido . . 36

Jordi Castellà-Roca, Macià Mut Puigserver, M. Magdalena Payeras Capellà, Alexandre Viejo y Carles Anglès-Tafalla

Utilización de un generador con distribución gaussiana para aumentar la complejidad lineal de las m-secuencias 42

Guillermo Cotrina, Alberto Peinado y Andrés Ortiz

Survey of network based attacks to the Bitcoin P2P network 47

Sergi Delgado Segura, Cristina Pérez Solà, Guillermo Navarro-Arribas, Jordi Herrera y Joan Borrell

Cryptocurrency P2P networks: a comparison analysis 52

Joan Antoni Donet Donet y Jordi Herrera-Joancomartí

Generación de primos demostrables: implementación y resultados 58

Raúl Durán Díaz, Víctor Gayoso Martínez y Luis Hernández Encinas

Los spammers no piensan: usando reconocimiento de personalidad para el filtrado de spam en mensajes cortos 64

Enaitz Ezpeleta, Urko Zurutuza y José María Gómez Hidalgo

Códigos de fingerprinting binarios. Nuevos paradigmas de identificación 70

Marcel Fernandez, Elena Egorova y Grigory Kabatyanskiy

Comparación de métodos de diagnóstico de anomalías en monitorización estadística multivariante de redes 75

Noemí Marta Fuentes García, José Camacho y Gabriel Maciá-Fernández

Anomaly detection in smart city parking data: A case study 81

Victor Garcia-Font, Carles Garrigues y Helena Rifà-Pous
Detección Colaborativa Multi-nivel de Anomalías en Entornos Móviles 86
Pedro García Teodoro y José Camacho Páez
On the Linear Complexity and k-Error Linear Complexity over $\mathbb{F}_{p^m}$ of Sequences of Period tp^v . . 91
Domingo Gomez-Perez
Evaluación de librerías criptográficas externas de Android 96
David Gonzalez, Oscar Esparza, Jose L. Muñoz, Jorge Mata y Juanjo Alins
Códigos Grupos No-Abelianos 101
Santos Gonzalez y Consuelo Martinez
Detección de Black holes en VANETs basada en auditoría a nodos 105
Jose Grimaldo, Ruben Martinez-Vidal y Ramon Marti
Generación distribuida y verificable de códigos de retorno para voto electrónico 111
Sandra Guasch, Víctor Mateu y Magda Valls
Distinguiendo entre perturbaciones de proceso e intrusiones en sistemas de control: caso de estudio con el proceso Tennessee-Eastman 117
Mikel Iturbe, José Camacho, Iñaki Garitano, Urko Zurutuza y Roberto Uribeetxeberria
Manifold alignment approach to cover source mismatch in steganalysis 123
Daniel Lerch-Hostalot y David Megías
Impacto de los ataques PUE y Bizantino en la conectividad de redes ad hoc de radio cognitiva . . 129
Olga León y Juan Hernández
Modelo epidemiológico para la propagación del jamming aleatorio en redes de sensores inalámbricos 134
Miguel López, Alberto Peinado y Andrés Ortiz
Sistema de Reconocimiento de Dinámicas de Firmas Manuscritas en Dispositivos Móviles 140
Andrea Alexandra Martin, Jorge Maestre Vidal, Luis Javier García Villalba, Jordi Iñigo y David Ruana
Un modelo para simular la propagación de código malicioso en redes inalámbricas 146
Angel Martin Del Rey, José Diamantino Hernández Guillén y Gerardo Rodriguez Sanchez
Mejora de la Gestión de Nodos Revocados en Redes Vehiculares 151
Francisco Martín-Fernández, Pino Caballero-Gil y Cándido Caballero-Gil
Ocultando la distribución de probabilidad en criptosistemas ordenables 157
Santi Martínez, Daniel Sadornil y Josep Conde
Uso de técnicas Big Data para evaluar seguridad 163
Julio Moreno, Manuel Serrano y Eduardo Fernandez-Medina
Arquitectura funcional para la cadena de custodia digital en objetos de la IoT 168
Ana Nieto, Rodrigo Román y Javier López
Nuevas nociones de seguridad y transformaciones genéricas para criptosistemas de recifrado delegado 174
David Nuñez, Isaac Agudo y Javier Lopez
Herramientas gráficas de la criptografía caótica para el análisis de la calidad de secuencias pseudo-aleatorias 180
Amalia Beatriz Orúe López, Amparo Fúster Sabater, Verónica Fernández Marmol, Fausto Montoya Vitini, Luis Hernández Encinas y Agustin Martín Muñoz
A SPAM Filtering Scenario Using Constrained Bit-Parallel Approximate Search 186
Slobodan Petrovic
Una Herramienta para el Control de la Privacidad contra el Rastreo en la Web 191
Jagdish Prasad Achara, Javier Parra-Arnau y Claude Castelluccia

Seguridad en Redes Definidas por Software: Desafíos y Soluciones 197
Jesús Antonio Puente Fernandez, Angel Leonardo Valdivieso Caraguay y Luis Javier García Villalba
Peer-to-peer Content Distribution Using Anonymous Fingerprinting - Proof of Concept 203
Amna Qureshi, Jordi Casas-Roma, David Megías y Helena Rifà-Pous
Evolución y nuevos desafios de privacidad en la Internet de las Cosas 209
Ruben Rios y Javier Lopez
Una Propuesta para la Mejora de la Seguridad y Eficiencia en la Gestión de Pacientes a través de m-Health 214
Alexandra Rivero-García, Candelaria Hernández Goya, Iván Santos-González y Pino Caballero-Gil
Protección de la privacidad en trayectorias para estudiar la propagación de epidemias 220
Cristina Romero-Tris, Joan Melia y David Megias
Theia: Una Herramienta para el Análisis Forense de Imágenes Digitales de Dispositivos Móviles . . 226
Jocelin Rosales Corripio, Anissa El-Khattabi, Ana Lucila Sandoval Orozco y Luis Javier García Villalba
Uso de Características en la Identificación de la Fuente de Imágenes de Dispositivos Móviles 232
Jocelin Rosales Corripio, Ana Lucila Sandoval Orozco y Luis Javier García Villalba
Multiplicación escalar en dos familias de curvas elípticas usando endomorfismos 238
Daniel Sadornil, Josep M. Miret Biosca y Juan Tena
Sistema de Comunicaciones con Autenticación Distribuida para Situaciones de Emergencia 242
Iván Santos-González, Pino Caballero-Gil, Jezabel Molina-Gil y Alexandra Rivero-García
Group Key Establishment: Compilers for Deniability 247
Rainer Steinwandt y Adriana Suárez Corona
HSTS y HPKP: Un estudio cuantitativo y cualitativo de su implementación en servidores 252
Carmen Torrano, Sergio de Los Santos y Antonio Guzmán
Uso actual de la criptografía sobre curva elíptica 258
Manuel Trujillo Vanrell, Macià Mut Puigserver, Magdalena Payeras-Capellà, Jordi Castellà-Roca y Llorenç Huguet
Hacia la optimización de un generador pseudoaleatorio matricial 264
Antonio Zamora Gómez, Rafael Alvarez y Francisco-Miguel Martínez
Uso de NFC para Gestionar con Seguridad el Equipaje 270
Néstor Álvarez-Díaz y Pino Caballero-Gil

Contenido de los artículos presentados en el congreso

Uso de técnicas Big Data para evaluar seguridad

Julio Moreno
Grupo de investigación GSyA
Universidad de
Castilla-La Mancha (UCLM)
Email: julio.moreno@uclm.es

Manuel Serrano
Grupo de investigación Alarcos
Universidad de
Castilla-La Mancha (UCLM)
Email: manuel.serrano@uclm.es

Eduardo Fernández-Medina
Grupo de investigación GSyA
Universidad de
Castilla-La Mancha (UCLM)
Email: eduardo.fdezmedina@uclm.es

***Resumen*—Los datos cada vez son más importantes para cualquier tipo de organización, la información extraída de ellos, les ayuda a alcanzar sus objetivos de negocio y a la toma de decisiones estratégicas. La seguridad de dicho recurso debe ser una prioridad. En la actualidad, cada vez se generan más datos lo que provoca que las técnicas tradicionales de análisis de datos no puedan gestionarlos de forma eficiente. Una solución a este problema es el uso de Big Data.**
Sus principales características la convierten en una técnica notable a la hora de analizar la seguridad de cualquier tipo de sistema. Así, el objetivo de este artículo es explicar una forma de aprovechar la versatilidad de Big Data para la seguridad, implementando una serie de algoritmos, que hacen uso del paradigma de programación MapReduce, con el propósito de evaluar problemas relacionados con la seguridad de sistemas de bases de datos relacionales.

***Palabras clave*—Bases de datos (*Database Systems*), Big Data, MapReduce, Seguridad de la información (*Information Security*)**

I. Introducción

Es una realidad que vivimos en la era del Big Data. Los datos tienen cada vez más protagonismo en el día a día de cualquier empresa u organización, no sólo en el ámbito de la tecnología sino también en organizaciones de gobierno, sanitarias, de educación o del sector industrial. Por ello, estos datos se han convertido en algo vital para las organizaciones, ya que, la información extraída de ellos les ayuda a la hora de alcanzar sus objetivos de negocio y en la toma de decisiones [1]. Así, cada vez generamos una mayor cantidad de datos por día. Se estima que de todos los datos generados por el ser humano el 90 por ciento han sido creados en los últimos años, por ejemplo, en 2003 se crearon 5 exabytes de datos, mientras que en la actualidad esa cantidad de datos se genera en dos días [2].

Esta tendencia de incrementar el volumen y detalle de los datos almacenados por las organizaciones no va a cambiar en un futuro próximo. Al contrario, el aumento del uso de las redes sociales, los archivos multimedia, y el inminente Internet de las Cosas (*IoT*) producen un inmenso flujo de datos en el que la mayoría de estos datos suelen tener un formato no estructurado [3]. Las organizaciones quieren obtener beneficio de este volumen y variedad de datos, pero las técnicas tradicionales de análisis de datos no son suficientes para abordar esta problemática [4]. Por ello, para analizar y entender mejor estos datos, y así, obtener información valiosa para la organización, nació un nuevo paradigma de análisis: Big Data [5].

Con cada nueva tecnología disruptiva surgen nuevos problemas. Con Big Data, estos problemas no sólo están relacionados con el tamaño de los datos a tratar, sino que también aumentan y se crean nuevos desafíos relacionados con la calidad de los datos, con la privacidad o con la seguridad [6].

En nuestro caso, nos centraremos en aquellos problemas relacionados con la seguridad. Existen principalmente dos formas de tratar la seguridad en Big Data: utilizar Big Data para lograr seguridad en un sistema o asegurar el sistema Big Data en sí mismo. En este artículo, se explica una forma de obtener valor de la potencia de Big Data para evaluar la seguridad de un sistema de bases de datos de tipo relacional. Para ello, se han implementado una serie de algoritmos usando el paradigma de programación MapReduce, habitual en Big Data, cuyo propósito es evaluar problemas vinculados con la seguridad en un sistema de bases de datos relacional. Dichos problemas se basan en distintos controles y recomendaciones realizados por los principales estándares y metodologías internacionales en el campo de la seguridad, como la familia de normas ISO/IEC 27000 o el marco internacional COBIT 5 para seguridad de la información.

Por tanto, el objetivo de este artículo es por un lado destacar las posibilidades que nos brinda el uso de Big Data en cuanto a la seguridad, y por otro, mostrar una serie de ejemplos sobre cómo aplicar técnicas típicas de Big Data para evaluar la seguridad de una base de datos. Este artículo tiene la siguiente estructura: primero se hará una introducción al tema de Big Data (incluyendo una explicación del paradigma de programación MapReduce), seguido por la explicación de las evaluaciones de seguridad que se han implementado y, finalmente, una sección en la que se abordarán las conclusiones y trabajo futuro.

II. Big Data

Big Data es un paradigma que permite el análisis y la gestión de una mayor cantidad de datos que las técnicas de procesamiento de datos tradicionales [7]. Big Data supone un cambio respecto a las técnicas tradicionales en, principalmente, tres características: la cantidad de datos (volumen), el alto ratio de generación de datos y su transmisión (velocidad) y los tipos de datos estructurados y no estructurados que puede manejar (variedad) [8]. Estas propiedades son conocidas

como las tres V's básicas de Big Data. Algunos autores han añadido nuevas características a este grupo inicial, como la variabilidad, la veracidad o el valor de los datos [9].
La forma más extendida de hacer uso de la tecnología Big Data es mediante Apache Hadoop [10], un entorno desarrollado por Apache que permite el procesamiento paralelo de grandes conjuntos de datos mediante clusters de ordenadores. Está diseñado para ser escalable desde un único servidor a miles, cada uno de los cuales, ofrece computación y almacenamiento local. Para ello, Hadoop dispone de su propio sistema de almacenamiento distribuido denominado HDFS, el cual, almacena los datos en diferentes servidores y distintas funciones, como el NameNode para almacenar los metadatos o los DataNodes utilizados para almacenar los datos generados por las aplicaciones [11]. Aunque, la principal característica de Hadoop es ser una implementación de código abierto que permite hacer uso del paradigma de programación MapReduce [12].
MapReduce es un entorno de desarrollo cuyo principal objetivo es procesar y generar grandes conjuntos de datos. Para ello, MapReduce define dos tipos de funciones distintas [13]:

- La función **Map** que se encarga de procesar un par clave/valor necesario para crear un conjunto de pares intermedios.
- La función **Reduce** que se encarga de procesar los valores intermedios generados por la función Map y fusionarlos para generar la solución.

En la Figura 1, se puede observar un ejemplo de cómo generalmente trabaja un algoritmo MapReduce. Como se puede ver, existe una etapa intermedia denominada *Shuffle* cuyo objetivo es ordenar las salidas de la función del Map para que las reciba la función Reduce. Esta capacidad de paralelizar el procesamiento de los datos en un entorno altamente distribuido, como es Hadoop, junto con su gran flexibilidad para analizar tanto datos estructurados como no estructurados, convierten a este paradigma de programación en una herramienta muy potente y a tener en consideración cuando el análisis se va a realizar sobre grandes conjuntos de datos.

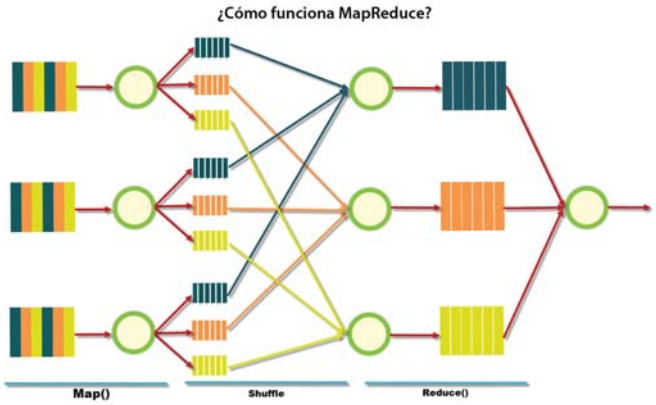

Figura 1. Funcionamiento de MapReduce. Adaptado de [14].

Normalmente, el uso de Big Data se relaciona con empresas de Internet como Google o Facebook que gestionan una inmensa cantidad de datos del orden petabytes por mes. Pero la realidad es que el uso de Big Data no se encuentra limitado a este tipo de empresa, sino que organizaciones de ámbito tan diverso como industrial, sanitario, meteorológico o incluso el control de tráfico, también se aprovechan de los beneficios de utilizar esta tecnología [15].
En cuanto al ámbito de la seguridad, existen numerosas herramientas que aprovechan las ventajas de Big Data desde sistemas que investigan el tráfico de red para detectar ataques [16], hasta sistemas cuyo objetivo es asegurar la integridad de los datos [17]. En nuestro caso, hemos definido una serie de evaluaciones sobre seguridad en sistemas de bases de datos relacionales implementados mediante el paradigma MapReduce, que nos permitirán hacer un primer acercamiento a la utilidad que se puede obtener del uso de de este tipo de técnicas en el campo de la seguridad.

III. Evaluaciones implementadas

Como se ha comentado en la introducción, el objetivo de este artículo es difundir una forma de utilizar Big Data para crear evaluaciones de seguridad en sistemas de bases de datos relacional. Así, en este apartado, se explicarán cuáles son las evaluaciones implementadas, además se definirá la motivación detrás de la elección de dichas evaluaciones, la cual, se basa en un estudio de los principales estándares y recomendaciones sobre seguridad de la información. Además, se explicará a modo de ejemplo cómo se ha implementado una de las evaluaciones, incluyendo el código empleado.
Para llevar a cabo estas evaluaciones es necesario analizar las tablas de un sistema de bases de datos o sus archivos de log, los cuales pueden ser enormes.

III-A. Evaluación 1 - Cifrado

Esta evaluación comprueba si la columna que introduce el usuario se encuentra encriptada. Como resultado de realizar esta evaluación, se obtiene el porcentaje de registros cifrados. Esta evaluación surge de la necesidad expresada dentro de la norma ISO/IEC 27002, sobre tener una política de cifrado de los datos sensibles o relevantes para así, protegerlos.
Para ello, se ha decidido tener dos diccionarios en el sistema: uno en castellano y otro en inglés. Una vez incluidos estos diccionarios, se crea el algoritmo MapReduce que compara cada una de las palabras de todo los registros contenidos en la columna con los diccionarios. Para obtener la columna de la base de datos que servirá de entrada al algoritmo, es necesario almacenar dicha columna en el sistema HDFS, para ello, se utiliza la herramienta Apache Sqoop. Si una de las palabras del registro no es reconocida como una palabra válida, se considerará que dicho registro se encuentra cifrado. En caso contrario, el registro se considerará como no cifrado.

III-B. Evaluación 2 - Permisos de usuario

Esta evaluación fue desarrollada con el objetivo de evaluar los permisos de los usuarios. Puede ser considerada como una implementación del principio de menor privilegio expresado en COBIT 5.
Para alcanzar ese propósito se implementó un algoritmo MapReduce cuya entrada es la tabla donde se almacenan los datos con los permisos de los usuarios. Por ejemplo, en una base de datos de tipo MySQL dicha tabla es USER_PRIVILEGES y se encuentra localizada con los metadatos de la base de datos. Para almacenarla en el HDFS hay que hacer uso de la herramienta Apache Sqoop, tal como se ha explicado en la anterior evaluación.
Una vez se ejecuta el algoritmo, como salida se obtiene el porcentaje de usuarios que tienen el permiso que ha sido especificado por el usuario. Una gran cantidad de usuarios con un privilegio particular puede significar una vulnerabilidad en el sistema. En la Figura 2, se puede ver un diagrama de proceso de cómo se realiza la evaluación 2.

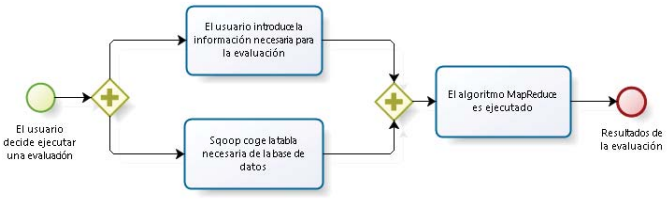

Figura 2. Diagrama de proceso de la evaluación 2.

III-C. Evaluación 3 - Usuarios eliminados

El propósito detrás de la implementación de esta evaluación es alcanzar un objetivo en concreto: detectar si a los usuarios que ya no pertenecen al sistema se les han revocado los diferentes permisos que tenían. Por ejemplo, se puede dar la situación de que un trabajador sea despedido, pero sus permisos de acceso o modificación no han sido eliminados del sistema. Según el estándar ISO/IEC 27002, tener antiguos empleados con privilegios expone al sistema a ataques de dichas personas generando una gran amenaza.
Se incluye en el Listado 1 la función Map utilizada, y por otro lado, en el Listado 2 la función Reduce. Como se puede observar ambas han sido implementadas usando el lenguaje de programación Python.

Listado 1. Función Map de la evaluación 3.

```
#!/usr/bin/env python

import sys

for line in sys.stdin:
        line = line.strip()
        words = line.split(',')
        print '%s\t%s' %(words[0], words[2])
```

Para ejecutar esta evaluación es también necesario que el usuario añada una lista con los usuarios que ya no pertenecen al sistema, además como entrada también tendrá la tabla con los permisos de los usuarios (USER_PRIVILEGES) que será obtenida usando la herramienta Apache Sqoop, como se explicó en la primera evaluación.

Listado 2. Función Reduce de la evaluación 3.

```
#!/usr/bin/env python

from operator import itemgetter
import sys

f = open(sys.argv[1])
users = f.read()
userActual = None
contador = 0

total = len(users.split('\n'))

for line in sys.stdin:
        line = line.strip()
        user, permiso = line.split("\t")

        if users.find(user) != -1:
                if (user != userActual):
                        userActual = user
                        contador = contador + 1

f.close()

print "Usuarios con permisos no revocados:
%d\nUsuarios sin permisos:%d" %
(contador, (total-contador))
```

Finalmente, la función Reduce compara cada uno de los usuarios de la tabla USERS_PRIVILEGES, con la lista introducida con los usuarios que no deben tener permisos en el sistema. En caso de que se localice un usuario que no deba tener permisos, se aumentará el contador y se notificará.

III-D. Evaluación 4 - Acceso en horario laboral

En general, los usuarios sólo deberían acceder al sistema cuando estén trabajando. Por ello, si un usuario accede de forma repetida al sistema cuando no se encuentra en su horario laboral, puede ser indicativo de que está realizando una actividad inapropiada. Así, el objetivo de esta evaluación será comprobar la cantidad de veces que esto ha ocurrido y cuáles son los usuarios que lo han realizado.
Para alcanzar este objetivo, será necesario que el usuario introduzca en el sistema dos archivos: por un lado, el archivo de log será necesario para observar en qué hora acceden al sistema los usuarios, y por otro lado, se requiere conocer cuál es el horario normal de cada usuario. Una vez se tienen las dos entradas, el algoritmo MapReduce compara las horas de cada uno de los accesos al sistema por parte de los usuarios con el horario laboral indicado en el archivo. Si un usuario ha accedido fuera de su horario laboral se contabiliza. Como

resultado de esta evaluación se obtiene una lista de los usuarios que acceden fuera de horario, la cual, se encuentra ordenada en función de la cantidad de veces que haya sucedido esta incidencia.
En la Figura 3, se puede observar un ejemplo de las distintas fases que componen esta evaluación y las distintas salidas parciales que van generando. En el ejemplo, como entrada se tiene una muestra de archivo de log simplificado.

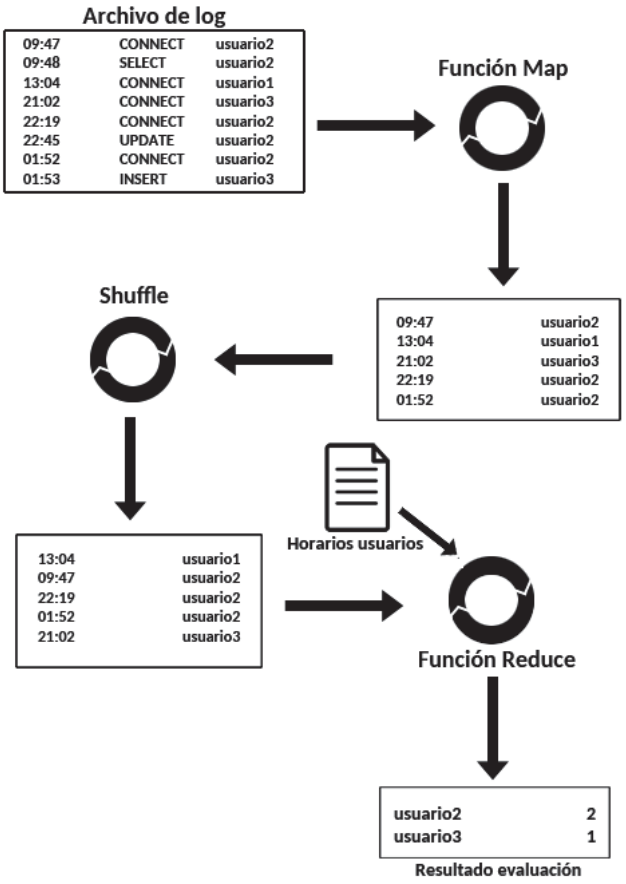

Figura 3. Funcionamiento de la evaluación 4.

III-E. Evaluación 5 - Accesos erróneos

El principal propósito de esta evaluación es comprobar la cantidad de accesos erróneos realizados por cada usuario. La motivación detrás de esta evaluación es que si algún usuario ha tenido una gran cantidad de fallos, puede ser indicativo de que se ha producido un intento de ataque al sistema. Este tipo de ataques puede haber sido perpetrado tanto por el propio usuario, como por un usuario externo. Esta evaluación se encuentra justificada por la recomendación hecha dentro de la norma ISO/IEC 15408 sobre la necesidad de implementar controles apropiados de autenticación e identificación para los usuarios del sistema.
Para alcanzar este objetivo será necesario volver a hacer uso del archivo log generado por el sistema de bases de datos. Una vez el log ha sido añadido al sistema de ficheros HDFS de Hadoop, este será analizado mediante un algoritmo MapReduce. Como resultado de ejecutar esta evaluación se obtendrá una lista ordenada por el número de veces que cada usuario ha fallado al intentar autenticarse en el sistema de bases de datos.
Finalmente, en la Tabla I se muestra un resumen de las evaluaciones realizadas incluyendo su nombre, su propósito y su motivación por la cuál ha sido implementada.

Tabla I
EVALUACIONES DESARROLLADAS

Id	Evaluación	Objetivo	Motivación
EV 1	Evaluar el cifrado de una columna en una tabla de una base de datos.	Comprobar que los campos de una columna se encuentran cifrados.	ISO/IEC 27002 explica la necesidad de tener una política para que los datos se encuentren cifrados.
EV 2	Evaluar los permisos de los usuarios	Comprobar que los usuarios tienen los permisos adecuados.	COBIT 5 insta a la aplicación del principio de menor privilegio para asegurar la seguridad de un sistema.
EV 3	Evaluar que los usuarios eliminados no disponen de privilegios.	Comprobar que a los usuarios, que ya no son parte del sistema, les han revocado los permisos.	ISO/IEC 27002 insta a las organizaciones a proteger sus intereses cuando los usuarios ya no pertenecen al sistema.
EV 4	Evaluar que un número de usuarios sólo acceden al sistema durante su horario laboral.	Comprobar que los usuarios sólo acceden al sistema cuando lo tienen permitido.	Los ataques internos realizados por usuarios que forman o formaban parte del sistema es una de las principales fuentes de amenaza.
EV 5	Evaluar la cantidad de accesos fallidos que han ocurrido.	Comprobar cuántos accesos fallidos ha realizado cada usuario.	ISO/IEC 15504 explica la necesidad de tener controles de autenticación e identificación para los usuarios.

IV. CONCLUSIÓN Y TRABAJO FUTURO

En este artículo se ha descrito cómo se puede aprovechar las características de la tecnología Big Data en el campo de la seguridad. Se ha especificado cómo se implementaron cinco evaluaciones distintas sobre la seguridad de un sistema de bases de datos, siguiendo el paradigma distribuido de programación MapReduce.
Para implementar estas evaluaciones se ha diseñado y construido un entorno de Big Data, utilizando como base la tecnología Apache Hadoop. Alrededor del mismo se han instalado los diferentes módulos necesarios, como por ejemplo Apache Sqoop.
También cabe destacar el hecho de que el sistema fue realizado teniendo en mente que sea fácilmente extensible con nuevas evaluaciones. Por ello, como trabajo futuro nos replanteamos la idea de implementar nuevas evaluaciones y funciones dentro del sistema que aumenten sus capacidades.

Finalmente, también nos gustaría probar nuestro prototipo en un *cluster* mayor que nos permita probar diferentes composiciones de los nodos, para observar el distinto rendimiento obtenido al modificar dichos parámetros, y obtener la configuración más eficiente.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el proyecto SEQUOIA (Ministerio de Economía y Competitividad y Fondo Europeo de Desarrollo Regional FEDER, TIN2015-63502-C3-1-R) y por el proyecto SERENIDAD (Consejería de Educación, Ciencia y Cultura de la Junta de Comunidades de Castilla La Mancha, y Fondo Europeo de Desarrollo Regional FEDER, PEII-2014-045-P).

REFERENCIAS

[1] M. S. Kulkarni and D. S. Urolagin, "Review of attacks on databases and database security techniques," *International Journal of Emerging Technology and Advanced Engineering, ISSN*, pp. 2250–2459, 2012.

[2] S. Sagiroglu and D. Sinanc, "Big data: A review," *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, pp. 42–47, May 2013.

[3] I. Hashem, I. Yaqoob, N. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of "big data.ºn cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98–115, 2015.

[4] S. Sharma, "Rise of Big Data and related issues," *2015 Annual IEEE India Conference (INDICON)*, pp. 1–6, Dec. 2015.

[5] R. Eynon, "The rise of Big Data: What does it mean for education, technology, and media research?," *Learning, Media and Technology*, vol. 38, no. 3, pp. 237–240, 2013.

[6] V. Rijmenam, *Think Bigger: Developing a Successful Big Data Strategy for Your Business*. New York: Amacom, May 2014.

[7] X. Meng and X. Ci, "Big data management: Concepts, techniques and challenges," *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*, vol. 50, no. 1, pp. 146–169, 2013.

[8] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[9] M. Ali-ud-din Khan, M. F. Uddin, and N. Gupta, "Seven V's of Big Data understanding Big Data to extract value," in *American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the*, pp. 1–5, IEEE, 2014.

[10] "Apache Hadoop."

[11] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop distributed file system," 2010.

[12] D. Jiang, B. Ooi, L. Shi, and S. Wu, "The performance of mapreduce: An indepth study," *Proceedings of the VLDB Endowment*, vol. 3, no. 1, pp. 472–483, 2010.

[13] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2004.

[14] Pinal Dave, "Big Data - Buzz Words: What is MapReduce - Day 7 of 21," Oct. 2013.

[15] V. Mayer-Schonberger and K. Cukier, *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.

[16] L. Lan and L. Jun, "Some special issues of network security monitoring on big data environments," *Proceedings - 2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing, DASC 2013*, pp. 10–15, 2013. cited By 1.

[17] D. Gligoroski, H. Mihajloska, S. Samardjiska, H. Jacobsen, R. Jensen, and M. El-Hadedy, "-cipher: Authenticated encryption for big data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8788, pp. 110–128, 2014. cited By 0.