# Ciencia de Datos en Salud: Minería de Procesos con Preservación de Privacidad de Datos Médicos

Heidy M. Marin-Castro[1], Héctor A. De la Fuente-Anaya[2], Miguel Morales-Sandoval[2], Ana B. Ríos-Alvarado[3], and Tania Y.Guerrero-Meléndez[3]

[1] Conacyt - Facultad de Ingeniería y Ciencias, Universidad Autónoma de Tamaulipas, México
hmarisol@docentes.uat.edu.mx
[2] Cinvestav Unidad Tamaulipas, Cd. Victoria, Tamps, México
{hector.delafuente,miguel.morales}@cinvestav.mx
[3] Facultad de Ingeniería y Ciencias, Universidad Autónoma de Tamaulipas, México
{arios,tyguerre}@docentes.uat.edu.mx

**Resumen** Diariamente, los sistemas de información generan una gran cantidad de registros de eventos a partir de la ejecución de procesos. Esta situación ha incentivado gran interés por parte de las organizaciones por realizar análisis de estos datos a fin de generar conocimiento que apoye en la toma de decisiones. La Minería de Procesos, en combinación con la Preservación de Privacidad de Datos, es una disciplina orientada a descubrir, monitorear y mejorar el desempeño de los procesos de negocio, garantizando preservar la confidencialidad de los datos sensibles o personales de eventos producidos por la ejecución de un proceso a partir de un sistema de información. La gran mayoria de los procesos clínicos relacionados con el diagnóstico, tratamiento y organización de personal de la salud y de los pacientes requieren de algoritmos que puedan realizar análisis y estudio de sus datos y procesos asegurando la privacidad y confidencialidad de estos mientras son usados. En este capítulo se describen algunas de las características más relevantes de los procesos en el dominio de salud, así como la importancia de la Minería de Procesos como un servicio, y se presenta una estrategia de confidencialidad para proveer privacidad en los datos de las bitácoras de eventos relacionados con un proceso clínico. Una de las ventajas de la estrategia propuesta es que permite mantener la utilidad de los datos para las tareas de Minería de Procesos sin que exista alguna perdida de información o la posibilidad de revelar a terceros información que se considera confidencial.

**Palabras clave:** Minería de Proceso · Bitácora de Eventos · Modelo de Proceso · Confidencialidad · Cifrado · Utilidad de Datos

## 1. Introducción

En nuestra sociedad actual, el tema de protección y privacidad de datos ha ganado mucha atención en los últimos años debido a los frecuentes ataques

cibernéticos o filtraciones de datos que comúnmente se presentan contra los sistemas de información y las regulaciones relacionadas con Reglamento General de Protección de Datos (GDPR) de Europa [7]. En general, la privacidad puede ser descrita como el derecho de las personas a controlar cómo se recopilan, utilizan y/o divulgan sus datos personales a otros individuos, organizaciones o gobiernos [21]. La privacidad de los datos se ha vuelto mucho más crítica, especialmente para aquellas empresas que trabajan con datos sensibles, como las organizaciones en el sector salud. Los proveedores de atención médica deben asegurarse de administrar adecuadamente los datos de los pacientes para crear una cultura de confianza y transparencia mientras cumplen con las estrictas normas legales y de privacidad de datos. La privacidad de los datos en el cuidado de la salud está en constante evolución, con leyes y regulaciones continuamente actualizadas. De esta forma, los pacientes obtienen la privacidad de datos que esperan. Mientras que ha habido mucha investigación sobre lo que constituye la privacidad de datos y su importancia en los sistemas de información en el sector salud [14], [3], [13], [18], existe una clara brecha en la investigación sobre privacidad en el campo de la Minería de Procesos (MP), la cual se centra en el estudio de los procesos de negocio descritos como un conjunto de actividades interrelacionadas y desempeñadas por un grupo de participantes para lograr un objetivo de negocio. Esta disciplina conecta técnicas de la Ciencia de Datos y de la Ciencia de Procesos para llevar a cabo tareas de descubrimiento, verificación de la conformidad y mejora de los procesos de negocio a partir de extraer conocimiento de colecciones de eventos llamadas bitácoras de eventos [1].

La MP ha evolucionado a través de los años. Al inicio, los algoritmos y herramientas de MP fueron desarrolladas por grupos de investigación [2] y paulatinamente han sido utilizados por la industria a través del análisis y estudio de casos y proyectos. Actualmente, MP se ha convertido formalmente en una disciplina ante la IEEE en el 2011 [1], cuyo objetivo principal es el de contribuir a mejorar el desempeño de los procesos de las organizaciones, a fin de descubrir su verdadero comportamiento, proporcionando información de lo que se está realizando bien, diagnosticando problemas y sugiriendo automáticamente acciones o medidas correctivas o de mejora del proceso. Para ello, se apoya en dos de los elementos medulares que sirven de entrada a la mayoría de los algoritmos de MP: las bitácoras de eventos y los modelos de proceso. Por un lado, las bitácoras de eventos son creadas a partir de la ejecución de los procesos de negocio disponibles a través de sistemas de información de las organizaciones. Cada *evento* en la bitácora corresponde a una actividad o tarea que forma parte del proceso de negocio realizada por algún participante del proceso, y un conjunto de eventos conforma un *caso* o *instancia*. Por ejemplo, en la Tabla 1 se muestran tres casos y siete eventos relacionados con las actividades de un proceso en el dominio de salud. Los eventos están conformados por diversos atributos: una estampa de tiempo (periodo de tiempo en que se ejecutó dicho evento), un nombre de la actividad ejecutada, un costo (representa el costo tomado por la actividad asociada al evento), un recurso o participante (rol o nombre de la persona que llevó a cabo dicho evento), y nombre del paciente. Cada instancia corresponde

a una *traza*, por ejemplo, la secuencia de actividades {*Registro, Triaje, Examen de Sangre*} en la Tabla 1 representa una traza sobre la atención a un paciente, en este caso el paciente de nombre Brenda.

Tabla 1: Ejemplo del segmento de una bitácora de eventos del dominio de salud.

| **Id Caso** | **Estampa de Tiempo** | **Actividad** | **Costo** | **Recurso** | **Paciente** |
|---|---|---|---|---|---|
| 1 | 01-01-2018 15:20:15 | Registro | 100 | Pedro | Brenda |
| 1 | 01-01-2018 15:22:02 | Triaje | 50 | Ana | Brenda |
| 1 | 01-01-2018 15:25:43 | Ex. sangre | 800 | Julio | Brenda |
| 2 | 01-01-2018 15:43:08 | Registro | 100 | Jorge | Isidro |
| 2 | 01-01-2018 15:43:50 | Rayos X | 500 | Pedro | Isidro |
| 3 | 01-01-2018 15:46:27 | Registro | 100 | Pedro | Marta |
| 3 | 01-01-2018 15:48:14 | Triaje | 50 | Ana | Marta |

Por otro lado, un modelo de proceso puede verse como la representación abstracta y gráfica de las actividades llevadas a cabo por un proceso de una organización. Existen diversos lenguajes de notación de los modelos de procesos, siendo el estándar de Modelado y Notación de Procesos de Negocio (BPMN) [8] uno de lenguajes más usados. Un ejemplo sencillo de un modelo de proceso BPMN en el área de salud sobre la atención médica de un paciente puede verse en la Figura 1. El proceso comienza con la actividad *Registrar al paciente*, seguido por las actividades paralelas *Registro de signos vitales* y *Recolectar síntomas*, seguida por *Triaje*, es decir, actividades que se ejecutan en cualquier orden. Después, se realiza la actividad de *Actualización del expediente médico* y se crean dos caminos alternativos con las actividades *Hacer prueba de sangre* y *Realizar análisis de hemoglobina*, donde alguna de ellas es ejecutada. Los dos símbolos en forma de diamante con el signo '+' adentro denotan compuertas paralelas. El primero corresponde a una compuerta paralela de división comenzando con dos ramas concurrentes y el segundo es la compuerta de la unión. Los dos símbolos en forma de diamante con el signo 'X' adentro denotan compuertas exclusivas, es decir, se tienen caminos alternativos. En el modelo de proceso, el evento inicial (mostrado como un círculo rojo) es activado por el paciente a partir de su llegada y termina con un evento final (mostrado por un círculo blanco). Durante la ejecución del proceso en salud pueden intervenir diferentes actores, como personal de salud (médicos, enfermeras, asistentes, etc.) y pacientes, realizando una o varias actividades del proceso.

### 1.1. Requerimiento de confidencialidad

Tanto los modelos de procesos como las bitácoras de eventos son una valiosa fuente de información que permite identificar el comportamiento real de un proceso y no únicamente una vista idealizada. Sin embargo, en un dominio de estudio como el de salud, el trabajar con bitácoras con datos de eventos sensibles,

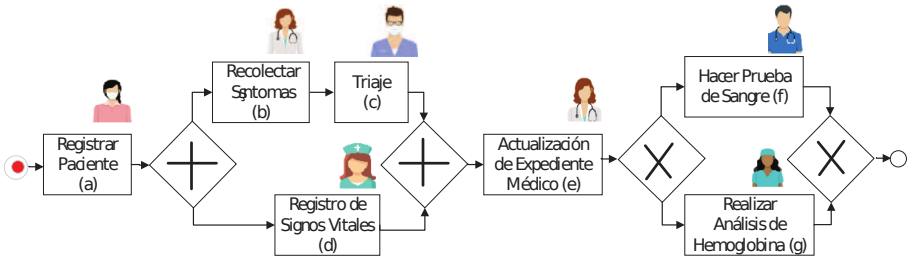



Figura 1: Modelo de proceso BPMN del seguimiento de pacientes.

así como la falta de confianza entre los participantes que ejecutan un proceso, inspira a la MP a desarrollar y utilizar métodos seguros de los datos de eventos que garanticen la confidencialidad de éstos. Sin embargo, las investigaciones que se han reportado en la literatura sobre tareas de MP pocas veces consideran cuestiones de confidencialidad de los datos, siendo éste un problema de gran relevancia, principalmente en dominios donde es necesario garantizar seguridad y privacidad de los datos.

En MP, la confidencialidad de las bitácoras de eventos no se puede lograr únicamente cifrando los datos; es necesario contar con mecanismos de seguridad y privacidad adecuados, de forma que las tareas de descubrimiento, validación de conformidad y mejora de procesos no sean afectadas y puedan seguir realizándose. Uno de los problemas comunes al intentar únicamente de anonimizar algunos atributos o registros de eventos es la posibilidad de vincular los eventos a un caso especifico e identificar atributos iguales. Esta situación de vinculación puede ocasionar que se descubra la identidad del registro. Por ejemplo, suponiendo que se tiene una bitácora de eventos de pacientes de un hospital con algunos atributos pseudo anonimizados (nombre del paciente, actividad, empleo) y un atacante desea conocer la enfermedad de cierto paciente del cual únicamente conoce su edad y algunas de las fechas en las que visitó el hospital en un período de tiempo determinado. Al identificar esta información en la bitácora de eventos puede inferir los eventos correspondientes de este paciente y, así, poder conocer su enfermedad. En este escenario se presenta un problema de ataque a la privacidad en los datos de los registros de eventos a partir de la vinculación de información.

Muchas organizaciones en salud son conscientes de la necesidad de gestionar y mejorar sus procesos, los cuales constantemente están evolucionando y cambiando dinámicamente. Algunos de los trabajos de investigación reportados en la literatura [17], [4], [11], [20], [16], [12] sobre la MP en salud, han identificado que aún no se cuenta con soluciones prácticas capaces de adaptarse a los diferentes entornos de los procesos en salud, los cuales carecen de mecanismos que garanticen la seguridad de las bitácoras de eventos construidas. En este sentido, resulta importante desarrollar y mejorar los algoritmos de MP para trabajar con datos sensibles de acuerdo con las normas internas o regulaciones externas establecidas

por instituciones médicas sin que exista algún riesgo de confidencialidad en los datos.

### 1.2. Objetivo y organización de este documento

En este documento se describen algunas de las características más relevantes de los procesos de negocio en el sector salud y se presenta una Estrategia de Confidencialidad para la Bitácora de Eventos (ECBE) basada en métodos criptográficos (haciendo uso de cifrado determinista, así como de técnicas de anonimización) sin perder la utilidad de los datos para su uso en algoritmos de Minería de Procesos. Esta estrategia construye una bitácora de eventos segura capaz de ser usada en tareas como el descubrimiento de procesos. La bitácora de eventos cifrada puede ser compartida por el equipo multidisciplinar que ejecuta el proceso sin que se comprometa la privacidad de los datos y el cumplimiento de las regulaciones, debido a que los participantes del equipo no podrán deducir información que no tenga que ver con la bitácora o el modelo de proceso cifrado.

El resto de este capítulo se encuentra organizado de la siguiente forma: en la Sección 2 se presentan algunas de las características más relevantes de los procesos de negocio en el dominio de la salud; en la Sección 3 se describe un escenario de las tareas principales, uso y explotación de la MP como servicio; en la Sección 4 se presenta una descripción de la estrategia de confidencialidad de la bitácora de eventos propuesta; en la Sección 5 se muestran los detalles de la experimentación y resultados obtenidos; finalmente, en la Sección 6 se presentan las conclusiones y trabajo futuro.

## 2. Procesos de negocio en salud

Uno de los intereses latentes de las organizaciones y, particularmente, en el sector salud consiste en aplicar MP para conocer la trayectoria de diferentes pacientes desde el momento de su ingreso hasta su alta hospitalaria. Cada visita de un paciente a un hospital constituye una instancia de proceso y los eventos individuales de cada caso se pueden obtener a partir del Sistema de Información. Este último normalmente registra información sobre las actividades logísticas y de tratamiento realizadas para pacientes específicos y el personal del hospital que los realizó. Además, una parte de un proceso en salud puede ser las interacciones con otras instituciones asistenciales y la solicitud de documentación médica previa. Por lo tanto, algunos de estos datos pueden ser compartidos sobre los límites de la organización o entre otras organizaciones y, así, realizar MP con éstos.

Típicamente, los procesos en salud se caracterizan por presentar altos niveles de variación debido a una vasta diversidad de actividades que pueden llevarse a cabo de forma secuencial o paralela por distintos participantes del proceso. Normalmente, los algoritmos de MP en salud pueden ser apoyados por guías y protocolos médicos para dar una referencia del orden de las actividades a seguir dentro del proceso. Sin embargo, en ocasiones los procesos en salud deben

considerar situaciones extraordinarias o de emergencia no previstas, que no necesariamente corresponden con el orden establecido en el proceso, por lo que en la mayoría de los casos el flujo de ejecución de las actividades en el proceso no se cumple adecuada o completamente de acuerdo con el orden previsto.

Los procesos de salud son llevados a cabo por un equipo multidisciplinar (médicos, enfermeras, especialistas, asistentes, etc.), quienes, de manera autónoma e independiente, ejecutan una o varias actividades del proceso y toman decisiones sobre determinadas tareas complejas sin apegarse completamente al proceso clínico establecido y sin limitarse al acceso de la información sensible que pueden manejar. La sensibilidad de los datos es un tema de gran importancia que necesita constantemente tomarse en cuenta, ya que los datos podrían incluir información tal como la condición médica actual del paciente, sus comorbilidades, tratamientos e información personal que no debería ser revelada a ninguna entidad externa o interna que no cuente con un acceso o permiso necesario. En este escenario, las bitácoras de eventos obtenidas a partir de la ejecución de un proceso en salud deberían ser cuidadosamente manejadas debido a la confidencialidad, privacidad, uso y almacenamiento de la información que éstas contienen. Además, el riesgo de privacidad se incrementa con el uso de proveedores de servicios externos, como la nube, para delegarle el costo computacional y de almacenamiento asociado con el uso de algoritmos, como los de MP, y de almacenamiento de las bitácoras de eventos. Sin embargo, las técnicas clásicas de MP no están preparadas para lidiar con problemas relacionados con la confidencialidad de las bitácoras de eventos dentro de un ambiente de trabajo multidisciplinario, así como la externalización de las bitácoras de eventos para tareas de MP en la nube.

## 3. La Minería de Procesos como servicio

A partir de la MP, muchas organizaciones pueden identificar cuellos de botella, desviaciones, anticipar y diagnosticar problemas de rendimiento y cumplimiento mediante el uso de las bitácoras de eventos y los modelos de procesos. Particularmente, las técnicas MP pueden ser agrupadas en tres tareas principales:

1. El **descubrimiento de modelos de procesos** tiene por objetivo descubrir automáticamente el modelo del proceso asociado a los eventos almacenados en la bitácora de eventos obtenidos a partir de la ejecución de un sistema de información, es decir, construye el modelo de proceso tomando como entrada la bitácora de eventos;
2. La **verificación de conformidad** consiste en reproducir cada una de las instancias o trazas contenidas en la bitácora de eventos en el modelo de proceso para comprobar si lo que se tiene registrado en la bitácora corresponde con la ejecución y orden de las actividades que se muestran en el modelo de proceso y viceversa; a partir de esta tarea es posible identificar desviaciones

que pueden ocurrir en el proceso o en su correspondiente bitácora de eventos;

3. La **mejora del proceso** se enfoca en mejorar el desempeño de los procesos, ya sea cambiando o extendiendo el modelo previo construido.

Con el auge del Big Data, diariamente se generan grandes cantidades de datos de eventos a partir de la ejecución de procesos de distintos dominios de los sistemas de información de las organizaciones, lo que hace necesario contar con técnicas de MP disponibles como un servicio para delegarles tareas de descubrimiento, conformidad y mejora. En este escenario, los datos (bitácoras de eventos) dejan de estar bajo el control del propietario del proceso de negocio y pueden ser accedidos por el proveedor del servicio, lo que puede ocasionar un riesgo de confidencialidad.

La privacidad se centra en el uso y el manejo de los datos personales de los individuos, así como las políticas que garantizan que la información personal de los usuarios se recopile, comparta y utilice de manera correcta [19]. Uno de los enfoques más usados para garantizar la confidencialidad en la bitácora de eventos sin perder utilidad en los datos, es el cifrado, el cual transforma un texto legible en texto ilegible y viceversa, utilizando sus respectivas claves de cifrado y descifrado. A continuación se describe la estrategia de confidencialidad propuesta.

## 4. Estrategia de Confidencialidad de la Bitácora de Eventos

La Estrategia de Confidencialidad para la Bitácora de Eventos (ECBE) descrita en este trabajo está basada en métodos criptográficos y técnicas de anonimización. Esta estrategia asegura la confidencialidad en los datos de la Bitácora de Eventos, ya que estos se transforman a un formato ilegible que no puede ser interpretado por los usuarios. La confidencialidad de los datos permite que la bitácora sea compartida con un servidor externo, como la nube, para llevar a cabo tareas de MP sin que se pueda revelar información que no esté relacionada con la tarea de descubrimiento del proceso de negocio cifrado. ECBE se asegura de preservar la utilidad de los datos en la Bitácora de eventos mediante técnicas de cifrado determinista, es decir, siempre se genera el mismo texto cifrado para un texto plano y una clave dados, lo cual mantiene la diferenciación en los datos para poder descubrir el modelo del proceso. El escenario de trabajo de la estrategia ECBE se compone de dos entornos, privado y externo (ver Figura 2), donde la bitácora de eventos puede moverse de un entorno privado (seguro) a un entorno externo (inseguro) en su versión cifrada (BE'). Una vez que el algoritmo de descubrimiento de modelos de proceso recibe como entrada una BE', el modelo resultante de esta tarea es un modelo de procesos cifrado (MP') del cual no se puede extraer información del contexto del proceso de negocio en estudio. El objetivo del uso de los entornos en la ECBE es garantizar la protección y determinar los niveles de seguridad que se tendrán en los registros que se tienen

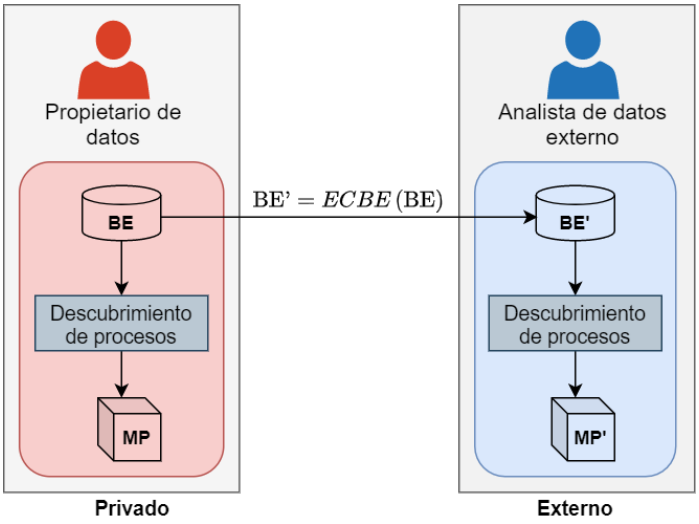



Figura 2: Entornos de seguridad considerados en ECBE.

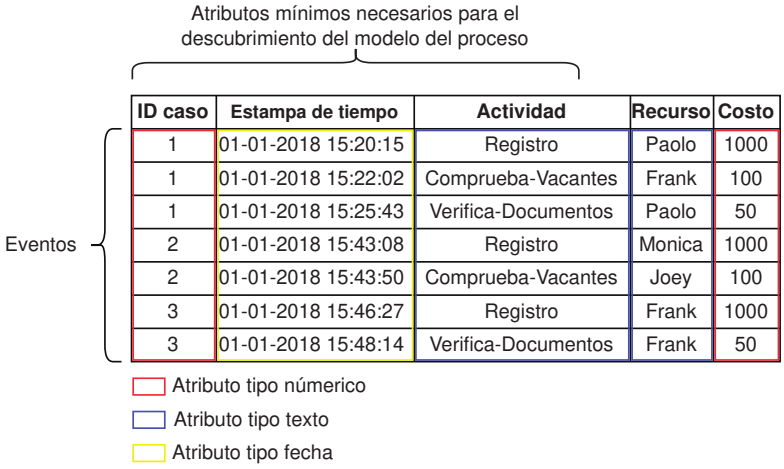
en las bitácoras. En el entorno externo se protege a la bitácora de eventos de analistas externos del proceso y de cualquier otro individuo que requiera hacer uso no permitido de la misma.

La bitácora de eventos cifrada (BE') evita cualquier riesgo de fuga de información que pueda ocurrir, aunque el atacante conozca el contexto en el que ocurre el proceso de negocio. ECBE asegura la bitácora de eventos considerando sus atributos, como puede verse en el ejemplo de la Figura 3, donde se presentan eventos que contienen atributos que almacenan valores del tipo numérico, texto o fecha. Comúnmente, la estrategia ECBE se enfoca en proteger los atributos más relevantes para la tarea de descubrimiento de procesos, como los que se presentan en la Figura 3 (estampa de tiempo, actividad, recurso, costo).

La estrategia de confidencialidad ECBE se compone de cuatro etapas (ver Figura 4), las cuales se enfocan en proteger los datos de los atributos de la bitácora de eventos e información que se puedan derivar de ésta. A continuación se describen cada una de las etapas.

**Etapa 1. Filtrado y modificación de la entrada.** Esta etapa consiste en preparar los datos relevantes para el análisis deseado. Incluye las siguientes actividades:

1. A partir de la bitácora de eventos $L$ en formato de tabla bidimensional, se retiran los atributos que son irrelevantes al análisis deseado (datos no sensibles). $L$ está conformado por una serie de trazas de la forma $(a_1, a_2, \cdots, a_n)$, donde $a_i$ representa una actividad. Si la traza se repite k veces, se indica como $(a_1, a_2, \cdots, a_n)^k$.




| ID caso | Estampa de tiempo | Actividad | Recurso | Costo |
|---|---|---|---|---|
| 1 | 01-01-2018 15:20:15 | Registro | Paolo | 1000 |
| 1 | 01-01-2018 15:22:02 | Comprueba-Vacantes | Frank | 100 |
| 1 | 01-01-2018 15:25:43 | Verifica-Documentos | Paolo | 50 |
| 2 | 01-01-2018 15:43:08 | Registro | Monica | 1000 |
| 2 | 01-01-2018 15:43:50 | Comprueba-Vacantes | Joey | 100 |
| 3 | 01-01-2018 15:46:27 | Registro | Frank | 1000 |
| 3 | 01-01-2018 15:48:14 | Verifica-Documentos | Frank | 50 |

Figura 3: Atributos típicos de una BE.

2. Se realiza un filtrado de trazas removiendo de la bitácora de eventos aquellas trazas con menor frecuencia de acuerdo a un umbral $\theta$. Los casos que formen la misma traza se agrupan, de esta manera se forma un conjunto de trazas únicas. Por ejemplo, sea $L$ una bitácora de eventos con $L = \{(a,b,c,d,e,f)^{10}$,$(a,b,e)^{8}$, $(a,c,b,d,f,e)^{20}$,$(a,c,e,f)^{4}$, $(a,c,f,c,e)^{15}\}$, y $\theta = 10$. Después del filtrado se tendría $L' = \{(a,b,c,d,e,f)^{10}$, $(a,c,b,d,f,e\ )^{20}$, $(a,c,f,c,e)^{15}\}$.

**Etapa 2. Cifrado del texto en claro.** En esta etapa se proporciona confidencialidad sin pérdida de utilidad a los datos de la bitácora. Para mantener la capacidad de aplicar cálculos matemáticos básicos a los valores numéricos en la bitácora de eventos, se usa el algoritmo de cifrado homomórfico Paillier [15], mientras que para datos de tipo texto, se aplica un algoritmo de cifrado determinista AES [5]. Los atributos *Id Caso* y *Estampa de tiempo* no son cifrados en esta etapa; éstos son protegidos mediante técnicas que les ayudan a conservar su utilidad. La realización de esta etapa se ilustra en la transformación de los atributos "*Actividad*", "*Recurso*" y "*Costo*" de la bitácora de eventos en claro en la Tabla 1 y la bitácora de eventos cifrada (BE') en la Tabla 2.

**Etapa 3. Conversión a tiempos relativos.** En esta penúltima etapa, el atributo "*Estampa de tiempo*" es modificado a fin de evitar que los periodos de tiempo de la bitácora de eventos sean identificados. Para ello, se selecciona otra fecha (que se mantiene secreta junto a las claves de cifrado) para que todos los eventos sean relativos a ésta. La estampa de tiempo de cada evento en la bitácora de eventos se sustituye por su diferencia con la fecha secreta seleccionada. En las Tablas 1 y 2, en el atributo "*Estampa de tiempo*" se muestra la conversión

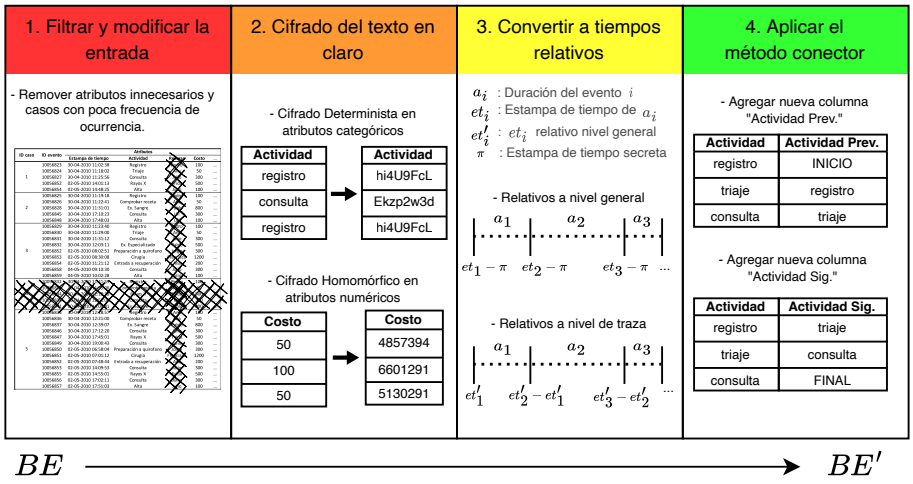



Figura 4: Etapas de la estrategia ECBE.

realizada.

**Etapa 4. Aplicación de método conector**. La última etapa consiste en aplicar un método que permite de forma segura obtener y extraer la estructura de un grafo dirigido a partir de la bitácora de eventos cifrada (BE'). Esta estructura sirve de entrada de muchos de los algoritmos de MP. En esta misma etapa se realiza la reconstrucción de la bitácora de eventos original una vez que se cuente con las claves de descifrado y los valores de fecha relativos. Las actividades consideradas en el método conector son las siguientes:

1. Se agrega el atributo llamado "*Actividad Prev.*", el cual indica la actividad previa a cada evento en la bitácora de eventos. Esto permite identificar cuáles actividades están directamente conectadas. En el caso del primer evento de cada traza, éste tendrá una actividad previa artificial "Start" cifrada con los mismos parámetros con los que fue cifrado el atributo "*Actividad*".
2. Se asegura la información contenida en el atributo "*Id Caso*" para no permitir el agrupamiento de eventos en trazas, manteniendo la posibilidad de recrear la bitácora de eventos original. Esto es realizado dando a cada evento un ID aleatorio ("*ID*") e indicando el ID previo ("*ID P.*"). Estas nuevas columnas en BE' permiten identificar el evento siguiente a nivel de traza. El "*ID P.*" de la actividad inicial en una traza siempre será 0.
3. Enseguida, los atributos "*ID*" e "*ID P.*" se ocultan, ya que son utilizados solo para reconstruir la bitácora de eventos. Estos se concatenan y se cifran usando el algoritmo de cifrado determinista AES (igual que en la Etapa 2) y se agrega como nuevo atributo llamado "*Conector*" en la bitácora de eventos.

Tabla 2: Bitácora de eventos de la Tabla 1 con ECBE aplicada.

| Estampa de tiempo | Actividad | Actividad Prev. | Recurso | Costo | Conector |
|---|---|---|---|---|---|
| 00-00-0000 00:00:42 | y7y4PUi2 | NM7Jgoum | XRCDyLgS | 59301 | 5q8aL2at |
| 00-00-0000 00:01:47 | UGdnk8fh | y7y4PUi2 | hLrq2mYD | 46012 | KQBindVr |
| 00-00-0000 15:46:27 | bvS(28op | UGdnk8fh | 4hIDYn0q | 98744 | CKl07FSq |
| 00-00-0000 00:01:47 | y7y4PUi2 | NM7Jgoum | tpwUTcAl | 58430 | N9a1qeto |
| 00-00-0000 00:03:41 | jhg!676 | y7y4PUi2 | XRCDyLgS | 81023 | XIQ7ZnqA |
| 00-00-0000 15:20:15 | y7y4PUi2 | NM7Jgoum | XRCDyLgS | 59015 | M4qAwqqz |
| 00-00-0000 15:43:08 | UGdnk8fh | y7y4PUi2 | hLrq2mYD | 42110 | z5Zb56jY |

4. Posteriormente, se usa el atributo "*Estampa de tiempo*" para proteger los valores de tiempo en eventos que cuenten con el mismo "*Id Caso*". Éstos se hacen relativos con respecto a la estampa de tiempo precedida, a excepción del primer evento de cada traza, el cual se mantiene con el mismo valor. Esto permite el cálculo de la duración de cada uno de los arcos en un grafo dirigido de seguimiento directo (DFG), pero hace complicado identificar eventos basándose en la parte temporal en la que ocurrieron.
5. Para finalizar, se retira el atributo "*Id Caso*" y se desorganiza el orden de todas las filas y se obtiene como resultado una bitácora de eventos protegida para el entorno externo (BE') (ver Tabla 2), capaz de ser usada para descubrimiento de procesos basándose en las causalidades de los eventos haciendo uso de un grafo DFG (generado de "*Actividad*" y "*Actividad Prev.*").

## 5. Evaluación de la estrategia ECBE

La estrategia ECBE se implementó en el lenguaje de programación Java, considerado como un lenguaje de alto nivel que permite expresar los algoritmos en un nivel más abstracto del lenguaje máquina. ECBE se evaluó usando datos reales extraídos de sistemas ERP (Enterprise Resource Planning) de tres bitácoras de eventos del campo de salud (ver detalles en la Tabla 3) con una clave de cifrado AES de una longitud de 128 bits. Con estas bitácoras se comprobó que los modelos de proceso DFG resultantes de las bitácoras de eventos cifradas fueran iguales a los resultantes de sus versiones en claro. De esta manera se comprueba que no existe pérdida en la utilidad de los datos al protegerlos con la estrategia propuesta.

En la evaluación de la estrategia propuesta se consideró el correcto descubrimiento de los modelos procesos y se comprobó que la solución propuesta permite obtener modelos con una nula pérdida de precisión, lo que demuestra que la solución propuesta basada en una estrategia criptográfica no produce pérdida de utilidad en los datos de entrada para los algoritmos de descubrimiento de procesos.

Tabla 3: Bitácoras de eventos de datos médicos reales.

| Nombre | No. eventos | No. actividades | No. trazas |
|---|---|---|---|
| Sepsis Cases [9] | 15,214 | 16 | 1,050 |
| BPIC11 [6] | 150,291 | 624 | 1,143 |
| Hospital Billing [10] | 451,359 | 18 | 100,000 |

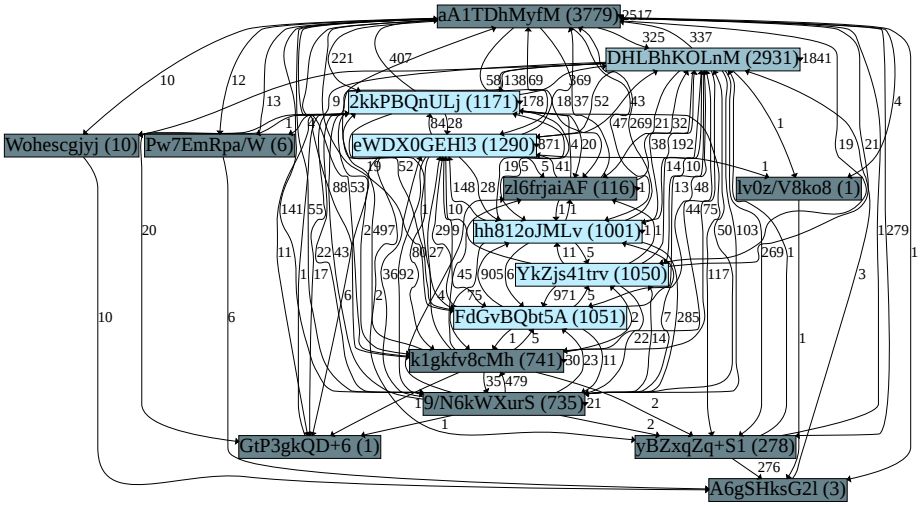



Figura 5: Modelo de proceso DFG de la bitácora de eventos cifrada Sepsis Cases resultante de ECBE.

La Figura 5 muestra el modelo descubierto usando la bitácora de eventos de Sepsis Cases cifrada usando la estrategia ECBE propuesta. Al comparar el modelo de proceso cifrado con el obtenido del modelo de proceso en su versión en claro, ambos modelos presentan la misma cantidad de aristas (137) y el mismo peso en cada una, lo que demuestra que no existe pérdida de utilidad al usar ECBE. Por lo tanto, el modelo obtenido a partir de la bitácora de eventos cifrada (BE') puede convertirse en el modelo que se obtendría usando la bitácora de eventos (en claro) y las llaves de descifrado correspondientes.

## 6. Conclusiones

Los procesos en salud representan un gran desafío debido a las diversas características que éstos presentan, como altos niveles de variación, diversidad en la ejecución y flujo de las actividades de los procesos, trabajo colaborativo de un equipo multidisciplinar (médicos, enfermeras, especialistas, asistentes, etc), así como el manejo de datos sensibles. Ésto hace necesario contar con algoritmos de MP capaces de trabajar con temas de confidencialidad, privacidad, uso y al-

macenamiento de la información, particularmente de las bitácoras de eventos y modelos de procesos. En este trabajo se describieron algunas de las características más relevantes de los procesos en salud y se presentó una estrategia llamada ECBE para garantizar la confidencialidad de la bitácora de eventos de un proceso de negocio en salud, sin pérdida de su utilidad en tareas de MP. ECBE está basada en la protección de los campos de una bitácora de eventos usando un algoritmo de cifrado determinista y técnicas de anonimización para mantener la diferenciación en los datos a fin de compartirlos con un servidor externo como la nube y, así, poder llevar a cabo tareas de MP sin que se pueda revelar información del proceso de negocio. A partir de la bitácora de eventos cifrada resultante por la estrategia propuesta no es posible relacionar los eventos, permitiendo que no se descubran trazas o secuencias de eventos, manteniendo la capacidad de realizar las tareas de MP. Una de las limitaciones de ECBE es que solo permite la creación de modelos DFG, lo que no permite usar algoritmos de MP basados en consecuencias. Como parte del trabajo futuro se considera el d esarrollo de un mecanismo de control de acceso y compartición segura tanto de la bitácora de eventos como del modelo de proceso para entidades autorizadas.

## Referencias


[1] Wil M. P. van der Aalst. *Process Mining: Data Science in Action*. 2.ª ed. Heidelberg: Springer, 2016. ISBN: 978-3-662-49850-7. DOI: 10.1007/978-3-662-49851-4.

[2] Wil M. P. van der Aalst y A. J. M. M. Weijters. "Process mining: a research agenda". En: *Comput. Ind.* 53.3 (2004), págs. 231-244. DOI: 10.1016/j.compind.2003.10.001. URL: https://doi.org/10.1016/j.compind.2003.10.001.

[3] Karim Abouelmehdi, Abderrahim Beni-Hessane y Hayat Khaloufi. "Big healthcare data: preserving security and privacy". En: *Journal of Big Data* 5 (ene. de 2018). DOI: 10.1186/s40537-017-0110-7.

[4] Elisabetta Benevento et al. "Evaluating the Effectiveness of Interactive Process Discovery in Healthcare: A Case Study". En: *Business Process Management Workshops*. Ed. por Chiara Di Francescomarino, Remco Dijkman y Uwe Zdun. Cham: Springer International Publishing, 2019, págs. 508-519.

[5] Joan Daemen y Vincent Rijmen. "The Advanced Encryption Standard Process". En: *The Design of Rijndael: AES — The Advanced Encryption Standard*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, págs. 1-8. ISBN: 978-3-662-04722-4. DOI: 10.1007/978-3-662-04722-4_1.

[6] Boudewijn van Dongen. *Real-life event logs - Hospital log*. 2011. DOI: 10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54.

[7] EU General Data Protection Regulation. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*

*(General Data Protection Regulation) (Text with EEA relevance)*. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[8] Business Process Management Initiative. *BPMN Specification - Business Process Model and Notation*. 2005. URL: https://www.bpmn.org.

[9] Felix Mannhardt. *Sepsis Cases - Event Log*. 2016. DOI: 10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460.

[10] Felix Mannhardt. *Hospital Billing - Event Log*. 2017. DOI: 10.4121/uuid:76c46b83-c930-4798-a1c9-4be94dfeb741.

[11] Niels Martin et al. "Interactive Data Cleaning for Process Mining: A Case Study of an Outpatient Clinic's Appointment System". En: *Business Process Management Workshops*. Ed. por Chiara Di Francescomarino, Remco Dijkman y Uwe Zdun. Cham: Springer International Publishing, 2019, págs. 532-544. ISBN: 978-3-030-37453-2.

[12] Jorge Munoz-Gama et al. "Process mining for healthcare: Characteristics and challenges". En: *Journal of Biomedical Informatics* 127 (2022), pág. 103994. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2022.103994. URL: https://www.sciencedirect.com/science/article/pii/S1532046422000107.

[13] Blake Murdoch. "Privacy and artificial intelligence: challenges for protecting health information in a new era". En: *BMC Medical Ethics* (ene. de 2021). DOI: https://doi.org/10.1186/s12910-021-00687-3.

[14] Sharyl Nass, Laura Levit y Lawrence Gostin. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Feb. de 2009. ISBN: 978-0-309-12499-7. DOI: 10.17226/12458.

[15] Pascal Paillier. "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes". En: *Advances in Cryptology — EUROCRYPT '99*. Ed. por Jacques Stern. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, págs. 223-238. ISBN: 978-3-540-48910-8.

[16] Marco Pegoraro et al. "Analyzing Medical Data with Process Mining: a COVID-19 Case Study". En: *arXiv e-prints*, arXiv:2202.04625 (feb. de 2022), arXiv:2202.04625. arXiv: 2202.04625 [cs.DB].

[17] Eric Rojas et al. "Process mining in healthcare: A literature review". En: *Journal of Biomedical Informatics* 61 (2016), págs. 224-236. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2016.04.007. URL: https://www.sciencedirect.com/science/article/pii/S1532046416300296.

[18] Sudhakar Sengan et al. "Secured and Privacy-Based IDS for Healthcare Systems on E-Medical Data Using Machine Learning Approach". En: *International Journal of Reliable and Quality E-Healthcare* 11 (oct. de 2021), págs. 1-11. DOI: 10.4018/IJRQEH.289175.

[19] H. Smith, Tamara Dinev y Heng Xu. "Information Privacy Research: An Interdisciplinary Review". En: *MIS Quarterly* 35 (dic. de 2011), págs. 989-1015. DOI: 10.2307/41409970.

[20] Zoe Valero-Ramon et al. "Dynamic Models Supporting Personalised Chronic Disease Management through Healthcare Sensors with Interactive Pro-

cess Mining". En: *Sensors* 20.18 (2020). ISSN: 1424-8220. DOI: 10.3390/s20185330. URL: https://www.mdpi.com/1424-8220/20/18/5330.

[21] Alan F. Westin. "Privacy And Freedom". En: *Washington and Lee Law Review* 25 (ene. de 1968), págs. 166-170.