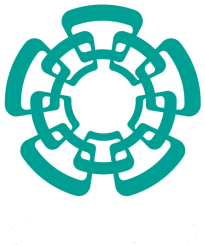

Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional

Cinvestav Unidad Tamaulipas

Método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles

Tesis que presenta:

Víctor Ríos Barrientos

Para obtener el grado de:

Maestro en Ciencias en Ingeniería y Tecnologías Computacionales

Dr. Víctor Jesús Sosa Sosa, Co-Director
Dr. José Luis González Compeán, Director

Cd. Victoria, Tamaulipas, México. Diciembre, 2019

La tesis presentada por Víctor Ríos Barrientos fue aprobada por:

_____

Dr. Iván López Arévalo

Dr. Miguel Morales Sandoval

Dr. Víctor Jesús Sosa Sosa, Co-Director

Dr. José Luis González Compeán, Director

Cd. Victoria, Tamaulipas, México., 8 de Diciembre de 2019

Vive

Agradecimientos

- A Gollita por estar ahí.
- A mis asesores, el Dr. José Luis González Compeán y el Dr. Víctor Jesús Sosa Sosa por sus enseñanzas y por guiarme en todo momento.
- A mis revisores, el Dr. Miguel Morales Sandoval y el Dr. Iván López Arévalo por sus recomendaciones, las cuales me ayudaron a mejorar esta presente tesis.
- A cada uno de mis compañeros de la maestría, por los gratos momentos que vivimos.
- También agradezco al personal administrativo de CINVESTAV Unidad Tamaulipas por su ayuda durante mi estadía.
- Agradezco a CONACyT por el apoyo económico brindado que me permitió concentrarme en mis estudios y al CINVESTAV Unidad Tamaulipas por la oportunidad de realizar estudios de posgrado.
- Agradezco también a todos lo que no he mencionado y que me brindaron ayuda y consejo.

Índice General

Índice de Figuras

Índice de Tablas

Resumen

Método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles

por

Víctor Ríos Barrientos
Cinvestav Unidad Tamaulipas
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, 2019
Dr. José Luis González Compeán, Director
Dr. Víctor Jesús Sosa Sosa, Co-Director

Las cadenas de valor son esquemas de manejo de información que permiten a organizaciones transformar contenidos digitales en versiones enriquecidas de los mismos a través de flujos de trabajo. En estas cadenas de valor se manejan o comparten datos que son adquiridos, transformados, transportados y consumidos o producidos por terceros. El cómputo en la nube representa una tecnología costo-beneficio que esta llegando a ser una solución para que las organizaciones desplieguen sus cadenas de valor y de esta forma puedan participar los involucrados desde cualquier lugar usando cualquier dispositivo a cualquier hora. A pesar de lo antes mencionado, en la literatura han sido reportados incidentes tales como violaciones a la confidencialidad, integridad y/o privacidad de los contenidos digitales en las diferentes etapas de procesamiento cuando las organizaciones subcontratan este tipo de servicios. Estos incidentes resultan críticos cuando las cadenas de valor que se despliegan en la nube involucran datos sensibles manejados por terceros (el proveedor de servicio).

En esta tesis se diseñó e implementó un método para el aseguramiento de contenidos que viajen a través de cadenas de valor desplegadas en la nube. Mediante patrones de paralelismo creados con contenedores virtuales se mejora la eficiencia del proceso de creación de sobres digitales, mientras que el seguimiento del ciclo de vida de los productos asegurados por los sobres se realiza con un esquema basado en tecnología de registro distribuido y bloques encadenados. Para mostrar la factibilidad del método propuesto se evaluaron escenarios de secuenciación de procesos y etapas de cadenas de valor construidas por las operaciones de compartición y manejo de archivos sensibles, así como

el cumplimiento de las reglas de negocio y las operación definidas por los interesados en dichas cadenas de valor. Para mostrar la eficiencia del método propuesto, se emularon escenarios donde las organizaciones construyan servicios de compartición de archivos sensibles en la nube para cadenas de valor reales (imágenes médicas y satelitales), con propiedades de verificabilidad de transacciones. La evaluación experimental y los estudios de caso revelaron la factibilidad de aplicar el método propuesto a escenarios reales de compartición de archivos sensibles.

Abstract

Transaction control method through distributed ledger for value chains in sensitive file sharing systems.

by

Víctor Ríos Barrientos
Cinvestav Tamaulipas
Center for Research and Advanced Studies of the National Polytechnic Institute, 2019
Dr. José Luis González Compeán, Co-advisor
Dr. Víctor Jesús Sosa Sosa, Co-advisor

Value chains are information management schemes that enable organizations to transform digital content into rich versions through the establishment of wrokflows. During the these chains the data/information is acquired, managed, transformed, transported, consumed by different entities (organizations, government agencies, partners, etc.). Cloud computing is cost-efficiency for organizations to build value chains where all the entities can participate in this chain any time, anyway by using any device. However, incidents such as violations of confidentiality, integrity and privacy still arise when organizations use this type of service. This is critical for organizations when the value chains manage sensitive data (e.g. health or strategic data).

This thesis document presents a method for securing sensitive contents managed in value chain deployed on the cloud. In this method, parallel patterns based on virtual containers improve the efficiency of the encapsulation of sensitive contents into secure digital envelopes (SDE). The verification and traceability of DSE lifecycle in value chains are performed by using intelligent contracts in the blockchain. To show the feasibility of this method, a software prototype that supports the deployment of real value chains, managing sensitive contents (medical and satellite imagery) with transaction verification and traceability, was implemented according to the proposed method. An experimental evaluation based on study cases revealed the feasibility of applying this method to the management of value chains of sensitive contents.

Nomenclatura

AES	Advanced Encryption Standard
CP-ABE	Ciphertext Policy - Attribute Based Encryption
ETL	Extract, transform and load
AES4SeC	Attribute based Encryption and Signing for Security in Cloud

1

Introducción

1.1 Antecedentes y motivación para proyecto

El cómputo en la nube representa una solución atractiva para aquellas organizaciones que buscan procesar contenidos digitales en la nube a través de flujos de trabajo. En este contexto, los flujos de trabajo representan un grupo de transacciones que automatizan procesos de transformación de los contenidos digitales. A cada punto del flujo de trabajo lo llamaremos $Etapa$. Cada etapa recibe de entrada un contenido digital para procesarlo (realiza cambios a la versión entrante) y produce una nueva versión de ese contenido, la cual se llamará en este documento *producto digital*. Las etapas pueden ser desplegadas en un solo computador (modelo único), en un grupo de computadores (cluster) o en la nube (Cloud).

Al proceso de transformación de un producto desde su versión original pasando por las diferentes versiones del mismo a través de diferentes etapas se le llama $ciclodevidadelproductodigital$.

Cuando las etapas de los flujos de trabajo se encadenan en forma secuencial agregando

características (valor) a los productos digitales se crean cadenas de valor, las cuales son cruciales para las organizaciones de diferentes ámbitos [36]. El valor agregado a los productos digitales considera características de contexto tales como anotaciones o metadatos (meta información) y de contenido tales (edición o transformación de datos, marcado). En la Figura 2.4 se muestra una cadena de valor en el ámbito médico (archivos sensibles), donde las tomografías (generadas por un tomógrafo) son el producto digital, las tomografías son enviadas al especialista que agrega un diagnóstico, y este a su vez envía la nueva versión (del producto digital) al médico, que también puede agregar alguna característica (valor), de esta forma el producto digital (tomografías) pasa por diferentes etapas (especialista, médico) en las cuales se les agregan características (diagnóstico, receta médica etc.) generando cadenas de valor.

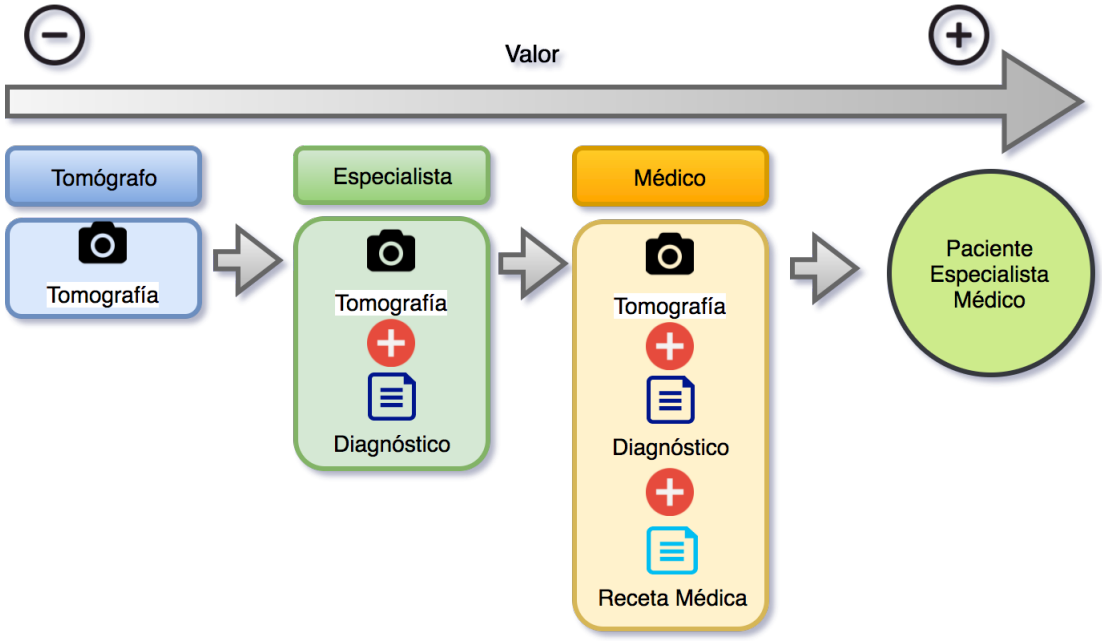

Figura 1.1: Cadena de valor sobre datos médicos.

Cuando las organizaciones despliegan cadenas de valor en una nube pública, deben hacer frente a dos desafíos: el primero relacionado a la seguridad, incidentes tales como violaciones a la confidencialidad de los datos, en las diferentes etapas de procesamiento se han observado en

la práctica y se han descrito en la literatura [2] [32] [37], mientras que el segundo tiene que ver con el seguimiento a la ejecución de cada etapa de un flujo de trabajo, incidentes de repudio de datos recibidos y entregados podrían presentarse produciendo conflictos entre las organizaciones que colaboran en las cadenas de valor de los productos digitales. En la cadena de valor de la Figura 2.4 los desafíos mencionados anteriormente se encuentra entre las diferentes etapas (flechas grises), ya que el procesamiento de cada etapa puede realizarse en diferente lugar geográfico, lo que genera que el producto viaje y se almacene en un medio inseguro entre cada etapa.

Por un lado la técnica de sobre digital ha mostrado resultados alentadores para responder a los problemas de seguridad en la nube [22] [23] , esta técnica fusiona estrategias de criptografía simétrica y asimétrica; en la criptografía simétrica se utiliza la misma clave para cifrar y descifrar sin embargo en dicha técnica se tiene el *problema de distribución de clave*, ya que al cifrar un dato es necesario enviar la clave al destinatario para que pueda descifrarlo, el problema radica en decidir por qué medio enviar la clave ya que es más fácil para el atacante interceptar la clave que intentar descifrar el contenido por algún otro método. Este problema es resuelto por medio de la criptografía asimétrica en la cual se tiene un par de claves, una pública y una privada [21] de esta forma la clave privada, que es necesaria para acceder a los datos se mantiene resguardada por el usuario.

A pesar de que la criptografía asimétrica permite que dos usuarios utilicen diferentes llaves, produce un rendimiento inferior en comparación con el rendimiento producido por los métodos simétricos [30]. La técnica de sobre digital es un concepto criptográfico que permite fusionar los esquemas simétricos y asimétricos. Cada esquema se utiliza de manera complementaria realizando una función diferente. En la Figura 1.2 se muestra el proceso de generación del Sobre Digital donde el algoritmo simétrico (cifrador simétrico AES) crea una llave secreta que se utiliza para el cifrado de datos en masa y el algoritmo asimétrico (Cifrador asimétrico CP-ABE) utiliza la llave pública del receptor para cifrar la llave secreta del algoritmo simétrico [42]. La técnica de sobre digital ha mostrado resultados alentadores para responder a los problemas de seguridad en la nube [22] [23].

La necesidad de llevar un registro seguro de transacciones que se aplican a un contenido digital ha

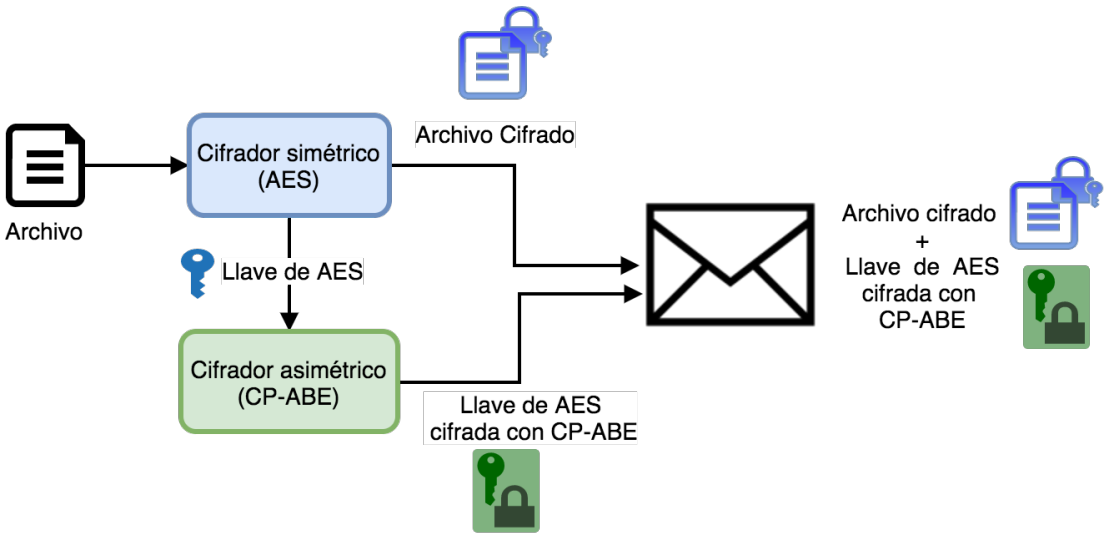

Figura 1.2: Proceso de generación del Sobre Digital.

sido parte de la motivación que da paso a la tecnología de cadena de bloques o blockchain, la cual fue creada junto a los bitcoins en el año 2008 por Satoshi Nakamoto [24]. La blockchain es esencialmente una base de datos distribuida de registros o libro mayor público de todas las transacciones o eventos digitales que se han ejecutado y compartido entre las partes participantes. Cada transacción en el libro mayor público se verifica por consenso de la mayoría de los participantes en el sistema. Y, una vez ingresada, la información no puede ser borrada [9]. La tecnología de cadena de bloques ha mostrado ser una solución para que las organizaciones den seguimiento a las transacciones (o cadenas de valor en el contexto de productos digitales).

Sin embargo, en escenarios del mundo real, se requiere unir ambas técnicas y resolver los problemas de factibilidad al realizar dicha unión. El primer problema tiene que ver con desplegar sistemas basados en sobres digitales en forma eficiente, mientras que el segundo tiene que ver con coordinar y sincronizar el manejo de sobres digitales con un esquema de manejo de transacciones mediante el uso de tecnologías como la de bloques encadenados.

En esta tesis se diseñó e implementó un método para el aseguramiento de contenidos que viajen a través de cadenas de valor desplegadas en la nube. Mediante patrones de paralelismo creados con contenedores virtuales se mejora la eficiencia del proceso de creación de sobres digitales, mientras

que el seguimiento del ciclo de vida de los productos asegurados mediante los sobres se realiza con un esquema basado en tecnología de registro distribuido y bloques encadenados.

Para mostrar la factibilidad del método propuesto se evaluaron escenarios de secuenciación de procesos y etapas de cadenas de valor construidas por las operaciones de compartición y manejo de archivos sensibles, así como el cumplimiento de las reglas de negocio y las operación definidas por los interesados en dichas cadenas de valor. Para mostrar la eficiencia del método propuesto, se emularon escenarios donde las organizaciones construyan servicios de compartición de archivos sensibles en la nube para cadenas de valor reales (imágenes médicas y satelitales) y se realizaron experimentos para evaluar el rendimiento en términos de experiencia del usuario final.

1.2 Planteamiento del problema

En la actualidad existe la necesidad de crear servicios de cadenas de valor en productos digitales a través de conectar cadenas de procesamiento en forma de flujos de trabajo (workflows). Como se mencionó en la sección anterior, existen técnicas para compartir productos digitales de manera segura basadas en sobre digital [33] que ha mostrado resultados alentadores. Sin embargo, la definición original de esta técnica sufre de problemas de eficiencia, cuando se trata de procesar grandes y diversas fuentes de datos, y no contempla en su diseño un método que garantice el seguimiento fiable (registro distribuido) de las etapas por donde van transitando los datos en una cadena de valor. Hacer frente a estos desafíos en forma eficiente motivo el desarrollo de esta tesis.

1.3 Hipótesis

Aplicando patrones de paralelismo con contenedores es posible mejorar la eficiencia de las soluciones criptográficas para el aseguramiento de contenidos que viajan a través de cadenas de valor creadas en la nube, permitiendo también hacer factible, en escenarios reales, establecer control sobre

las transacciones realizadas con productos digitales mediante la tecnología de cadena de bloques.

1.4 Objetivos

1.4.1 Objetivo general

Desarrollar un método eficiente de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles.

1.4.2 Objetivo particular

- Definir un esquema de creación de sobres digitales eficiente mediante patrones de paralelismo basados en contenedores virtuales.
- Crear un mecanismo de registro de seguimiento de transacciones en cadenas de valor mediante bloques encadenados.
- Crear mecanismo para fusionar esquemas de manejo de productos digitales en cadenas de valor con la creación de sobres digitales y la red de registro de transacciones.

1.5 Metodología

Las siguientes etapas se consideran necesarias para la realización de la tesis y se describen en el orden de realización.

- Etapa 1: Introducción y estado del arte

  Durante esta etapa se realizó una investigación del estado del arte que abarca los temas de seguridad (sobre digital) y cadena de bloques. Además, se establecieron los objetivos (objetivo general y objetivos particulares), así como la hipótesis del trabajo de tesis a realizar. El producto final de esta etapa es el protocolo de tesis.

- Etapa 2: establecimiento y desarrollo de las fases del método propuesto

  En es etapa se contempla realizar lo siguiente:

  - Encapsular el proceso de generación del Sobre Digital por medio de contenedores virtuales, para ello fue necesario revisar e implementar en contenedores virtuales el proceso de cifrado de contenido digital (uso de algoritmo AES), el proceso de cifrado de la llave simétrica utilizando cifrado asimétrico (ABE) y el proceso de resumen digital (Hash).
  - Definición del esquema de paralelismo a utilizar: se investigó sobre los patrones de paralelismo basado en tareas y se realizó su implementación por medio de los contenedores virtuales creados en el punto anterior.
  - Definición de una estrategia para implementar un registro distribuido a través del uso de la tecnología de cadena de bloques.
  - Creación de mecanismos para la unión del Sobre Digital y el uso de cadenas de bloques.

- Etapa 3: Evaluación experimental

  La metodología que se siguió para realizar la evaluación experimental del método propuesto fue dividida en cuatro etapas:

  - En la primera etapa se realizó el desarrollo del prototipo que se utilizó durante la experimentación.
  - En la segunda etapa se definieron las configuraciones del prototipo, las métricas y el ambiente donde se llevó acabo la experimentación.
  - En la tercera etapa se realizó una evaluación experimental controlada para validar el desarrollo del método propuesto y el prototipo.
  - En la etapa final, se llevó a cabo la evaluación de un caso de estudio basado en escenarios de compartición de datos médicos y satelitales.

- Etapa 4: análisis de resultados

  Esta tarea consistió en la interpretación de los resultados obtenidos de la experimentación del método propuesto, mediante una exploración estadística de las métricas mencionadas anteriormente.

- Etapa 5: redacción de Tesis

  A lo largo del desarrollo de las etapas anteriores se llevo a cabo la redacción del documento de tesis donde se reportaron cada uno de los resultados obtenidos en cada una de las etapas mencionadas en este documento.

1.6 Organización de la tesis

Esta tesis contiene 5 capítulos que se encuentran organizados de la siguiente manera: En el Capítulo 1 se da la introducción y el contexto del trabajo de tesis. En el Capítulo 2 se describe el fundamento teórico necesario para el desarrollo de esta tesis, de igual manera se presenta la revisión del estado del arte relacionado a cifrado de datos, patrones de paralelismo y cadena de bloques. En el Capítulo 3 se describe el método propuesto detallando cada una de las fases que lo componen, permitiendo de esta forma alcanzar con los objetivos. En el Capítulo 4 se describe el diseño experimental y se discuten los resultados obtenidos. Finalmente en el Capítulo 5 se da la conclusión de este trabajo de tesis, el cual contiene las contribuciones, limitaciones y trabajo futuro.

2

Marco Teórico y Estado del Arte

En el presente capítulo se aborda el marco conceptual necesario para el desarrollo de esta tesis, se describen los conceptos de cadena de valor, seguridad informática, criptografía, computo en la nube, patrones de paralelismo, contenedores y la cadena de bloques. Así mismo se revisan los trabajos mas representativos sobre criptografía, y cadena de bloques, para el aseguramiento de archivos sensibles.

2.1 Cadenas de valor

El concepto de "cadena de valor" fue introducido por Porter [29] (1985) para describir la serie completa de actividades que se requieren para llevar un producto o servicio desde la creación, a través de las diferentes etapas de producción, distribución a los consumidores y disposición final después del uso. A medida que el producto se mueve de una posición de la cadena a otra, va adquiriendo valor.

En la Figura 2.1 se muestra un ejemplo de una cadena de valor de la información [12], donde se tienen diferentes etapas y en cada una de ellas se va agregando valor.

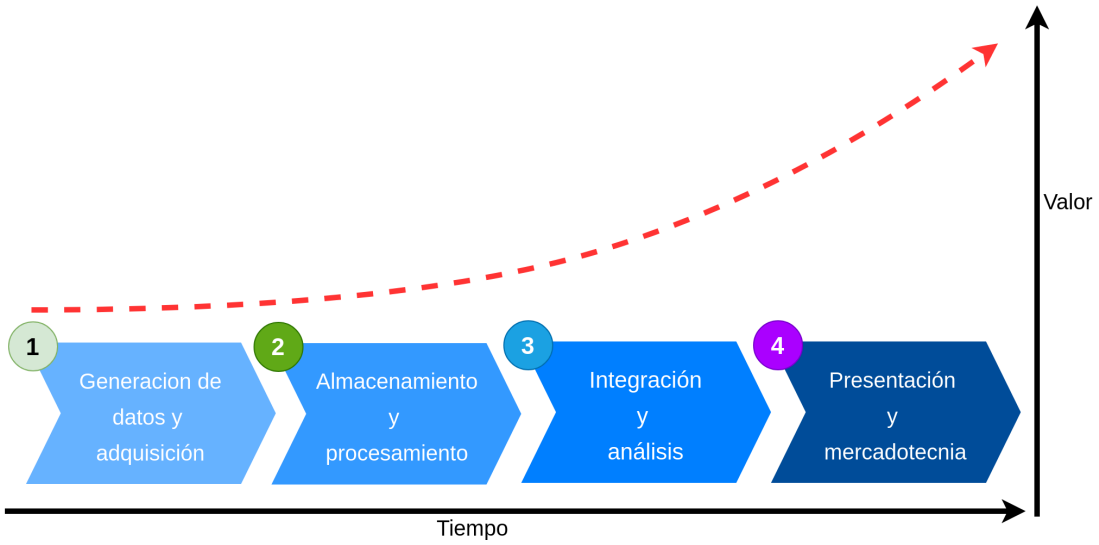

Figura 2.1: Cadena de valor de la información.

Las cadenas de valor se pueden encontrar en diferentes ámbitos como es el ámbito médico [28], [26], (se puede observar un ejemplo en Figura 2.2) en el cual está enfocado este trabajo de tesis, centrándose en las cadenas generadas por productos digitales (tomografías, radiografías, etc).

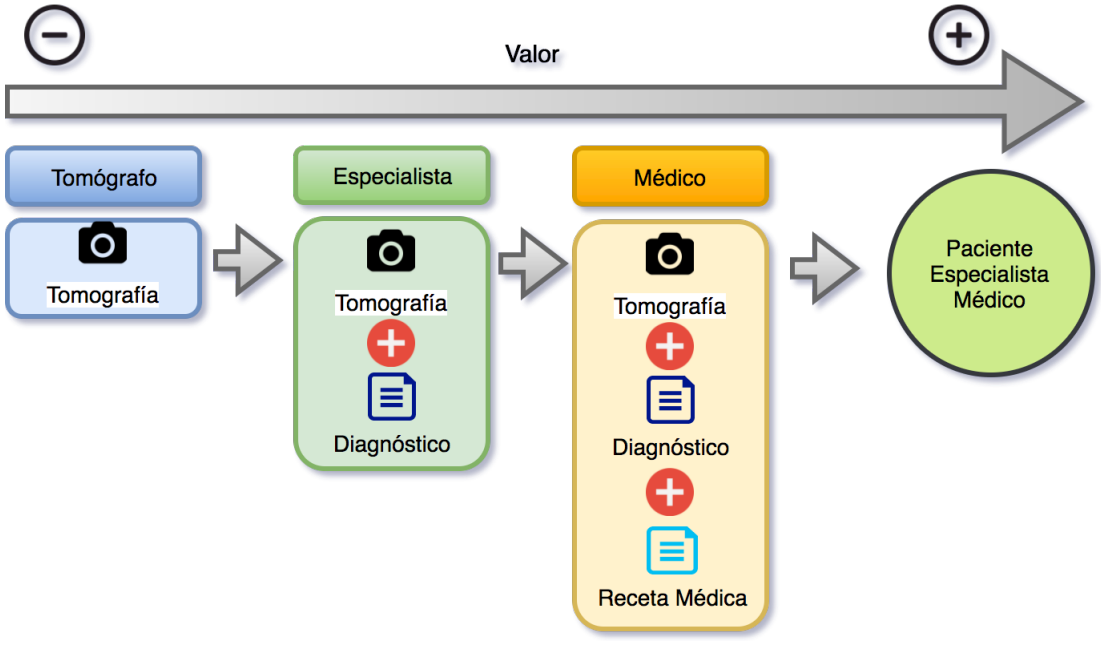

Figura 2.2: Cadena de valor en el ámbito médico.

2.2 Seguridad de la información

En general, la seguridad es la calidad o el estado de estar seguro o estado libre de peligro. En otras palabras, el objetivo es la protección contra acciones realizadas por adversarios que podrían dañar, intencionalmente o sin intención [40] a una entidad. La seguridad por tanto puede variar dependiendo de lo que se proteja, por ejemplo podemos tener seguridad física, personal, de operaciones, comunicación, redes, información, siendo esta última de interés para este tema de tesis.

El Comité de Sistemas de Seguridad Nacional (CNSS, Committee on National Security Systems) define la seguridad de la información como la protección de la información y sus elementos críticos, incluidos los sistemas y el hardware que usan, almacenan y transmiten esa información; sirve para para proteger la confidencialidad, integridad y disponibilidad de los activos de información, ya sea en almacenamiento, procesamiento o transmisión. Se logra mediante la aplicación de políticas, educación, capacitación y sensibilización, y tecnología.

La seguridad de la información se manifiesta de muchas maneras según la situación y los requisitos. Independientemente de quién esté involucrado, en un grado u otro, todas las partes en una transacción deben tener la confianza de que se han cumplido ciertos objetivos (como los arriba mencionados) asociados con la seguridad de la información.

A lo largo de los siglos, se ha creado un elaborado conjunto de protocolos y mecanismos para tratar los problemas de seguridad de la información [7]. Especialmente enfocados a escenarios cuando la información es transmitida por documentos físicos. A menudo, los objetivos de seguridad de la información no pueden lograrse únicamente a través de algoritmos y protocolos matemáticos solos, sino que requieren técnicas de procedimiento y el cumplimiento de leyes para lograr el resultado deseado [25].

Conceptualmente, la forma en que se registra la información no ha cambiado drásticamente con el tiempo. Mientras que la información generalmente se almacenaba y transmitía en papel, gran parte de ella ahora reside en medios magnéticos y se transmite a través de sistemas de telecomunicaciones.

Lo que ha cambiado dramáticamente es la capacidad de copiar y alterar información. Uno puede hacer miles de copias idénticas de una información almacenada electrónicamente y cada una es indistinguible del original. Realizar este tipo de tarea con información en papel, no resultaría trivial. Lo que se necesita entonces para una sociedad donde la información se almacena y transmite principalmente en forma electrónica es un medio para garantizar la seguridad de la información, que es independiente del medio físico que la graba o transmite, y que los objetivos de la seguridad de la información dependen únicamente de la información digital.

Lograr la seguridad de la información en una sociedad electrónica requiere una amplia gama de habilidades técnicas y legales. Sin embargo, no hay garantía de que se puedan cumplir adecuadamente todos los objetivos de seguridad de la información que se consideren necesarios para escenarios específicos [39] [15]. Es por lo anterior que los medios técnicos tradicionales para hacer frente a este tipo de problema se proporcionan a través de la criptografía y se aplican principalmente a datos llamados $sensibles$.

2.2.1 Datos sensibles

En el articulo 3 de la La ley de protección de datos personales en posesión de los particulares (LFPDPPP), define a los datos sensibles como: "aquellos datos personales que afecten a la esfera más íntima de su titular, o cuya utilización indebida pueda dar origen a discriminación o conlleve un riesgo grave para éste. En particular, se consideran sensibles aquellos que puedan revelar aspectos como origen racial o étnico, estado de salud presente y futuro, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas, preferencia sexual".

Los escenarios donde se manejan datos médicos [3] son especialmente propicios para generar una gran cantidad de datos sensibles (tomografías, radiografías, historiales, estudios etc.), los cuales representan el principal sujeto de estudio de esta tesis.

2.2.2 Criptografía

En término general, la criptografía es la ciencia de la escritura secreta con el objetivo de ocultar el significado de un mensaje [27], particularmente en la computación se le define como el estudio de técnicas matemáticas relacionadas con aspectos de la seguridad de la información, como la confidencialidad, la integridad de los datos, la autenticación de la entidad y la autenticación del origen de los datos [21].

La criptografía no es el único medio para proporcionar seguridad de la información, sino un conjunto de técnicas generalmente propuestas en la forma de servicios.A continuación se describen los tipos de servicios que se pueden brindar por medio de la criptografía [21] [27]:

- El servicio de *confidencialidad* es utilizado para mantener en secreto el contenido de la información de todos menos aquellos autorizado para tenerlo.

- La *integridad* de los datos es un servicio aborda la alteración no autorizada de los datos. Para garantizar la integridad de los datos, este servicio debe detectar la manipulación de datos por partes no autorizadas. La manipulación de datos incluye acciones tales como inserción, eliminación y sustitución de datos en un contenido o archivo.

- La *autenticación* es una propiedad que un servicio puede entregar mediante la identificación de aquellos que manejan datos sensibles. Este servicio se aplica tanto a las entidades como a la información misma. Dos partes que inician una comunicación deben identificarse entre sí. La información entregada a través de un canal debe autenticarse en cuanto a origen, fecha de origen, contenido de datos, hora de envío, etc. Por estas razones, este aspecto de la criptografía generalmente se subdivide en dos clases principales: autenticación de entidad que se crea mediante el uso de credenciales (e.g. contraseñas, certificados, etc), mientras que la autenticación de origen de datos se proporciona implícitamente mediante la determinación de la integridad de los datos (al modificar un mensaje, la fuente cambiado, se pierde la integridad

y por tanto no se puede identificar un dato).

- El *no repudio* es un servicio que evita que una entidad niegue compromisos o acciones anteriores. Cuando surgen disputas debido a que una entidad niega que se hayan tomado ciertas acciones, es necesario un medio para resolver la situación. Por ejemplo, una entidad puede autorizar la compra de propiedades por otra entidad y luego negar que se haya otorgado dicha autorización. Se necesita un procedimiento que involucre a un tercero de confianza para resolver la disputa.

2.2.2.1. Criptografía simétrica

La criptografía simétrica (en inglés symmetric key cryptography), también llamada criptografía de clave secreta (en inglés secret key cryptography) es un método criptográfico en el cual se usa una misma clave para cifrar y descifrar mensajes. Las dos partes que se comunican (Alicia, Benito) han de ponerse de acuerdo de antemano sobre la llave secreta k a usar. Una vez que ambas partes tienen acceso a esta llave secreta k, el remitente cifra un mensaje m con al algoritmo de cifrado E usando la llave secreta k, obteniendo el mensaje cifrado $c = E(k, m)$ y lo envía al destinatario, éste lo descifra con la misma llave secreta k [16]

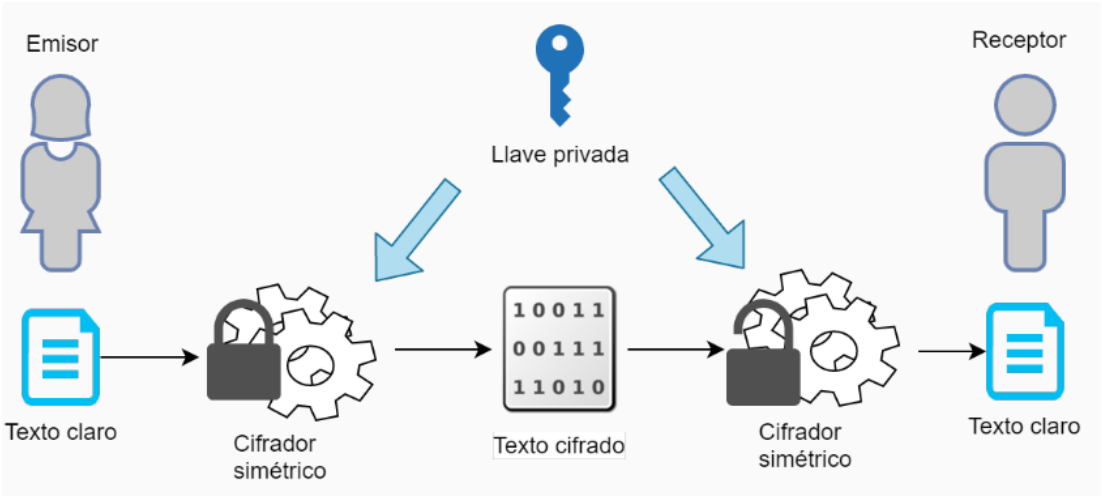

Figura 2.3: Criptografía simétrica.

Estándar de cifrado avanzado AES (por sus siglas en ingles Advance Encryption Standar) es un algoritmo de cifrado simétrico desarrollado por los estudiantes Vincent Rijmen y Joan Daemen de la Katholieke Universiteit Leuven en Bélgica, bajo el nombre "Rijndael" fue presentado en 1997 al concurso organizado por el Instituto Nacional de Normas y Tecnologías (NIST) para elegir el mejor algoritmo de cifrado; el algoritmo gano el concurso transformándose en un estándar en el año 2002, con algunos cambios fue posteriormente renombrado AES (Advanced Encryption Standard) y se convirtió en uno de los algoritmos más utilizados en la actualidad.

AES es un algoritmo de cifrado por bloques y contiene diferentes modos de operación entre ellos se encuentra el CBC (cipher block chaining).

El cifrado por bloques (block cipher en inglés) es una unidad de cifrado de clave simétrica que opera en grupos de bits de longitud fija, llamados bloques, aplicándoles una transformación invariante. Cuando realiza cifrado, una unidad de cifrado por bloques toma un bloque de texto plano o claro como entrada y produce un bloque de igual tamaño de texto cifrado. La transformación exacta es controlada la clave secreta. El descifrado es similar: se ingresan bloques de texto cifrado y se producen bloques de texto plano.

En el modo cipher-block chaining (CBC), a cada bloque de texto plano se le aplica la operación lógica XOR con el bloque cifrado anterior antes de ser cifrado. De esta forma, cada bloque de texto cifrado depende de todo el texto en claro procesado hasta este punto. Para hacer cada mensaje único se utiliza asimismo un vector de inicialización (iv).

AES es obligatorio en varios estándares de la industria y se usa en muchos sistemas comerciales. Entre los estándares comerciales que incluyen AES se encuentran el estándar de seguridad de Internet IPsec, TLS, el estándar de cifrado de Wi-Fi IEEE 802.11i, el protocolo de red de shell seguro SSH (Secure Shell), el teléfono de Internet Skype y numerosos productos de seguridad en todo el mundo. Hasta la fecha, no hay ataques mejores que la fuerza bruta conocida contra AES.

Los algoritmos simétricos como AES son muy seguros, rápidos y de uso generalizado. Sin embargo, existen deficiencias asociadas con los esquemas de clave simétrica, a continuación se describen

algunos.

- **Problema de distribución de claves**: La clave debe establecerse entre Alicia y Benito utilizando un canal seguro. El enlace de comunicación para el mensaje no es seguro, por lo que no se puede enviar la clave directamente por el canal, que sería la forma más conveniente de transportarlo.

- **Número de claves** En una red grande por ejemplo con 2000 personas, esto requiere más de 4 millones de pares de claves que deben generarse y transportarse a través de canales seguros, esto genera el problema de administrar esa cantidad de claves.

Una forma en la que se han resuelto estas deficiencias es por medio de la criptografía asimétrica la cual se describe a continuación.

2.2.2.2. Criptografía asimétrica

La criptografía asimétrica (en inglés asymmetric key cryptography), también llamada criptografía de clave pública (en inglés public key cryptography) o criptografía de dos claves (en inglés two-key cryptography), es el método criptográfico que usa un par de claves para el envío de mensajes. Las dos claves pertenecen a la misma persona que ha enviado el mensaje. Una clave es pública y se puede entregar a cualquier persona, la otra clave es privada y el propietario debe guardarla de modo que nadie tenga acceso a ella. Además, los métodos criptográficos garantizan que esa pareja de claves sólo se puede generar una vez, de modo que se puede asumir que no es posible que dos personas hayan obtenido casualmente la misma pareja de claves. Si el remitente usa la clave pública del destinatario para cifrar el mensaje, una vez cifrado, sólo la clave privada del destinatario podrá descifrar este mensaje, ya que es el único que la conoce. Por tanto se logra la confidencialidad del envío del mensaje, nadie salvo el destinatario puede descifrarlo.

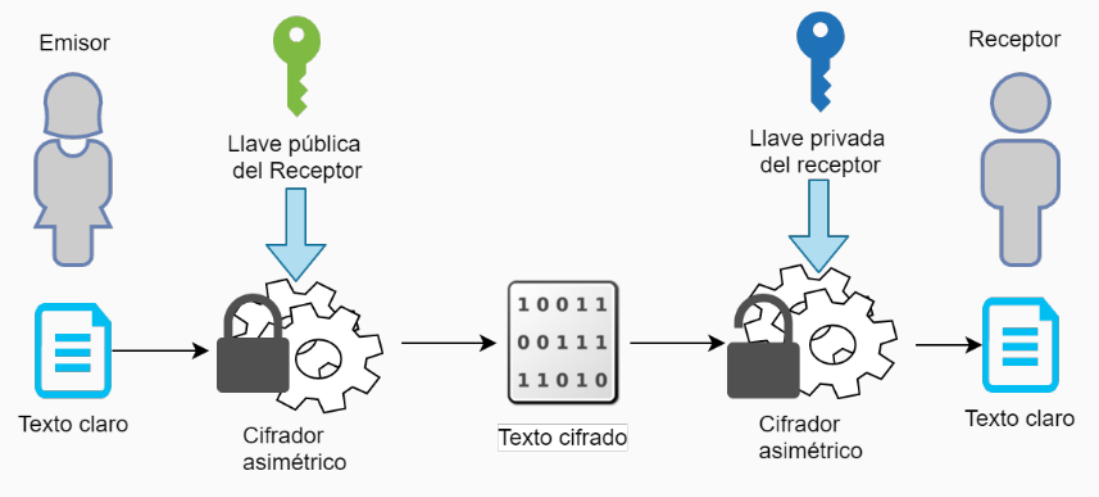

Figura 2.4: Criptografía asimétrica.

En [31] se introduce el término cifrado basado en atributos (ABE, Attribute-based Encryption) como un nuevo medio para el control de acceso, En un sistema de criptografía basado en atributos, los textos cifrados no están necesariamente cifrados para un usuario particular como en la criptografía de clave pública tradicional. En cambio, las claves privadas y los textos cifrados de ambos usuarios se asociarán con un conjunto de atributos o una política sobre tributos. Un usuario puede descifrar un texto cifrado si hay una coincidencia entre su clave privada y el texto de cifrado.

CP-ABE presentado en [6] las claves privadas se identifican con un conjunto S de atributos. la parte que desee descifrar un mensaje especificará a través de una estructura de acceso de árbol una política que las claves privadas deben cumplir. Cada nodo interior del árbol es una puerta de umbral y las hojas están asociadas con atributos. Podemos representar un árbol con compuertas "AND" y "OR" , esta compuertas también puedes representarse por compuertas de umbral usándose $nofn$ como AND y $1ofn$ como OR. Un usuario podrá descifrar un texto cifrado con una clave dada si y solo si hay una asignación de atributos de la clave privada a los nodos del árbol de modo que el árbol esté satisfecho.

En la Figura 2.5 se observan dos usuarios (Benito y Alicia), y cada uno con sus atributos (Paciente, Hospital), para acceder al texto cifrado se tiene que satisfacer la política, en este caso el paciente

Alicia del Hospital general o que sea cardiólogo del Hospital general, por lo tanto Benito no puede acceder al texto, caso contrario, Alicia satisface la política y tiene acceso al texto cifrado.

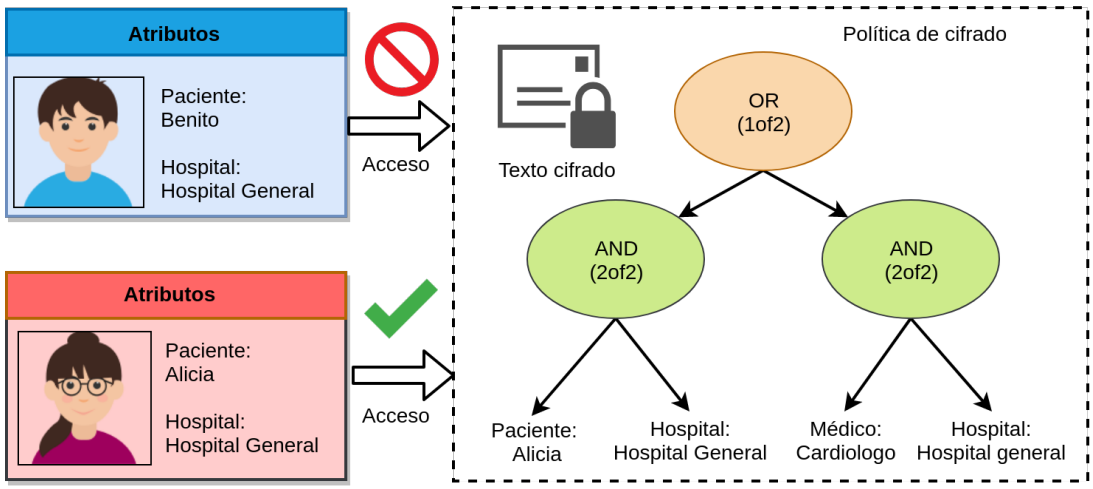

Figura 2.5: Cifrado basado en atributos.

CP-ABE contiene los siguientes algoritmos:

Configuración. El algoritmo de configuración no toma ninguna entrada que no sea el parámetro de seguridad implícito. Emite los parámetros públicos PK y una clave maestra MK.

Cifrado (PK; M; A). El algoritmo de cifrado toma como entrada los parámetros públicos PK, un mensaje M y una estructura de acceso A sobre el universo de atributos. El algoritmo cifrara M y producirá un texto cifrado CT de tal manera que solo un usuario que posea un conjunto de atributos que satisfaga la estructura de acceso podrá descifrar el mensaje.

Generación clave (MK; S). El algoritmo de generación de claves toma como entrada la clave maestra MK y un conjunto de atributos S que describen la clave. Emite una clave privada SK.

Descifrado (PK; CT; SK). El algoritmo de descifrado toma como entrada los parámetros públicos PK, un texto cifrado CT, que contiene una política de acceso A, y una clave privada SK, que es una clave privada para un conjunto S de atributos. Si el conjunto S de atributos satisface la estructura de acceso A, el algoritmo descifrará el texto cifrado y devolverá un mensaje M.

Aunque la criptografía asimétrica resuelve algunas deficiencias de la criptografía simétrica, el

rendimiento para los métodos de cifrado de clave pública (asimétricos) son varios órdenes de magnitud más lentos que los esquemas de clave simétrica más conocidos.

En resumen, los puntos importantes en la práctica de la criptografía simétrica y asimétrica son:

1. La criptografía de clave pública facilita las firmas eficientes (particularmente el no repudio) y la gestión de claves; y

2. La criptografía de clave simétrica es eficiente para el cifrado y algunas aplicaciones de integridad de datos

Estos puntos son importantes para entender el concepto de Sobre digital, el cual se describe más adelante en la sección 2.2.2.5

2.2.2.3. Función Hash

Una función hash (también llamada función picadillo o resumen), es una función h que tiene, como mínimo, las dos propiedades siguientes:

- Compresión: h asigna una entrada x de longitud de bits finita arbitraria, a una salida $h(x)$ de longitud de bits fija n.

- Facilidad de cálculo: dado h y una entrada x, $h(x)$ es fácil de calcular.

A continuación se describen las propiedades adicionales de la función Hash:

- Resistencia de preimagen para una salida dada z, es imposible encontrar una entrada x tal que $h(x) = z$, es decir, $h(x)$ es unidireccional.

- Segunda resistencia previa a la imagen Dado x_1, y por lo tanto $h(x_1)$ es computacionalmente encontrar cualquier x_2 tal que $h(x_1) = h(x_2)$.

- Resistencia a colisiones Es computacionalmente inviable encontrar pares $x_1 = x_2$ tal que $h(x_1) = h(x_2)$ [27].

Las funciones Hash proveen el servicio de integridad de los datos, y el uso criptográfico más común es con las firmas digitales, que se describen a continuación.

2.2.2.4. Firma digital

La firma digital es una cadena de datos (resumen) que asocia un mensaje (en forma digital) con alguna entidad de origen y provee los servicios de autenticación y principalmente, no repudio. Para generar la firma digital primero se obtiene un resumen de la información electrónica que se firmará usando un algoritmo hash, el cual aplica una función unidireccional a cada bit del mensaje o documento electrónico y produce como salida una cadena binaria, que puede interpretarse como la huella digital de los bits de entrada. La función hash es tal que a partir del resumen o huella digital es prácticamente imposible poder deducir el mensaje o documento electrónico que lo produce. Esta última aseveración depende del número de bits que se usen para representar al resumen o huella digital que la función hash produce. El actual estándar para calcular funciones hash es la familia SHA-2, donde el resumen del mensaje puede ser de entre 200 a 600 bits. La cadena binaria correspondiente al resumen del documento (hash) se cifra con la llave privada del firmante 2.6, resultando una nueva cadena binaria que representa la firma digital del documento. Entonces el documento junto con la firma se distribuye o almacena. Después, para realizar el proceso de verificación, se descifra la cadena binaria correspondiente a la firma digital usando la llave pública del firmante. Este valor descifrado debe corresponder al valor hash original del archivo firmado. Entonces, solo resta calcular nuevamente el valor hash del documento y compararlo con el valor resultante del descifrado. Si los valores coinciden, la firma digital es considerada auténtica, de lo contrario, la firma es rechazada, por lo que quién verifica la firma considera como inválido el documento, ya que éste o ha sufrido modificaciones y no corresponde al mensaje/documento originalmente firmado, o se está intentando verificar la firma con la llave pública de un usuario distinto al que firmó el documento.

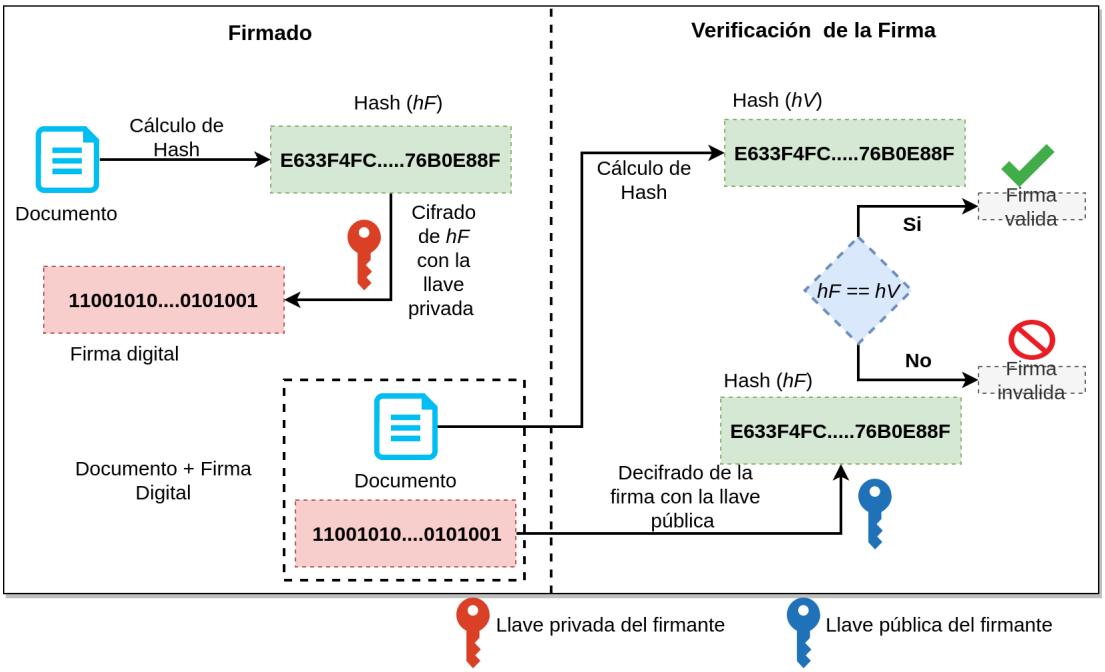

Figura 2.6: Proceso de firmado y verificación de firma digital.

2.2.2.5. Integración de servicios criptográficos: Técnica de sobre digital

El sobre digital es un objeto criptográfico que sirve para el intercambio seguro de la clave de un cifrador simétrico, explotando las ventajas de la criptografía simétrica y asimétrica, como se menciono en la sección **??**, la criptografía de clave pública facilita la gestión de claves y la criptografía de clave simétrica es eficiente para el cifrado.

Con un sobre digital se pueden garantizar las propiedades de confidencialidad de un archivo y se genera a partir del cifrado simétrico de un documento d y con una llave secreta k, denotado como $S_k(d)$, luego la llave secreta k se cifra asimétricamente con la llave pública k_2 de la persona a la que le vamos a enviar el sobre digital, esto se denota como $A_{k2}(k)$ y finalmente el sobre digital se forma concatenando $S_k(d)$ y $A_{k2}(k)$.

2.3 Cómputo distribuido

En esta sección se describe como se implementan los sistemas de seguridad informática en escenarios reales.

2.3.1 Cómputo y almacenamiento en la nube

La computación en la nube, es un servicio que resulta de la manipulación del hardware realizada por virtualización. La virtualización es la técnica que separa la arquitectura física para crear varios recursos dedicados, de modo que podamos utilizar la eficiencia de nuestro hardware por completo. Hoy en día se combinan estas dos tecnologías para obtener los mejores resultados, es decir, utilizando el entorno virtualizado sobre la nube. La virtualización le permite ejecutar más de un sistema operativo y ejecutar múltiples aplicaciones en el mismo servidor.

Según el NIST el cómputo en la nube [20] es un modelo para permitir el acceso de red ubicuo, conveniente y bajo demanda a un grupo compartido de recursos informáticos configurables (por ejemplo, redes, servidores, almacenamiento, aplicaciones y servicios) que se pueden aprovisionar y liberar rápidamente con un mínimo esfuerzo de gestión o Interacción del proveedor de servicios. Este modelo de nube se compone de cinco características esenciales, tres modelos de servicio y cuatro modelos de implementación.

Las características esenciales son:

- Auto servicio bajo demanda. Un consumidor puede aprovisionar unilateralmente capacidades informáticas, como la hora del servidor y el almacenamiento en red, según sea necesario automáticamente sin requerir la interacción humana con cada proveedor de servicios.

- Amplio acceso a la red. Las capacidades están disponibles a través de la red y se accede a ellas a través de mecanismos estándar que promueven el uso de plataformas heterogéneas de

clientes delgados o gruesos (por ejemplo, teléfonos móviles, tabletas, computadoras portátiles y estaciones de trabajo).

- Piscina de recursos. Los recursos informáticos del proveedor se agrupan para servir a múltiples consumidores utilizando un modelo de múltiples inquilinos, con diferentes recursos físicos y virtuales asignados y reasignados dinámicamente de acuerdo con la demanda del consumidor. Existe una sensación de independencia de ubicación en el sentido de que el cliente generalmente no tiene control o conocimiento sobre la ubicación exacta de los recursos proporcionados, pero puede ser capaz de especificar la ubicación en un nivel superior de abstracción (por ejemplo, país, estado o centro de datos). Ejemplos de recursos incluyen almacenamiento, procesamiento, memoria y ancho de banda de red.

- Rápida elasticidad. Las capacidades se pueden aprovisionar y liberar elásticamente, en algunos casos automáticamente, para escalar rápidamente hacia afuera y hacia adentro de acuerdo con la demanda. Para el consumidor, las capacidades disponibles para el aprovisionamiento a menudo parecen ser ilimitadas y pueden ser apropiadas en cualquier cantidad en cualquier momento.

- Servicio medido. Los sistemas en la nube controlan y optimizan automáticamente el uso de recursos al aprovechar una capacidad de medición1 en algún nivel de abstracción apropiado para el tipo de servicio (por ejemplo, almacenamiento, procesamiento, ancho de banda y cuentas de usuario activas). El uso de los recursos se puede monitorear, controlar e informar, proporcionando transparencia tanto para el proveedor como para el consumidor del servicio utilizado.

Modelos de servicios:

- Software como servicio (SaaS). La capacidad proporcionada al consumidor es utilizar las aplicaciones del proveedor que se ejecutan en una infraestructura en la nube2. Se puede

acceder a las aplicaciones desde varios dispositivos cliente a través de una interfaz de cliente ligero, como un navegador web (por ejemplo, correo electrónico basado en la web) o una interfaz de programa. El consumidor no gestiona ni controla la infraestructura de nube subyacente, incluidas las redes, los servidores, los sistemas operativos, el almacenamiento o incluso las capacidades de aplicaciones individuales, con la posible excepción de los ajustes de configuración de aplicaciones específicas del usuario.

- Plataforma como servicio (PaaS). La capacidad proporcionada al consumidor es desplegar en la infraestructura de la nube aplicaciones creadas o adquiridas por el consumidor creadas usando lenguajes de programación, bibliotecas, servicios y herramientas compatibles con el proveedor.3 El consumidor no administra ni controla la infraestructura de nube subyacente, incluida la red , servidores, sistemas operativos o almacenamiento, pero tiene control sobre las aplicaciones implementadas y posiblemente las configuraciones para el entorno de alojamiento de aplicaciones.

- Infraestructura como servicio (IaaS). La capacidad que se brinda al consumidor es el aprovisionamiento de procesamiento, almacenamiento, redes y otros recursos informáticos fundamentales donde el consumidor puede implementar y ejecutar software arbitrario, que puede incluir sistemas operativos y aplicaciones. El consumidor no administra ni controla la infraestructura de nube subyacente, pero tiene control sobre los sistemas operativos, el almacenamiento y las aplicaciones implementadas; y posiblemente un control limitado de componentes de red seleccionados (por ejemplo, firewalls de host).

Modelos de despliegue:

- Nube privada. La infraestructura de la nube está aprovisionada para uso exclusivo de una sola organización que comprende múltiples consumidores (por ejemplo, unidades de negocios). Puede ser propiedad, ser administrado y operado por la organización, un tercero o alguna combinación de ellos, y puede existir dentro o fuera de las instalaciones.

- Nube comunitaria. La infraestructura de la nube está provista para uso exclusivo de una comunidad específica de consumidores de organizaciones que tienen preocupaciones compartidas (por ejemplo, misión, requisitos de seguridad, políticas y consideraciones de cumplimiento). Puede ser propiedad, administrada y operada por una o más de las organizaciones de la comunidad, un tercero o alguna combinación de ellas, y puede existir dentro o fuera de las instalaciones.

- Nube pública. La infraestructura de la nube está aprovisionada para uso abierto por el público en general. Puede ser propiedad, ser administrado y operado por una organización comercial, académica o gubernamental, o alguna combinación de ellos. Existe en las instalaciones del proveedor de la nube.

- Nube Híbrida. La infraestructura de la nube es una composición de dos o más infraestructuras de nube distintas (privadas, comunitarias o públicas) que siguen siendo entidades únicas, pero están unidas por una tecnología estandarizada o patentada que permite la portabilidad de datos y aplicaciones (por ejemplo, la explosión de la nube para el equilibrio de carga entre nubes).

2.3.2 Contenedores virtuales

La implementación de la computación en la nube en formas tradicionales se realiza utilizando máquinas virtuales, pero hoy en día un nuevo concepto de contenedores también está ganando popularidad debido a sus características. Los contenedores virtuales en algunos casos son tratados como una técnica de virtualización ligera.

Los contenedores son plataformas para desarrollar e implementar aplicaciones ajenas a la infraestructura. La metodología de contenedor permite a los desarrolladores realizar una implementación rápida y reduce significativamente el retraso entre escribir el código y tenerlo en producción. Los contenedores brindan la capacidad de empaquetar y ejecutar una aplicación en un entorno aislado.

Los contenedores requiere menos recursos informáticos y menos gastos de virtualización en comparación con las máquinas virtuales, [13] de la misma manera es posible implementar un mayor número de contenedores que de máquinas virtuales en la misma maquina física [1, 8], a su vez que permiten una mejor portabilidad e interoperabilidad [41].

En la Figura 2.8 se muestra la comparación de las arquitecturas de una máquina virtual y un contenedor, el objetivo de una máquina virtual es aislar el sistema operativo huésped, en cambio en los contenedores están pensados para aislar aplicaciones, Cada contenedor no solo comparte el sistema operativo anfitrión (no necesitan un sistema operativo huésped) sino también las bibliotecas y los archivos binarios que son necesarios para que la aplicación se ejecute. Todos los componentes compartidos son generalmente de solo lectura, esto reduce el tamaño de los mismo, el tiempo de despliegue, y el el gasto en la comunicación desde las aplicaciones al sistema operativo anfitrión es menor que en las maquinas virtuales.

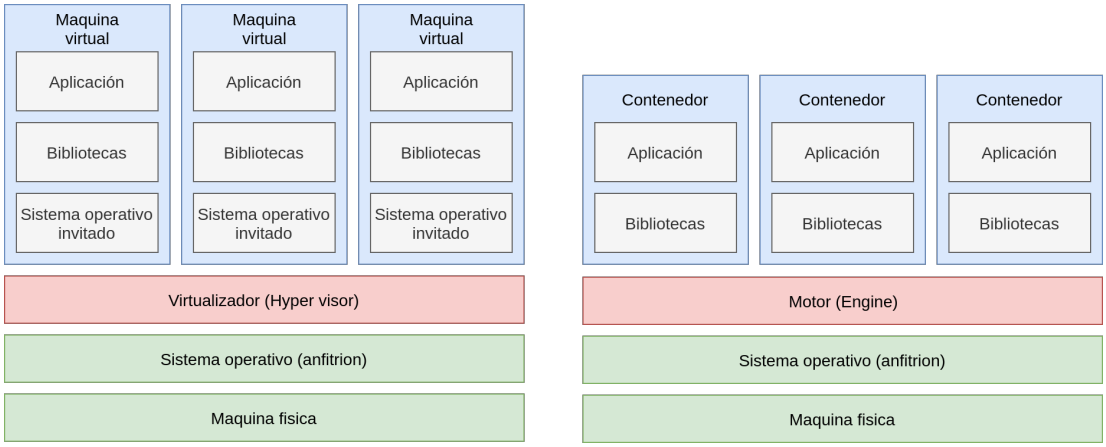

Figura 2.7: Máquinas virtuales vs contenedores.

2.3.2.1. Docker

Docker es una plataforma de código abierto para desarrollar, enviar y ejecutar aplicaciones, permite separar sus aplicaciones de su infraestructura para que pueda entregar software rápidamente. Con Docker, puede administrar su infraestructura de la misma manera que administra sus

aplicaciones. Al aprovechar las metodologías de Docker para enviar, probar e implementar código rápidamente, puede reducir significativamente el retraso entre escribir el código y ejecutarlo en producción.

Docker ofrece la capacidad de empaquetar y ejecutar una aplicación en un entorno aislado llamado contenedor. El aislamiento y la seguridad le permiten ejecutar muchos contenedores simultáneamente en una máquina determinada. Los contenedores son livianos porque no necesitan la carga adicional de un hipervisor, sino que se ejecutan directamente dentro del núcleo de la máquina anfitrión. Esto significa que puede ejecutar más contenedores en una combinación de hardware determinada que si estuviera utilizando máquinas virtuales. Incluso puede ejecutar contenedores Docker dentro de una máquina virtual, de esta forma el contenedor se convierte en la unidad para distribuir y probar sus aplicaciones.

2.4 Cómputo paralelo basado en Patrones

EL cómputo paralelo involucra el uso de múltiples recursos de cómputo de manera simultanea para resolver un solo problema de computo, el cual es dividido en múltiples series de instrucciones que puedes ser ejecutadas simultáneamente en diferentes procesadores, el resultado esperado es el cómputo mas rápido comparado con la ejecución en un solo procesador. El cómputo paralelo es tradicionalmente asociado al cómputo de alto rendimiento.

La estructura de los múltiples recursos a utilizar se logra por medio de patrones arquitecturales, los cuales se pueden describir como una estructura organizacional básica, comúnmente usada en cómputo paralelo, la selección de el patrón se basa en el particionamiento propuesto para algoritmos y datos.

2.4.1 Tuberías

Una tubería (pipeline) es una secuencia lineal de etapas. los flujos de datos fluyen a través de la tubería, de la primer etapa a la ultima etapa y la salida de una etapa es la entrada de la siguiente, cada etapa realiza una transformación en los datos [19].

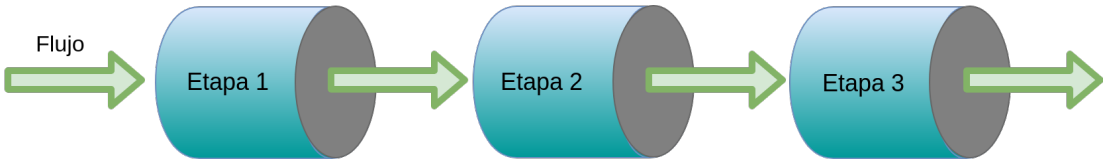

Figura 2.8: Tubería de procesamiento.

2.4.2 Paralelismo de tareas

El paralelismo de tareas (task parallel) es una forma de paralelización de código de computadora a través de múltiples procesadores en entornos de cómputo paralelo. El paralelismo de tareas se centra en la distribución de tareas, realizadas simultáneamente por procesos o subprocesos, en diferentes procesadores. A diferencia del paralelismo de datos que implica ejecutar la misma tarea en diferentes componentes de datos, el paralelismo de tareas se distingue al ejecutar muchas tareas diferentes al mismo tiempo en los mismos datos. Un tipo común de paralelismo de tareas es la canalización que consiste en mover un solo conjunto de datos a través de una serie de tareas separadas donde cada tarea puede ejecutarse independientemente de las demás.

2.4.3 Manejador-trabajador

Patron manejador-trabajador (Figura 2.9 Manager-Worker) considerando un enfoque de paralelismo de actividad donde las mismas operaciones se realizan en datos ordenados. Cada componente de procesamiento realiza simultáneamente las mismas operaciones, independientemente de la actividad de procesamiento de otros componentes. El manejador tiene la responsabilidad de

crear un número de trabajadores, repartir el trabajo entre ellos e iniciar la ejecución. Los trabajadores actúan como componentes de procesamiento, son idénticos entre ellos, tiene acceso a diferentes datos y las operaciones de cada uno son independientes de los otros.

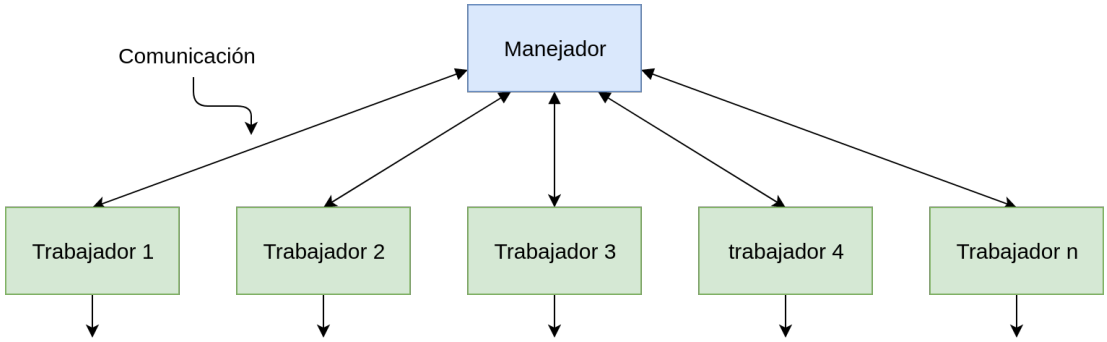

Figura 2.9: Patrón manejador-trabajador.

Beneficios:

1. Cada trabajador solicita una una parte diferente de los datos, esto forma una estructura que se muestra como un balanceo de carga natural.

2. Es posible cambiar o agregar nuevos trabajadores sin realizar cambios significativos en el manejador.

3. Es fácil de sincronizar, ya que la comunicación se realiza solo entre el manejador y cada uno de los trabajadores y es controlada por el.

4. Si se realiza de forma cuidadosa, el patrón manejador-trabajador ha demostrado aumentar el rendimiento sin cambios significativos en la implementación.

Consideraciones:

1. Puede presentar problemas de rendimiento si el número de trabajadores es grande y el procesamiento es muy simple o si los trabajadores reciben una pequeña cantidad de datos, esto generaría que algunos estén ociosos mientras el manejador trata de repartir el trabajo.

2. Puede generar rendimiento pobre si las tareas de procesamiento del manejador tardan mas que las tareas de los trabajadores.

3. Un bajo rendimiento en el manejador afecta el rendimiento de todo el sistema.

4. El punto importante es encontrar el número de trabajadores con el que se logra el rendimiento óptimo.

2.5 Sinergia entre servicios criptográficos y sistemas distribuidos: Cadena de bloques (Blockchain)

En el año 2008, un individuo o grupo que escribe bajo el nombre de Satoshi Nakamoto(nadie conoce a Satoshi Nakamoto hasta el día de hoy) publicó un trabajo titulado Bitcoin: un sistema de efectivo electrónico punto a punto [24]. Este documento describió una versión punto a punto del efectivo electrónico que permitiría que los pagos en línea se envíen directamente de una parte a otra sin pasar por una institución financiera. Bitcoin fue la primera realización de este concepto. Unos meses más tarde, se lanzó un programa de código abierto que implementa el nuevo protocolo que comenzó con el bloque Génesis de 50 monedas. Cualquiera puede instalar este programa de código abierto y formar parte de la red punto a punto. Ha crecido en popularidad desde entonces.

El sistema Bitcoin ordena las transacciones colocándolas en grupos llamados bloques y luego vinculando estos bloques a través de lo que se llama cadena de bloques (Blockchain). Se considera que las transacciones en un bloque ocurrieron al mismo tiempo. Estos bloques están vinculados entre sí (como una cadena) en un orden cronológico lineal adecuado con cada bloque que contiene el hash del bloque anterior.

La cadena de bloques o Blockchain es esencialmente una base de datos distribuida de registros o contabilidad pública de todas las transacciones o eventos digitales que se han ejecutado y compartido entre las partes participantes. Cada transacción en el libro público se verifica por consenso de

la mayoría de los participantes en el sistema. Y, una vez ingresado, la información nunca puede borrarse. La cadena de bloques contiene un registro cierto y verificable de cada transacción que se haya realizado. La cadena de bloques permite establecer un sistema para crear un consenso distribuido en el mundo digital. Esto permite a las entidades participantes saber con certeza que un evento digital sucedió al crear un registro irrefutable en un libro público [10].

La Figura 2.10 ilustra un ejemplo de una cadena de bloques. Con un hash de bloques anterior contenido en el encabezado del bloque, un bloque tiene solo un bloque padre. El primer bloque de una cadena de bloques se llama bloque de génesis.

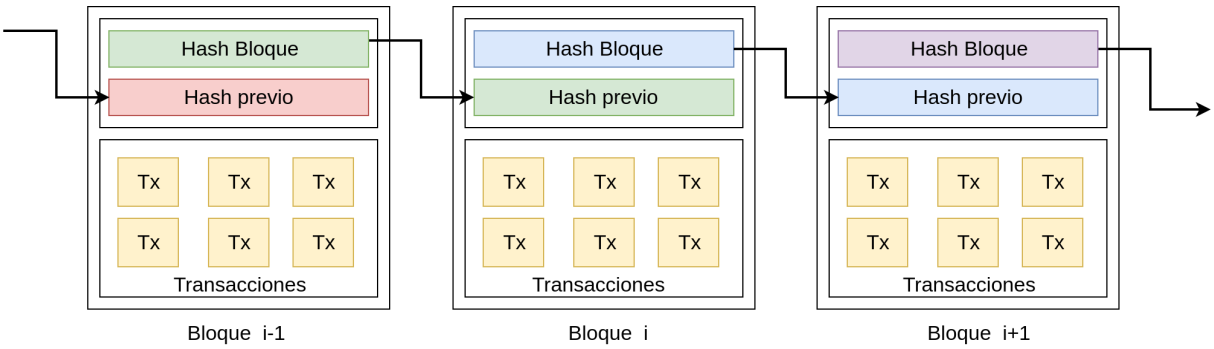

Figura 2.10: Ejemplo de la cadena de bloques

2.5.1 Características

Descentralización: En los sistemas de transacción centralizados convencionales, cada transacción necesita ser validada a través de una entidad central de confianza (por ejemplo, el banco central), lo que inevitablemente resulta en el costo y el rendimiento de los cuellos de botella en los servidores centrales. En contraste con el modo centralizado. En la cadena de bloques no se necesita de un tercero. Se utilizan algoritmos de consenso para mantener la consistencia de los datos en la red distribuida.

Persistencia: Las transacciones pueden validarse rápidamente y las transacciones inválidas no serían admitidas por los mineros honestos. Es casi imposible eliminar o deshacer transacciones una

vez que están incluidas en la cadena de bloques. Los bloques que contienen transacciones no válidas se pueden descubrir de inmediato.

Anonimato: Cada usuario puede interactuar con la cadena de bloques con una dirección generada, que no revela la identidad real del usuario.

Auditabilidad La cadena de bloques almacena todas las transacciones realizadas sobre un activo. Una vez que la transacción actual se registra en la cadena de bloques se pueden verificar y rastrear fácilmente.

2.5.2 Hyperledger-Fabrics

Hyperledger Fabric es una plataforma de tecnología de contabilidad distribuida autorizada (DLT, distributed ledger technology) de código abierto bajo la Fundación Linux, diseñada para su uso en contextos empresariales, asimismo tiene una arquitectura modular y configurable, que permite ser usada en una amplia gama de casos de uso de la industria, incluidos la banca, las finanzas, los seguros, la atención médica, los recursos humanos, la cadena de suministro etc.

Fabric admite contratos inteligentes creados en lenguajes de programación de propósito general como Java, Go y Node.js, en lugar de lenguajes restringidos específicos de dominio (DSL).

La plataforma Fabric también está autorizada, lo que significa que, a diferencia de una red pública sin permiso, los participantes se conocen entre sí, en lugar de ser anónimos y pueden no ser de confianza. Esto significa que si bien los participantes pueden no confiar completamente el uno en el otro (pueden ser, por ejemplo, competidores en la misma industria), una red puede funcionar bajo un modelo de gobernanza que se basa en la confianza que existe entre los participantes, como un acuerdo legal o marco para el manejo de disputas.

Esta plataforma cuenta con un soporte para protocolos de consenso conectables que permiten que la plataforma se personalice de manera más efectiva para adaptarse a casos de uso particulares y modelos de confianza, también puede aprovechar los protocolos de consenso que no requieren una criptomoneda nativa para incentivar la minería costosa o para impulsar la ejecución de contratos

inteligentes. Evitar una criptomoneda reduce algunos vectores de riesgo / ataque significativos, y la ausencia de operaciones de minería criptográfica significa que la plataforma se puede implementar con aproximadamente el mismo costo operativo que cualquier otro sistema distribuido.

2.6 Estado del arte

En esta sección se describen los avances mas importantes de la encontrados en la literatura relacionados a los tres pilares de este trabajo de tesis: seguridad, eficiencia y trazabilidad.

Yang et al. [43] realizan una revisión de las amenazas a la seguridad, confidencialidad, integridad y disponibilidad de los datos manejados en la Iniciativa de Medicina de Precisión (PMI, por sus siglas en inglés), enunciada en el año 2015 por el expresidente de los Estados Unidos, Barack Obama. En esta iniciativa se sugiere a las organizaciones participantes una forma de mitigar los desafíos de seguridad a través de una nueva arquitectura de sistema, en desarrollo en el Instituto de Tecnología de Massachusetts (MIT,Massachusetts Institute of Technology), conocida como proyecto OPAL Enigma. En este proyecto se crea una red de pares (P2P) que permite a las partes almacenar y analizar conjuntamente los datos con total privacidad. Una bitácora o libro mayor distribuido, auditable e inviolable (una cadena de bloques autorizada) registra y controla el acceso a través de contratos inteligentes e identidades digitales. OPAL Enigma visualiza la arquitectura del repositorio de datos distribuidos en una red P2P (peer-to-peer) donde los datos se cifran en su repositorio para que los datos sin procesar nunca se publiquen. Los datos permanecen seguros durante el almacenamiento y el análisis, porque se pueden consultar los datos, pero solo mediante consultas autorizadas por credenciales de identidad digitales para operaciones de datos específicas definidas por contratos inteligentes legalmente vinculantes. Se registra un registro inalterable y auditable de los patrones de comunicación entre los datos y los operadores, incluidas las credenciales y las operaciones de datos.

Azaria et al. [4] proponen un sistema descentralizado de gestión de registros para manejar los registros médicos electrónicos (EMR por sus siglas en ingles) llamado MedRec el cual utilizando la

tecnología de cadena de bloques *Ethereum*. En él brindan a los pacientes de un registro completo e inmutable y un acceso fácil a su información médica entre los proveedores y los sitios de tratamiento. También gracias al uso de la cadena de bloques, gestionan la autenticación, la confidencialidad, la responsabilidad y el intercambio de datos, que son cruciales cuando se maneja información confidencial. Así mismo incentivan a los interesados médicos (investigadores, autoridades de salud pública, etc.) para que participen en la red como mineros de la cadena de bloques. Esto les proporciona acceso a datos agregados y anónimos como recompensas mineras, a cambio de mantener y asegurar la red a través de un algoritmo de consenso distribuido llamado Prueba de Trabajo [14].

Dagher et al. [11] proponen Ancile, un marco de trabajo (framework) basado en cadena de bloques para el acceso seguro, interoperable y eficiente a los registros médicos por parte de pacientes, proveedores y terceros, al tiempo que se preserva la privacidad de la información confidencial de los pacientes. Ancile utiliza contratos inteligentes en una cadena de bloques basada en Ethereum para un mayor control de acceso y ofuscación de datos, y emplea técnicas criptográficas para mayor seguridad asignando el control de acceso a los usuarios.

Sukhodolskiy et al. [35] presentan un prototipo de sistema multi-usuario para control de acceso a conjuntos de datos almacenados en un entorno de nube no confiable. El enfoque proporciona un control de acceso a los datos almacenados en la nube sin la participación del proveedor. La herramienta principal del mecanismo de control de acceso es el esquema de cifrado basado en atributos. Utilizando cadena de bloques (Ethereum), el sistema proporciona un registro inmutable de todos los eventos de seguridad significativos, tales como generación de claves, asignación de políticas de acceso, cambio o revocación, solicitud de acceso. A su ves proponen un conjunto de protocolos criptográficos que garantizan la privacidad de las operaciones criptográficas que requieren claves secretas o privadas. Solo los códigos hash de los textos cifrados de se transfieren a través del cadena de bloques. El prototipo del sistema se implementa utilizando contratos inteligentes.

Harleen Kaur et al. [17] proponen un modelo de datos que integra el computo en la nube con la tecnología de cadena de bloques para almacenar , compartir, transferir y procesar datos médicos. El

modelo propuesto tiene tres componentes principales 1) Expertos de dominio (Doctores, técnicos de laboratorio, farmacéuticos etc.) 2) Proveedores de seguros de salud y 3) pacientes los cuales pueden acceder y/o generar datos en la nube usando la cadena de bloques, para ello se verifica su identidad usando credenciales criptográficas, una vez verificada su identidad pueden realizar peticiones de almacenamiento, procesamiento, transferencia o recuperación de datos médicos, las cuales también serán verificadas antes de ser agregadas a la cadena de bloques.

Hao Wang et al. [38] proponen un sistema que introducen una nueva primitiva criptográfica que combina el cifrado basado en atributos, el cifrado basado en identidad y el firmado digital, para asegurar los registros de salud electrónicos (EHR, Electronic Healt Record) para ello, comparten la configuración y el algoritmo de generación de llaves los cuales se encuentran en un solo sistema y utiliza los mismos parámetros públicos y llaves secretas para las 3 diferentes funciones, también utilizan la cadena de bloques para garantizar la integridad y trazabilidad de los datos médicos. El sistema cuenta con cinco tipos de entidades que son: El centro de generación de llaves, hospitales, pacientes, nubes médicas y consumidores de datos los cuales representan diferentes roles dentro del sistema.

Hongyu Li et al. [18] presentan DPS (blockchain-based data preservation system), que es un sistema de preservación de datos médicos con almacenamiento confiable para garantizar la seguridad de los datos y la privacidad de los usuarios, utiliza la cadena de bloques de Ethereum. En el sistema DPS se cifran los datos médicos , si son archivos de texto estos se almacenan directamente en la cadena de bloques, en cambio si son otro tipo de archivos estos se segmentan en archivos de un megabyte se cifran y se envían a diferentes lugares , solo las rutas de los archivos son almacenadas en la cadena de bloques al mismo tiempo la identidad de los usuarios se mantiene anónima.

Benhamouda et al. [5] usan Hyperledger Fabric, la cual es una arquitectura de cadena de bloques permisionada (uso de contratos), para proporcionar un registro consistente distribuido compartido por un conjunto de pares (peers). Los autores extienden Hyperledger para dar soporte al intercambio de datos privados implementando un esquema de procesamiento llamado multiparte (MPC, por sus

siglas en inglés).

Yanez-Sierra et al. [42] presentan un esquema de Sobre Digital sobre una arquitectura de flujo de trabajo configurable, que permite asegurar la compartición de documentos en ambientes de nube privada, pero sin llevar un registro distribuido que conlleve un consenso entre las partes.

Tabla 2.1: Resumen del estado del arte

Autor	Titulo	Año	Aportación	Alcance	Diferencias
Huihui Yang et al. [43]	A blockchain-based approach to the secure sharing of healthcare data	2016	OPAL / Enigma, crea una red de pares que permite a las partes almacenar y analizar los datos con total privacidad y usa la cadena de bloques para registrar y controla el acceso a través de contratos inteligentes e identidades digitales.	Seguridad Trazabilidad	Los usuarios tiene que pagar, no se centra en el rendimiento, no menciona como se cifran los datos.
A. Azaria et al. [4]	Medrec: Using blockchain for medical data access and permission management	2016	MedRec que es un sistema de gestión de registros para manejar los registros médicos electrónicos utilizando la cadena de bloques (Ethereum).	Seguridad Trazabilidad	no menciona temas de rendimiento
Gaby G. Dagher et al. [11]	Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology	2018	Ancile un marco de trabajo basado en cadena de bloques para el acceso seguro a los registros médicos , utiliza contratos inteligentes en una cadena de bloques (Ethereum)	Seguridad Trazabilidad	Cifra los datos con cifradores simétricos por lo que la llave puede ser vulnerada. El gobierno tiene que pagar para mantener el sistema.
I. Sukhodolskiy et al. [35]	A blockchain-based access control system for cloud storage	2018	Prototipo de sistema multi-usuario para control de acceso a conjuntos de datos almacenados en un entorno de nube no confiable. El enfoque proporciona un control de acceso utilizando el cifrado basado en atributos. Utiliza la cadena de bloques (Ethereum) , como registro de los eventos de seguridad significativos.	Seguridad Trazabilidad	No trata temas de rendimiento, solo los eventos de seguridad son almacenados en la cadena de bloques
Harleen Kaur et al. [17]	A Proposed Solution and Future Direction for Blockchain-Based Heterogeneous Medicare Data in Cloud Environment	2018	Un modelo de datos de cuidado a la salud en una arquitectura de cadena de bloques en un ambiente de computo en la nube.	Almacenamiento seguro Trazabilidad	No mencionan como se realiza el aseguramiento de datos, y no tratan temas de rendimiento.
Hao Wang et al. [38]	Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain	2018	Introducen una nueva primitiva criptográfica que combina el cifrado basado en atributos, el cifrado basado en identidad y el firmado digital, esto le permite manejar diferentes requerimientos de seguridad con un solo sistema criptográfico y utilizan la tecnología de la cadena de bloques para garantizar la integridad y trazabilidad de los datos médicos.	Seguridad Trazabilidad	Cifra los datos con cifradores asimétricos (son lentos en comparación con los cifradores simétricos), no menciona temas de rendimiento
Hongyu Li et al. [18]	Blockchain-Based Data Preservation System for Medical Data	2018	DPS Un sistema de preservación de datos médicos con almacenamiento confiable para garantizar la seguridad de los datos y la privacidad de los usuarios, utiliza la cadena de bloques de Ethereum.	Seguridad Trazabilidad Almacenamiento.	Almacena los archivos de texto directamente en la cadena de bloques, por lo cual debe enviar la información a todos los nodos de la red, esto genera un impacto negativo en la red y en el almacenamiento de los nodos de la red.
Yanez-Sierra et al. [42]	A digital envelope scheme for document sharing in a private cloud storage	2015	un esquema de Sobre Digital sobre una arquitectura de flujo de trabajo configurable, que permite asegurar la compartición de documentos en ambientes de nube privada	Seguridad	No tiene un registro distribuido que lleve un consenso entre las partes

En la literatura expuesta anteriormente se puede observar que es factible usar la cadena de bloques para el manejo de archivos sensibles sin embargo las propuestas revisadas usan la cadena de bloques Ethereum por lo que deben realizar sus implementaciones adaptándose a las reglas definidas en Ethereum lo cual genera un sesgo en la configurabilidad de la solución, de la misma manera no se centran en buscar la eficiencia de las soluciones para el cifrado de datos, como también no manejan el ciclo de vida de los mismos.

3

Método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles

En esta sección se describen el método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles detallando cada una de las fases del mismo.

3.1 Vista general de las fases del método propuesto

El método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles consta de cuatro fases: La primera considera la definición de la gestión de cadenas de valor. La segunda define el modelo de aplicación de servicios de seguridad para cadenas de valor. En la tercera se define el manejo de transacciones en cadenas de valor. Finalmente, en la última fase

se define el mecanismo de trazabilidad de transacciones (consultas) de las cadenas de valor.

3.2 Fase 1: Definición de la gestión de la cadena de valor

La primera fase del método consiste en construir la cadena de valor identificando cada una de las etapas (eslabones) y uniones entre etapas de igual forma es necesario definir su funcionamiento.

3.2.1 Construcción de la cadena de valor

Para la construcción de las cadena de valor es necesario definir cada una de las etapas y y su conexiones.

3.2.1.1. Etapas

Las etapas son cada una de las entidades por donde tendrá que pasar el producto digital (archivos sensibles), y en las cuales irá obteniendo valor, estas etapas se pueden ver como cajas negras (Figura 3.1), y la actividad que se realice dentro de ellas (Edición, agregado, análisis etc.) no es importante para este trabajo de tesis, el punto importante es la entrada y la salida de cada etapa, bajo el supuesto de que el interior de cada caja negra (etapa) es seguro para el producto digital.

Las etapas pueden encontrarse en diferente lugar geográfico, por ejemplo la etapa 1 de la Figura 3.1 puede estar en México y la etapa 2 puede encontrarse en Canadá por este motivo es necesario garantizar la seguridad de los productos digitales (sensibles) entre las etapas y poder tener el historial de cada producto digital (trazabilidad).

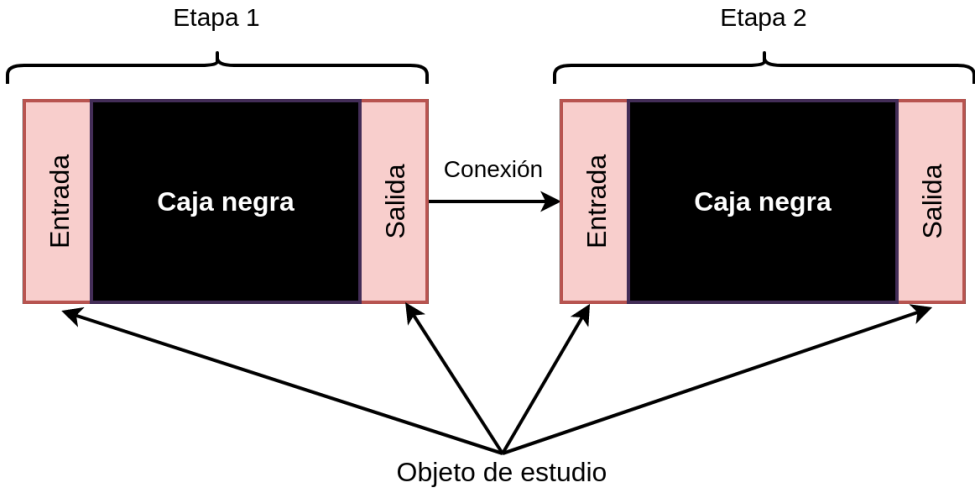

Figura 3.1: Etapas de la cadena de valor.

3.2.1.2. Conexiones

Las conexiones son las uniones entre las etapas, cada una de las conexiones representa el camino por donde deben viajar los productos digital para llegar de una etapa a otra, estas conexiones pueden ser de acceso público (internet) o ser administradas por terceros como por ejemplo subcontratando algún servicio de almacenamiento en la nube, esto expone a los productos digitales a múltiples amenazas a lo largo del camino, ya que no se puede garantizar la seguridad de las conexiones es obligación de cada entidad preparar a los productos para hacer frente a las amenazas que pueden encontrarse durante el viaje de una etapa a otra.

3.2.2 Modelo de procesamiento

El modelo de procesamiento de la solución es un modelo ETL (Extract, Transform and Load) Extracción, Transformación y carga, este modelo se aplica a cada etapa y al el mecanismo de seguridad que se encuentra a la entrada y salida de cada etapa de la cadena de valor (vea Figura 3.2).

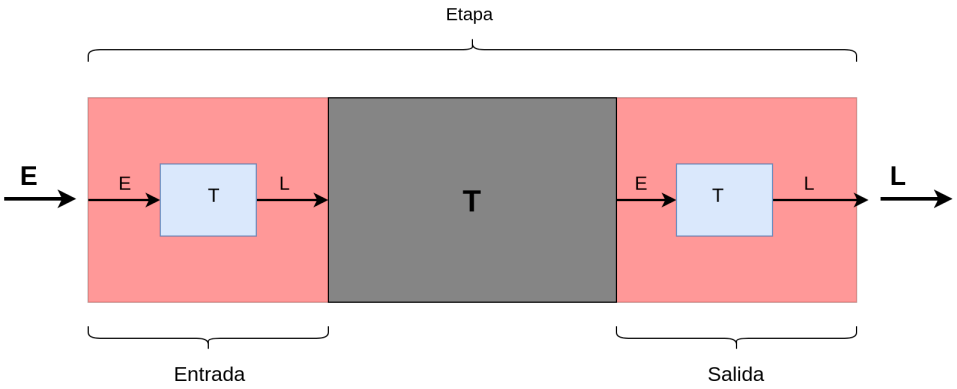

Figura 3.2: Etapas de la cadena de valor.

3.2.2.1. Procesamiento de la etapa

- Extracción (E_e): En esta parte se extraen los datos desde el origen, la cadena de valor inicia desde la creación del producto digital, por lo que en la primera etapa de la cadena la extracción E_{e1} seria la fuente de creación (tomógrafo, radiólogo, etc), en las etapas siguientes la extracción $E_{e2}, E_{e3}...E_{en}$ se hace desde el sistema de compartición de archivos al que fue enviado el producto digital, que puede estar en algún sistema de almacenamiento en la nube, el termino los productos son depositados en un medio de almacenamiento local (sistema de ficheros).

- Transformación(T_e): la transformación dentro de la etapa es lo que le agrega valor al producto (modificación, análisis, etc), este proceso queda fuera del alcance de este tema de tesis.

- Carga(L_e) : En esta fase los productos digitales que salieron de la fase de transformación T_e son enviados al destino elegido por ejemplo a un sistema de compartición de archivos fuera de la etapa.

3.2.2.2. Procesamiento del mecanismo de aseguramiento

El mecanismo de aseguramiento se encuentra en la entrada y salida de cada etapa, en la entrada es el responsable de recibir los datos de la fase de extracción de la etapa y pasarlos a la fase de

transformación de la misma,

- Extracción (E_s): en esta fase se extraen los productos del del sistema de ficheros para pasar a la fase de trasformación del mecanismo de seguridad T_s.
- Transformación (T_s): la transformación del mecanismo de aseguramiento s varia en la entrada in de la etapa e, $E_{s\_in\_e}$ (cifrado y firma) y en la salida out $E_{s\_out\_e}$ (descifrado y verificación de firma)
- Carga (L_s): esta fase es la salida de la etapa de transformación y consiste en colocar los elementos en algún directorio.

3.2.3 Creación de patrones de intercambio de datos

los patrones de intercambio de datos describen la forma que los productos pasan a través de diferentes etapas formando las cadenas de valor, para ello se utilizan patrones de publicación y suscripción para poder enviar y recibir los productos digitales, de esta forma cada entidad (etapa) puede publicar y suscribirse a alguna publicación, en la fase de publicación cada entidad genera contenido (productos digitales), este contenido es recibido por las entidad o entidades que estén suscritas a la publicación, la entidad que publica define los permisos para los datos, de esta forma puede elegir que suscripciones aceptar y solo esas podrás descargar los datos de esa publicación.

3.3 Fase 2: Aplicación de servicios de seguridad para cadenas de valor

En esta fase aplican los servicios de confidencialidad, control de acceso , integridad y no repudio a los productos digitales que viajan a lo largo de las cadenas de valor, esto se logra a través del proceso de generación del sobre digital, este proceso contiene diferentes mecanismos criptográficos, los cuales de listan a continuación indicando los servicios de seguridad que brindan.

- Cifrador simétrico (AES): Se usa para cifrar los productos digitales y brinda el servicio de confidencialidad.

- Cifrador asimétrico (CP-ABE) se usa para cifrar la llave de cifrado del cifrador simétrico y brinda los servicios de confidencialidad, control de acceso,

- Firma digital: Cada producto digital se firma para brindar los servicios de integridad y no repudio.

3.3.1 Aplicación de patrones paralelos en la construcción de sobres digitales

En esta sección se muestra el proceso de generación del sobre digital por medio de tres enfoques diferentes basados en patrones de paralelismo, En el primer enfoque se muestra el proceso de generación de sobre digital de forma secuencia , en el segundo se describe el proceso basado en paralelismo basado en tareas sobre un diseño manejador-trabajador, en el tercer enfoque se describe un enfoque de paralelismo sobre las operaciones de generación del sobre digital el cual llamamos solapado.

3.3.1.1. Primer enfoque: secuencial

En la Figura 3.3 se muestra el proceso de generación del sobre digital en forma secuencial, donde el flujo de datos pasa de una etapa a otra generando una tubería de procesamiento, la tubería recibe un archivo que pasa por la etapa AES, En AES se genera el archivo cifrado y la llave de cifrado que es enviada a la siguiente etapa para ser cifrada por CP-ABE, en la siguiente etapa se obtiene el hash del documento y se crea la firma, finalmente se obtiene un sobre digital al término de la tubería.

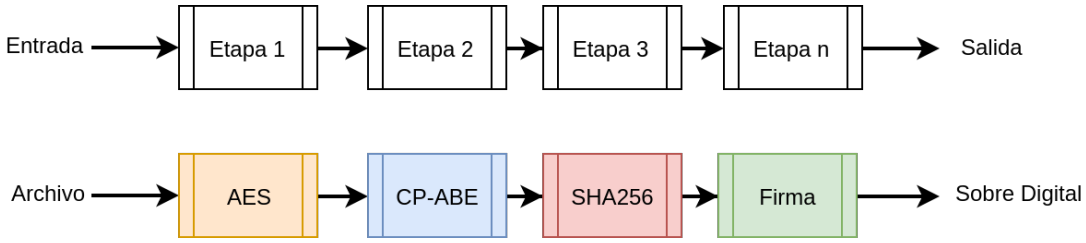

Figura 3.3: Tubería para la generación del sobre digital.

3.3.1.2. Segundo enfoque: manejador-trabajador

Es un enfoque donde el conjunto de tareas está dada por el número de archivos de la fuente, las tareas se procesan con un patrón manejador-trabajador donde el manejador es el encargado de crear un numero de trabajadores y de asignarles las tareas, los trabajadores son responsables de realizar las tareas asignadas por el manejador, las cuales consisten en la generación del sobre digital del archivo seleccionado. todos los trabajadores realizan el mismo procesamiento solo cambian los parámetros de entrada.. cada trabajador es independiente de los otros e internamente los trabajadores realizan el procesamiento secuencial definido en AES4SeC [23], el diseño se observa en la Figura 3.4

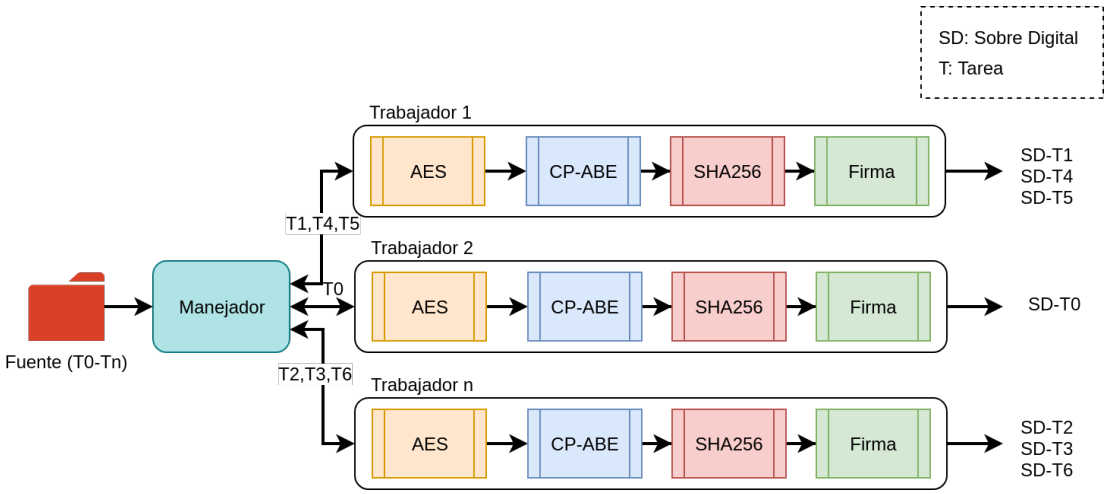

Figura 3.4: Diseño: manejador-trabajador con paralelismo basado en tareas.

3.3.1.3. Tercer enfoque: Solapado

En la tubería de generación del sobre digital existen etapas que no tiene dependencia secuencial y pueden ser ejecutadas de manera simultanea, El algoritmo de AES recibe un archivo de entrada e internamente genera la llave de cifrado y cifra el contenido, esa llave después es enviada a CP-ABE para ser cifrada, si se aísla el mecanismo de generación de llave de AES, una vez generada la llave AES puede continuar cifrando el contenido y al mismo tiempo CP-ABE puede cifrar la respectiva llave de AES. El algoritmo de hash (SHA256) necesita el archivo de entrada por lo que no tiene ninguna relación con AES y CP-ABE.

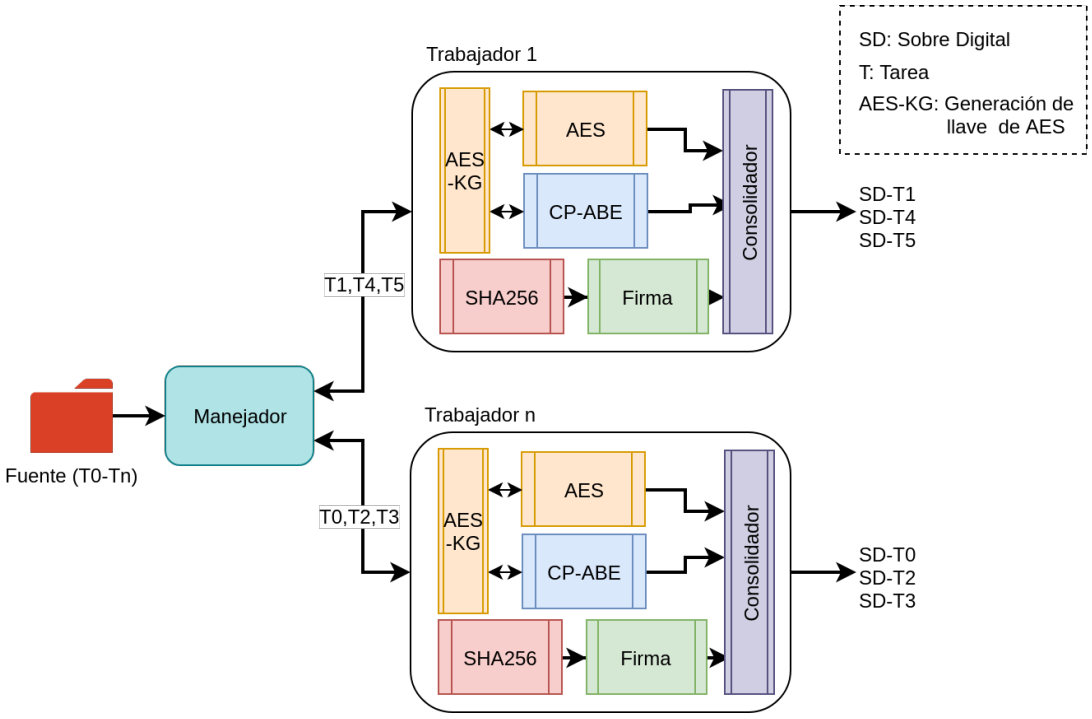

Figura 3.5: Registro de transacciones y creación de la cadena de valor.

3.4 Fase 3: Manejo de transacciones en cadenas de valor

En esta fase cada movimiento que se genera en los productos a través de la cadena de valor (transacción) se registra en la cadena de bloques, para ello es necesario definir cada uno de los

componentes necesarios para el manejo de las transacciones.

3.4.1 Definición del modelo de negocios

En esta etapa se define el modelo de negocios, esto involucra el que (activos) y quien (entidades),

3.4.1.1. Definición de activos

Definimos a los activos como: todos los productos digitales sensibles (tomografías, imágenes satélites) que transitan a lo largo de la cadena de valor.

3.4.1.2. Definición de entidades

Las entidades corresponden a cada una de las etapas en la cadena de valor y tienes los permisos para poder manipular los activos (productos digitales sensibles), siempre y cuando estén dentro de la etapa.

3.4.2 Definición del modelo de la red de la cadena de bloques

En este paso se definen los elementos que conforman la red de la cadena de bloques, estos son:

- Libro mayor en ingles Ledger en el se registra de manera inmutable todas las transacciones generadas por contratos inteligentes (chaincode), uno por canal, contiene la cadena de bloques y el estado actual del sistema (World state).
- Contratos inteligentes llamado en Hyperledger como chaincode, que son simplemente un fragmento de código que accede al libro mayor, escrito en uno de los lenguajes de programación compatibles.
- Nodos pares (peers) en ellos se alojan libros mayores y contratos inteligentes por este motivo son elementos fundamentales para la red.

- Nodos de pedido (orderer), son el mecanismo por el cual las aplicaciones y los pares interactúan entre sí para garantizar que el libro mayor de cada par se mantenga constante, estos nodos se encarga de agrupar las transacciones en bloques y distribuirlas a toda la red de pares, donde se pueden verificar antes de aplicarlas a la copia local del libro mayor de cada par.

- Canal: Es un mecanismo por el cual un conjunto de componentes dentro de la red (pares, aplicaciones) puede comunicarse y realizar transacciones de forma privada.Al unirse a un canal estos componentes, acuerdan colaborar para compartir y administrar colectivamente copias idénticas del libro mayor asociado con ese canal.

- Autoridades certificadoras: Emiten certificados digitales validados criptográficamente que cumplen con el estándar X.509, de esta forma un actor o un nodo puede participar en la red de la cadena de bloques, a través de esta identidad digital, se despliega una autoridad certificadora por organización.

- Aplicaciones: Estas interactúan con sus pares (peers) para acceder al libro mayor y a los contratos inteligentes, en este trabajo la aplicaciones son las encargadas de asegurar los activos por medio de la la generación de los sobres digitales.

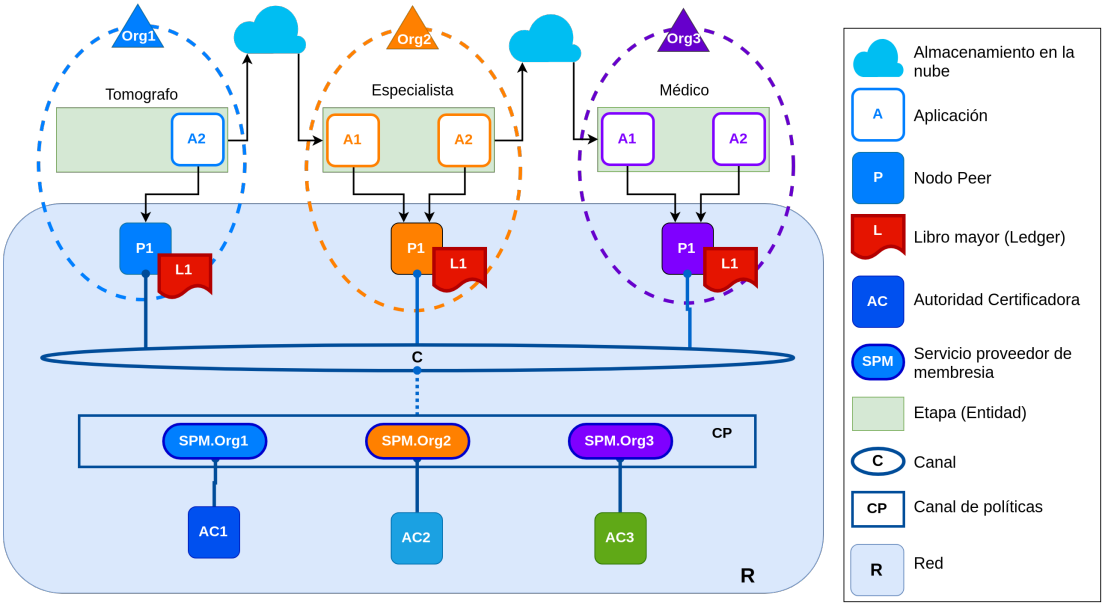

Figura 3.6: Modelo de la red de la cadena de bloques.

3.4.3 Transacción de activos en patrones de compartición

Cada Activo que viaja a través de la cadena de valor, es enviado de una etapa a otra por medio de un patrón de publicación/suscripción, cada movimiento del activo genera una transacción y esta es registrada en la cadena de bloques.

El proceso de registro de las transacción desde una aplicación hasta el libro mayor (ledger) se describe a continuación:

Los pares (peers), junto con los nodos de pedido (Orderer), se aseguran de que el libro mayor se mantenga actualizado en cada par. En el ejemplo de la Figura 3.7, la aplicación A se conecta a P1 e invoca el Contrato S1 para actualizar el libro mayor L1. P1 invoca a S1 para generar la actualización de libro mayor propuesta. La aplicación A recibe la respuesta de la propuesta y crea una transacción a partir de todas las respuestas, que la envía a O1 para ordenar. O1 recopila transacciones de toda la red en bloques y las distribuye a todos los pares, incluido P1. P1 valida la transacción antes de aplicar a L1. Una vez que L1 se actualiza, P1 genera un evento, recibido por A, para indicar la finalización.

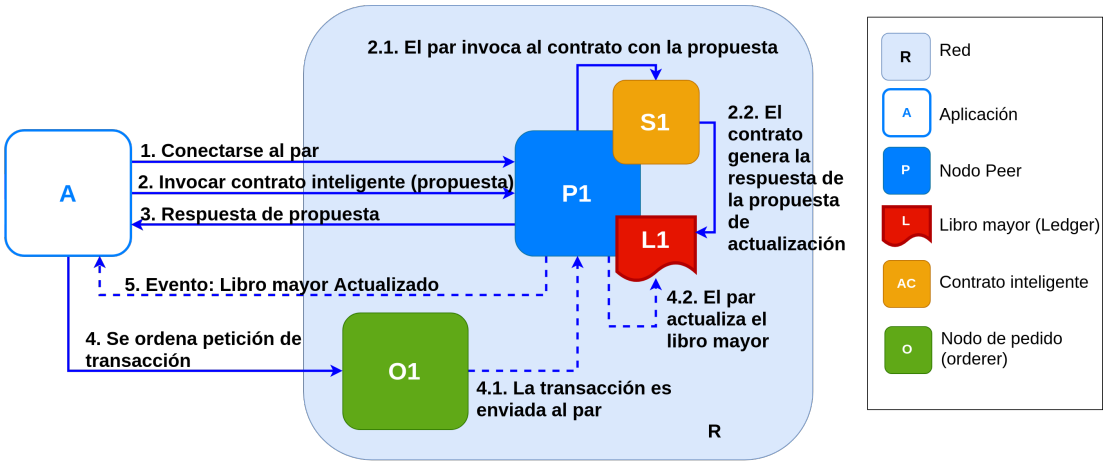

Figura 3.7: Registro de la transacción en el libro mayor.

3.5 Fase 4: Trazabilidad de transacciones (consultas)

En esta fase se describe como se realizan las consultas a la red de la cadena de bloques

En la Figura 3.8 se describe como se realizan las consultas en la cadena de bloques, la aplicación A se conecta a P1 e invoca el contrato S1 para consultar el libro mayor L1. P1 invoca a S1 para generar una respuesta de la propuesta que contenga un resultado de consulta. La aplicación A recibe la respuesta de la propuesta y con esto el proceso está completo.

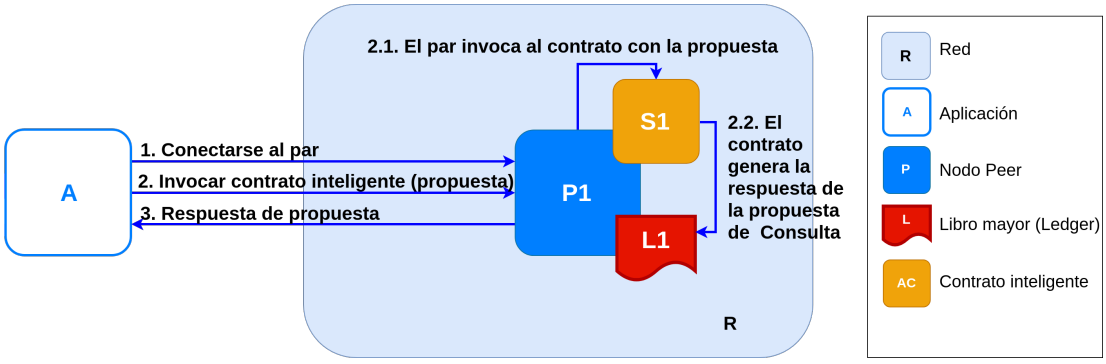

Figura 3.8: Consulta al libro mayor.

Para el registro de transacciones se utiliza el marco de trabajo (framework) hyperledger, este se

usa para configura la red y diseñar el modelo de las transacciones entre las entidades.

En la Figura 3.9 se muestra 2 etapas por las que pasara un producto digital, las etapas pueden estar en diferente lugar geográfico por lo que el producto digital viajaría por un medio inseguro, al final de cada etapa se utiliza el proceso de generación del sobre digital, y es necesario el registro de cada transacción en la cadena de bloques, para ello la transacción se define como un hash de entrada a la etapa $H - in$ y uno de salida $H - out$, el $H - in$ esta formado por hash del producto digital $Hashin1$ y el hash de la firma $Hashin2$, de esta forma el producto y el usuario quedan registrados, y el has de salida $H - out$ se obtiene del sobre digital creado al cifrar el producto digital Tn , de esta forma la transacción se forma al concatenar el hash de entrada y salida quedando : $transET1 = H - in, H - out$, la transacción de la siguiente etapa recibirá como hash de entrada $H - in$ el hash de la transacción en la etapa anterior anterior quedando $H - in = TransEt1, H - out = Hashin1, Hashin2$. La cadena de valor se forma con las transacciones de las etapas por donde pasa el producto digital $cadenadevalor = transET1, TransET2...TransETn$, de esta forma la transacción $TransETn$ estará ligadas a las transacciones anteriores.

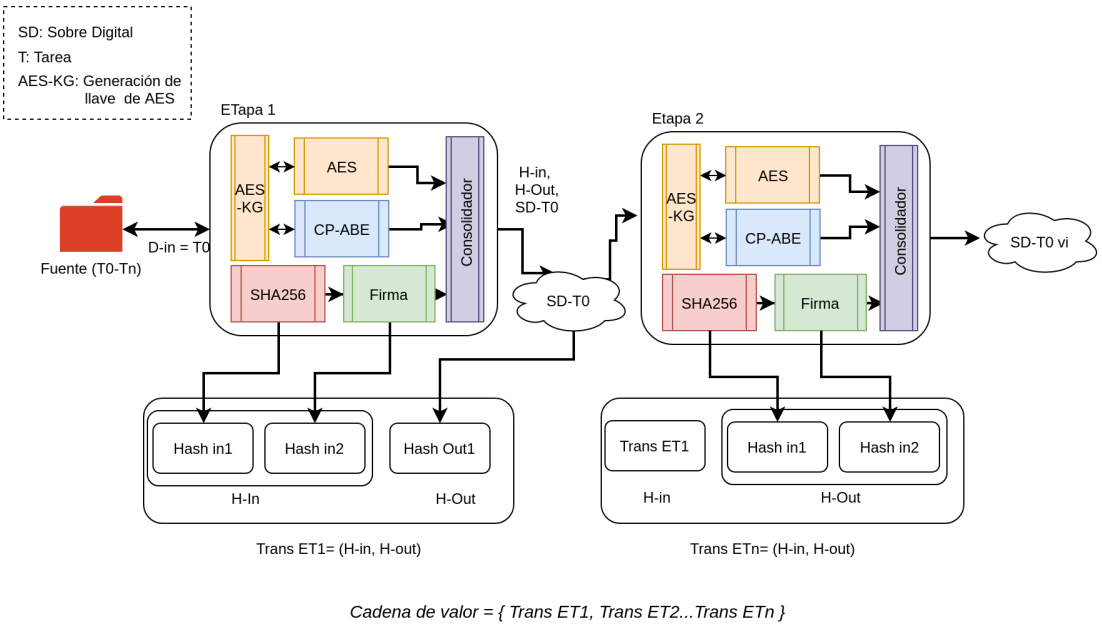

Figura 3.9: Etapas del proceso de generación de sobre digital en paralelo.

3.6 Aplicación del método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles

En la Figura 3.10 se muestra el diagrama de la solución propuesta, en el se muestra una cadena de valor en el área médica en la cual los productos digitales son tomografías que pasan por diferentes etapas , (tomógrafo, médico, paciente) intermedio a cada etapa las tomografías son almacenadas en la nube, antes de ser enviadas y salir de la etapa pasan por proceso de aseguramiento que involucra la generación del sobre digital (cifrado y firmado) de cada tomografía, este proceso es realizado por una aplicación encapsulada en contenedores los cuales son desplegados en patrones de paralelismo para mejorar la eficiencia, al entrar a cada etapa se realiza el proceso inverso que involucra descifrar el producto digital y verificar la firma digital, cada ves que un sobre digital es enviado o recibido se genera una transacción y esta es registrada en la cadena de bloques.

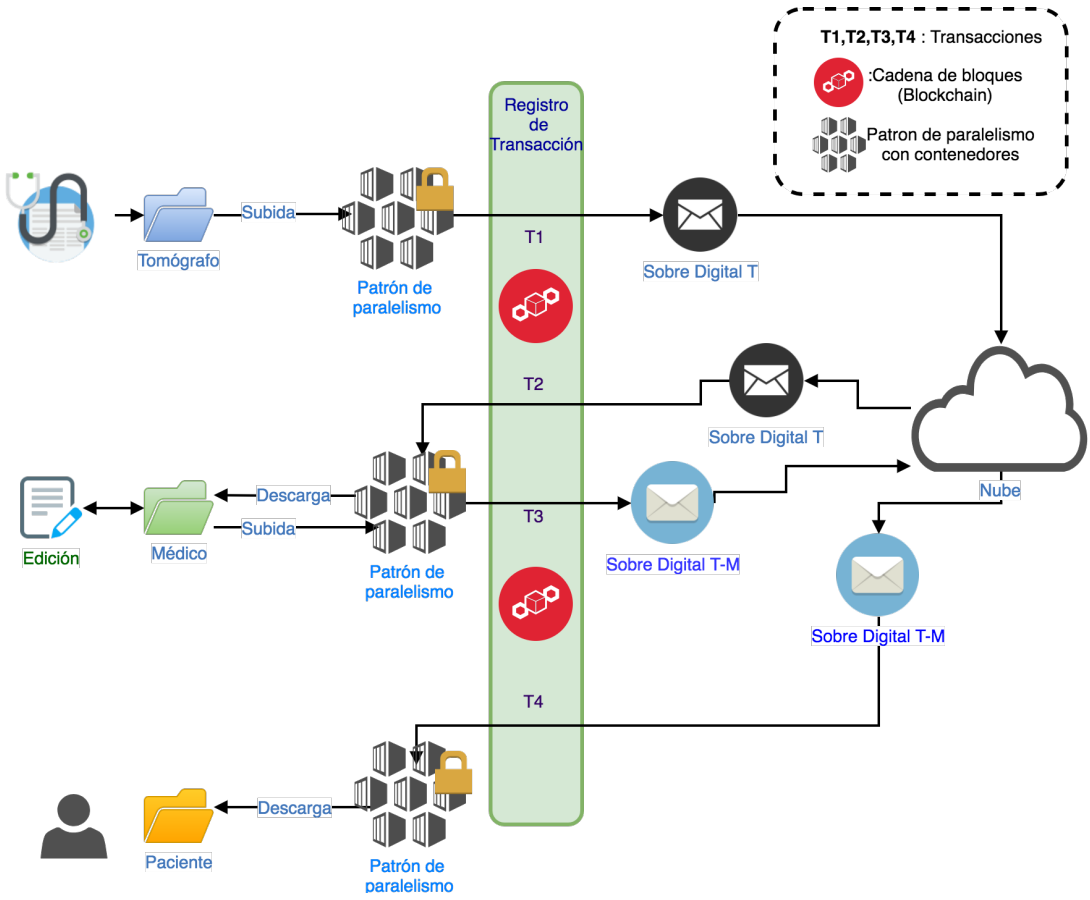

Figura 3.10: Despliegue del método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles.

4

Evaluación experimental y Resultados

4.1 Metodología de evaluación

La metodología para realizar la evaluación experimental esta dividida en tres etapas, en la primera se realiza una evaluación controlada de los enfoques propuestos del proceso de generación del sobre digital, con el objetivo de medir la eficiencia de los patrones de paralelismo y validar su eficacia. En la segunda etapa se realiza un par de estudios de caso para validar la factibilidad de la solución en un problema del mundo real basado en el manejo de imágenes satelitales y médicas. En la tercera etapa se realiza una comparación de la eficiencia del proceso de generación del sobre digital con un software existente en el mercado llamado Jenkins, que comúnmente se usa en la literatura para construir tuberías de software como las que se proponen este documento de tesis.

En la Figura 4.1 se muestra el diseño experimental de las primeras dos etapas, en la primera etapa se realiza la evaluación controlada de los patrones de paralelismo. La evaluación consideró cada una de las configuraciones soportadas por la aplicación para los diferentes enfoques, y están divididas en

tres bloques principales:

- Parámetros de sintonización del sistema de Criptografía.
  - Nivel de seguridad: 128, 192, 256.
  - Políticas de cifrado: AND (nofn), OR (1ofn).
- Parámetros de los Patrones de paralelismo(Numero de trabajadores): 2,4,6,8,10.
- Tipos de tamaño de datos(tamaño de archivos): 1MB, 10MB, 100MB, 1GB.

Para la evaluación se realizó la combinación de cada parámetro previamente menciondo. Por ejemplo (128, AND, 2, 1MB), (192, AND, 2, 1MB) y cada combinación de parámetros se ejecutó treinta y un veces obteniendo la media de cada combinación de parámetros.

4.1.1 Configuraciones estudiadas

- *Solapado*: Esta configuración representa el esquema en el cual los sistemas criptográficos que no presentan dependencia funcional entre ellos son ejecutados en forma concurrente.
- *Maestro-trabajador* $M - T$: Esta configuración representa el esquema donde los sistema criptográficos son ejecutas en la forma de una tubería simple donde los sistemas son alineados secuencialmente para agregar gradualmente las propiedades de seguridad al sobre digital. cada tubería es asignada a un trabajador diferente por el manejador, el cual lanza tantos trabajadores como cores disponibles en equipo de cómputo donde se ejecute la tubería.

La evaluación se condujo ejecutando las configuraciones para cada combinación de parámetros bajo las mismas condiciones y se realizó una comparación directa. Tabla 4.1 describe las características de la infraestructura usada para la evaluación descrita.

En la segunda y tercera etapas de la evaluaciónse conducen estudios de caso, donde solo se realizó la variación del número de trabajadores y para las configuraciones de criptografía (tamaño de

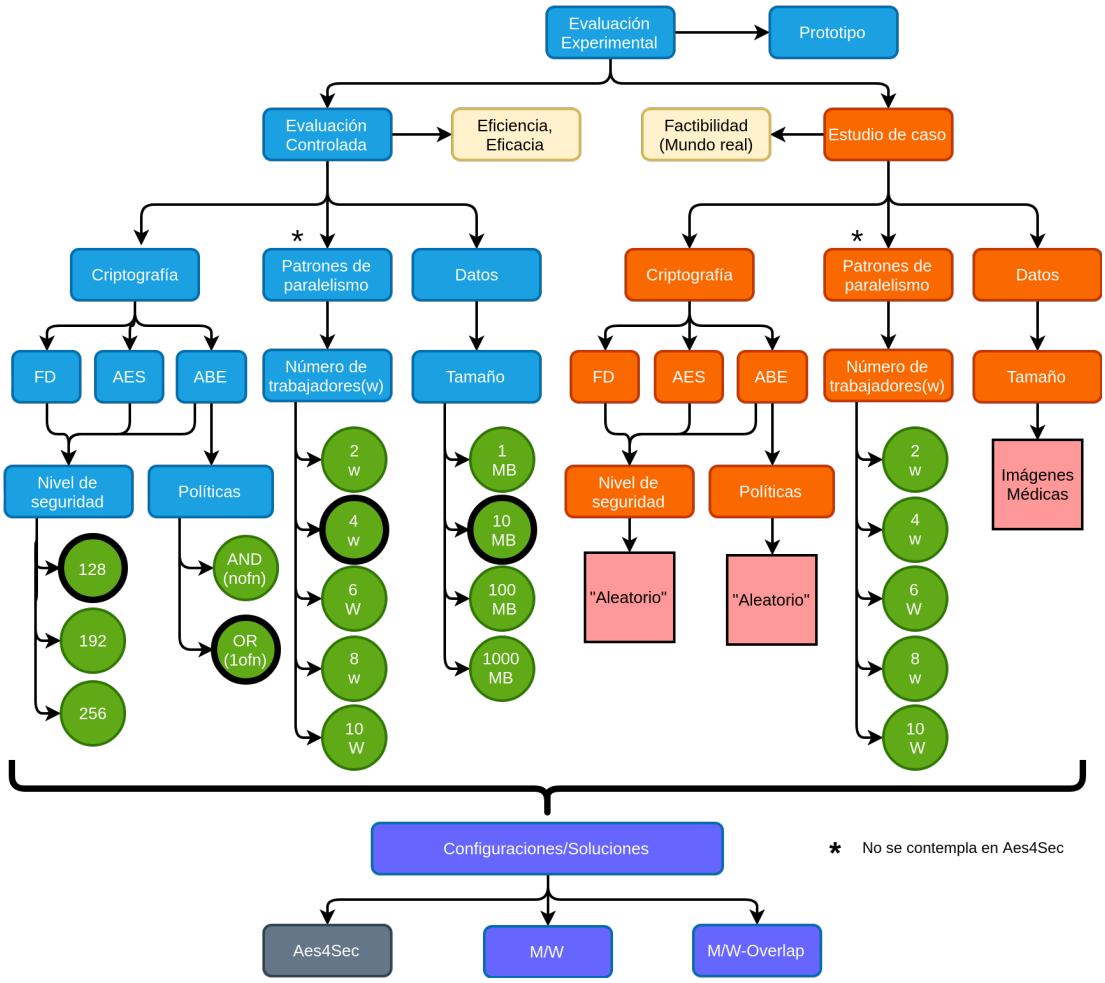

Figura 4.1: Diseño experimental.

llave) y datos (tamaño de archivos) se realizó la selección de forma pseudoaleatoria, las cuales serán descritas en su correspondiente sección.

4.2 Evaluación Controlada: Análisis del rendimiento de los patrones de paralelismo

Después de analizar la naturaleza de las soluciones diseñadas se espera que los patrones solapados produzcan un mejor rendimiento que los patrones $M-T$ cuando se aseguran archivos de gran tamaño

Tabla 4.1: Infraestructura usada para la evaluación

Tipo	OS	CPU	Cores	RAM	HD
Máquina física	CentOS 7	Intel Xeon CPU E5-2640 2.50GHz	24	64	3.5.T
Máquina física	CentOS 7	Intel Xeon CPU E5645 2.40GHz	6	12	2T
Máquina física	CentOS 7	Intel Xeon CPU E5645 2.40GHz	6	12	1.5T
Máquina física	CentOS 7	Intel Xeon CPU E5645 2.40GHz	6	12	1.5
Máquina física	CentOS 7	Intel Xeon CPU E5675 3.07GHz	6	24	1T
Máquina física	CentOS 7	Intel Xeon CPU E5-2650 2.60GHz	16	64	3T
Máquina física	CentOS 7	Intel Xeon CPU E5-26502.20GHz	24	251	3T

(centenas de MBs) usando llaves de largas. También se espera que los patrones $M - T$ ofrezcan un mejor rendimiento que los patrones solapados cuando se manejan lotes completos de archivos de tamaño pequeño (decenas de MBs).

El objetivo por tanto de esta evaluación es determinar si la intuición previamente descrita se puede corroborar con los resultados.

La Figura 4.2 muestra el tiempo de servicio (eje vertical) para las configuraciones estudiadas usando diferentes tamaños de llaves y variando el número de trabajadores usados por cada configuración (eje vertical). Se puede observar el comportamiento esperado, a mayor número de trabajadores, el tiempo de servicio es menor, con cada aumento en el numero de trabajadores el tiempo de servicio muestra una disminución llegando a estabilizarse en 8 trabajadores para niveles de 256 y en 5 para niveles de 192 y 128, la gráfica muestra una disminución del tiempo de hasta un 80 % en comparción con la versión secuencial (con 1 solo trabajador).

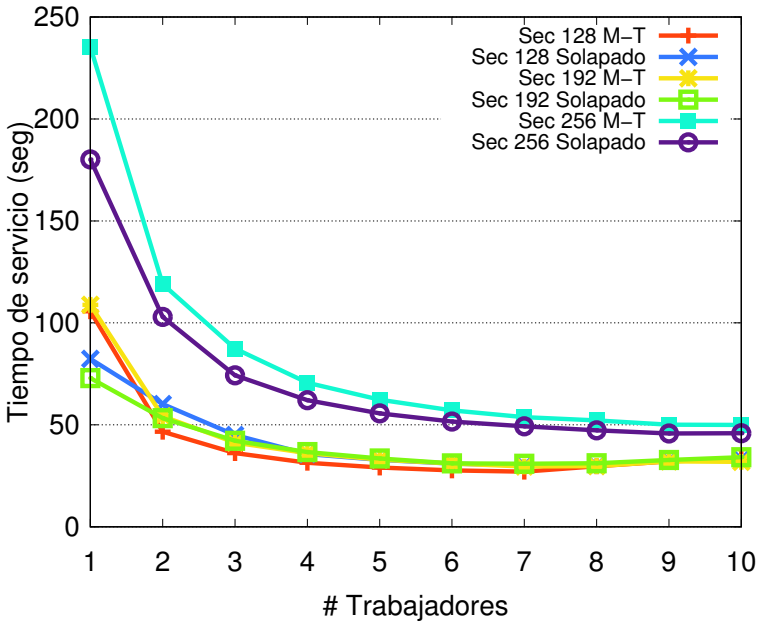

Figura 4.2: Tiempo de servicio del proceso de generación del sobre digital, enfoque 2 y 3.

En la Figura 4.3 se observa, en el eje vertical, el rendimiento (MB/segundos) para nivel de seguridad de 128 variando el número de trabajadores. Se la configuración manejador-trabajador alcanza un mayor rendimiento que la configuración desplegando el esquema solapado. Se observa, como esperado, que al agregar trabajadores aumenta el rendimiento hasta llegar a 7 trabajadores, al pasar los 7 trabajadores el rendimiento disminuye, esto es porque se ha llegado a utilizar la máxima capacidad de los recursos físicos, seguir aumentando trabajadores generara colas en los procesos disminuyendo el rendimiento de las soluciones estudiadas.

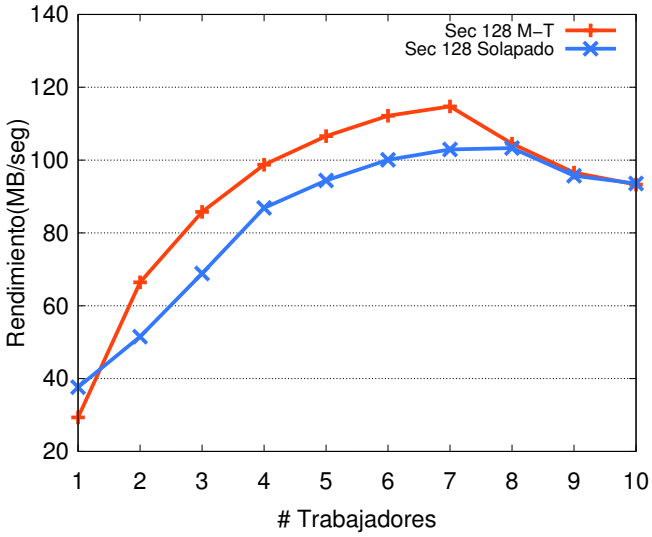

Figura 4.3: Comparación de Rendimiento nivel de seguridad 128 enfoque 2 y 3.

La Figura 4.4 muestra a las configuraciones usando un tamaño de llave de 192k. Se observa que el comportamiento de ambas configuraciones es similar al observado en Figura 4.3 (el punto más alto de mejora se observa con 7 trabajadores.

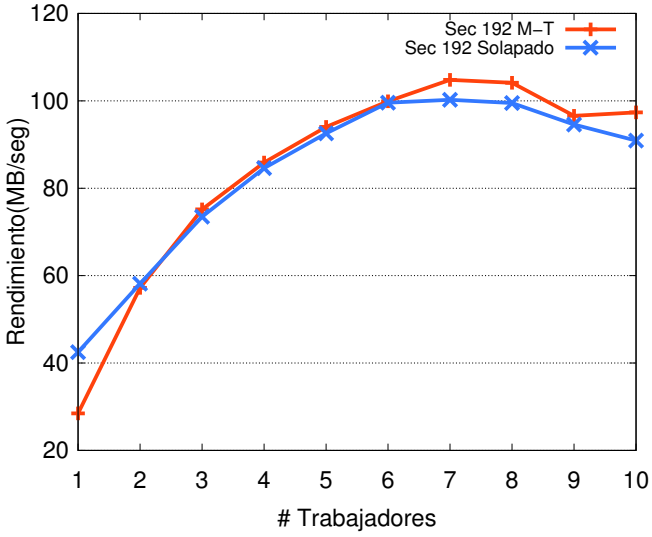

Figura 4.4: Comparación de Rendimiento nivel de seguridad 192 enfoque 2 y 3.

Figura 4.5 muestra el rendimiento de las configuraciones usando un tamaño de llave de 256. Se

observa que el rendimiento para este nivel de seguridad del esquema de patrones solapados produce un mejor rendimiento que la configuración manejador trabajador y este mejoramiento es sostenido hasta llegar a nueve trabajadores.

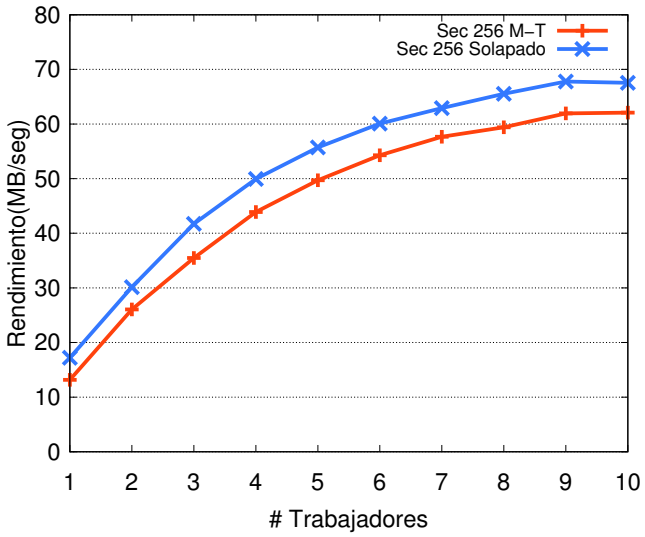

Figura 4.5: Comparación de Rendimiento nivel de seguridad 256 enfoque 2 y 3.

En las figuras 4.3, 4.4 y 4.5 se observa que a más trabajo (256) el esquema de patrones solapados produce el mejor rendimiento, mientras que para el nivel de seguridad definido por llaves de 128 la configuración manejador-trabajador produce el mejor rendimiento. Esto comprueba que la intuición descrita al inicio de la evaluación es soportada por los resultados de la evaluación experimental.

Se demuestra que ambos esquemas de procesamiento basado en patrones es eficaz para mejorar significativamente el rendimiento de esquemas disponibles en la literatura [23] ya que observaron rangos de mejora en los tiempos de respuesta entre 40 % y 85 % en comparación con la versión secuencial disponible en la literatura.

4.3 Estudio de caso basado en Imágenes Satelitales de la agencia espacial Europea (ESA)

En esta etapa se realiza la comparación del proceso de generación del sobre digital con patrones de paralelismo descrito en este documento de tesis con otro sistema disponible en el estado del arte llamado Jenkins [34]. Este generador de flujos de continuos de trabajo es una herramienta popular entre las organizaciones para construir tuberías de procesamiento de software.

Jenkins utiliza un conjunto de complementos de código abierto que admite la implementación e integración de tuberías de entrega continua. Esta herramienta se enfoca principalmente al desarrollo y pruebas de software y permite automatizar el proceso para llevar el software desde el control de versiones hasta sus usuarios y clientes.

Jenkins proporciona un conjunto extensible de herramientas para modelar tuberías de entrega simples a complejas "como código" a través de la sintaxis de lenguaje específico de dominio, el cual a su vez que permite desplegar las soluciones en contenedores virtuales. Jenkins además esta basado en un modelo de paralelismo implícito, el cual también es usado por los patrones de paralelismo propuestos en esta Tesis.

Lo anterior permite realizar una comparativa justa y directa entre Jenkins y los patrones de paralelismo propuesto en esta Tesis ya que la comparación se realizó bajo las mismas condiciones, hardware, mecanismos de aseguramiento y datos.

4.3.1 Datos

A lo largo de este documento se ha tomado de base el ámbito médico. En el ámbito de la gestión de datos estratégicos del territorio nacional, las imágenes satelitales son de uso restringido y son consideradas, en algunos casos militares, como contenidos de alta sensibilidad. Esto equipara este dominio satelital al dominio descrito en este documento de Tesis.

El repositorio de imágenes satelitales incluye un conjunto de 1620 archivos con un peso total de 59 GB y comprende tanto datos, metadatos como imágenes de la superficie de la tierra capturadas por la misión SMOS de la Agencia Espacial Europea (ESA).

4.3.2 Configuraciones y métricas

La evaluación se realizó para cada nivel de seguridad $\lambda \in \{128, 192, 256\}$ variando el numero de trabajadores $W \in \{1, 2, 4, 6, 8, 10\}$ y se midió el tiempo de servicio para la generación del sobre digital por lote de archivos para las siguientes soluciones:

- *AES4SeC*: Esta solución representa una implementación secuencial del constructor de sobres digitales disponible en el estado del arte.
- *Manejador-trabajador*: Esta configuración ha sido previamente descrita en este documento de Tesis.
- *Jenkins*: Esta configuración representa los componentes de $AES4SeC$ organizados en tuberías paralelas. Es importante notar que esto fue posible ya que Jenkis es nativo de Java y comúnmente usado para crear tuberías con código de este lenguaje de programación, lo cual es el caso de $AES4SeC$.

La Figura 4.6 muestra, en el eje vertical, el tiempo de servicio producido por las soluciones estudiadas al procesar el lote completo de de imágenes de satélite variando el número de trabajadores (eje horizontal) para los 3 niveles de seguridad (128,192,256).

Como se puede observar, al aumentar el número de trabajadores, el tiempo de servicio disminuye hasta llegar a un punto de estabilidad, para el nivel de 128 se observa el tiempo de servicio disminuye hasta 4 trabajadores, para 192 hasta 6 trabajadores y para el nivel 256 hasta 10 trabajadores. Dichos números de trabajadores representan la mejor opción para cada nivel de seguridad.

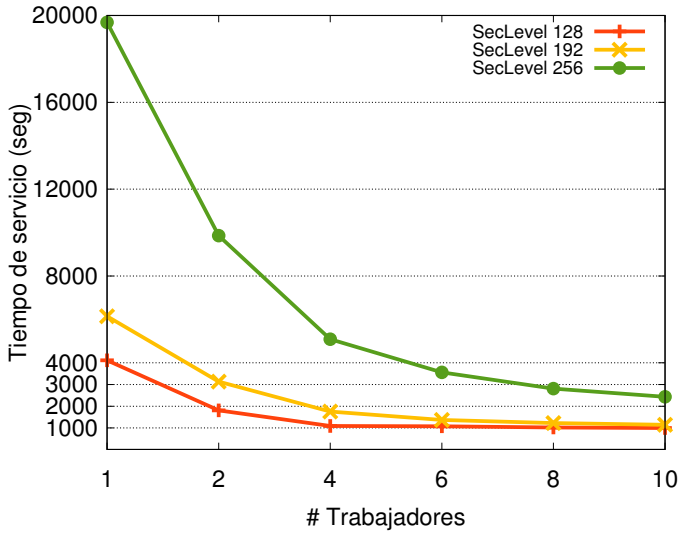

Figura 4.6: Tiempo de servicio en la generación del sobre digital.

La Figura 4.7 muestra el porcentaje de ganancia de la configuración manejador-trabajador sobre la versión secuencial AES4SeC para los diferentes niveles de seguridad variando el numero de trabajadores. Se observa una ganancia que gradualmente crece con el número de trabajadores hasta ser superior al %80.

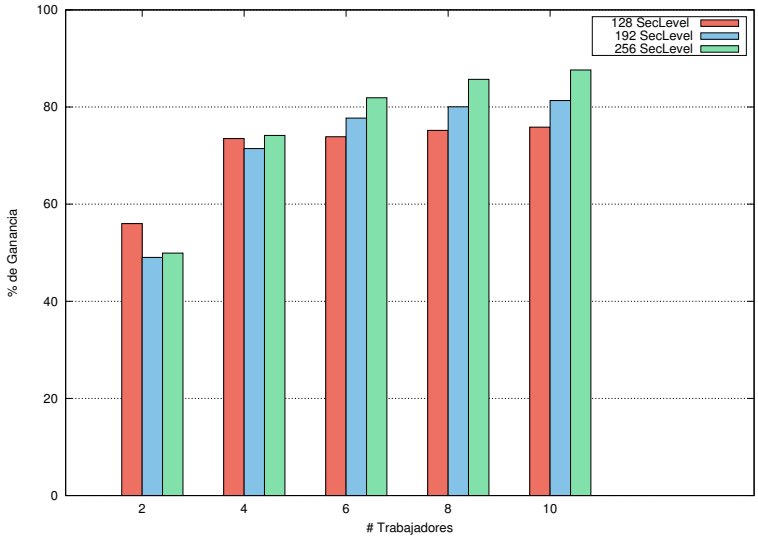

Figura 4.7: Porcentaje de ganancia de Manejador-trabajador sobre AES4SeC en la generación del sobre digital.

La Figura 4.8 muestra el tiempo de servicio del proceso de generación del sobre digital por lote de archivos, variando el nivel de seguridad y el número de trabajadores por cada prueba. Come se puede observar, la disminución del tiempo de servicio al aumentar los trabajadores y al disminuir el nivel de seguridad es evidente. De igual forma, en todos los casos el tiempo de servicio producido por la configuración manejador-trabajador es menor al obtenido por la tubería creada con Jenkins.

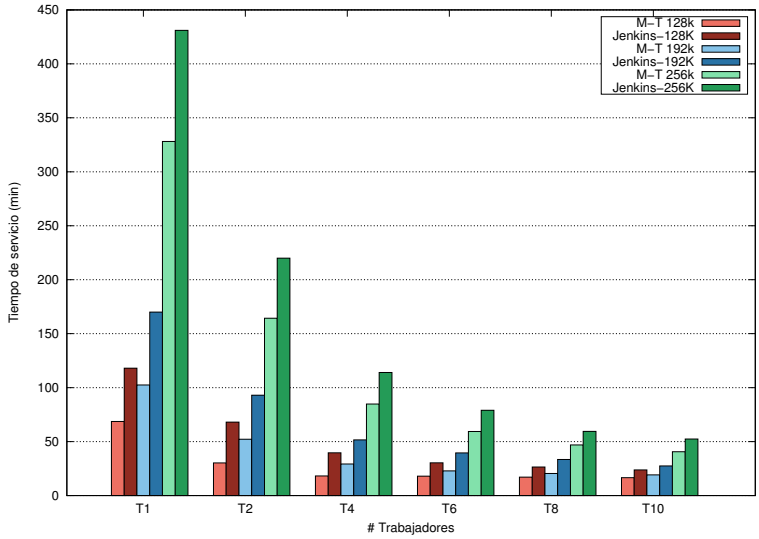

Figura 4.8: Tiempo de servicio manejador-trabajador vs Jenkins.

La Figura 4.9 muestra el porcentaje de ganancia de la configuración manejador-trabajador sobre la tubería creada con Jenkins. Se puede observar que para el nivel de 128 se obtuvo la mayor ganancia y para el 256 la ganancia fue menor.

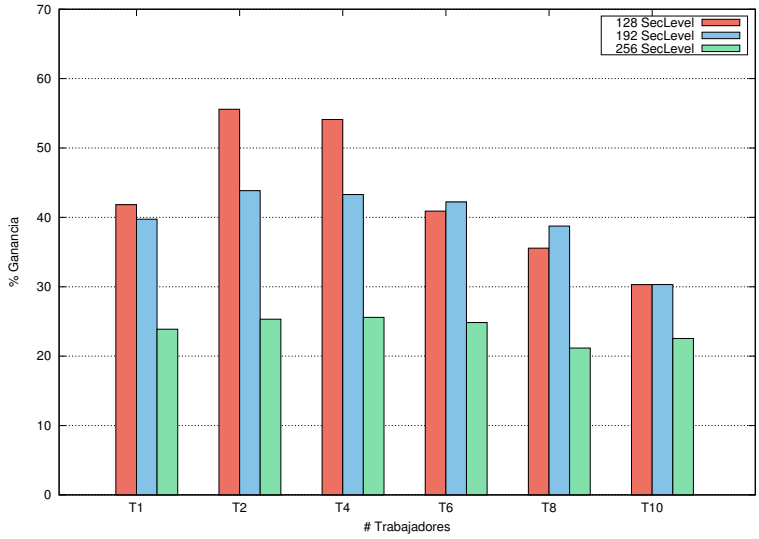

Figura 4.9: Porcentaje de ganancia manejador trabajador vs Jenkins

Se observó que Jenkins es eficiente al manejar carga de trabajo pesada (cifrado de archivos con nivel de seguridad 256). Sin embargo, la tubería creada mediante el método descrito en este documento de tesis fue también significativo ya que se observó un mejoramiento en el rango de (56 %-25 %) como mostrado en Figura 4.9.

Con lo anterior se muestra que el método no solo es eficaz para mejorar el rendimiento del proceso de aseguramiento de contenidos sensibles sino que también es bastante competitivo comparado con las opciones que existen actualmente para crear cadenas de procesamiento eficientes.

4.4 Estudio de caso basado en registro de transacciones de imágenes médicas

El estudio de caso sirvió para determinar la factibilidad del método en escenarios del mundo real. Este estudio ofrece elementos de incertidumbre (pseudoaleatorios) entre los que se contempla la definición mediante una distribución uniforme del nivel de seguridad, las políticas de cifrado y el tamaño de los archivos.

Para esta prueba se utilizaron 1864 imágenes médicas (mamografías) con un peso total de 50 GB con una media de 26 MB y desviación estándar 8.7 MB.

Para la prueba se definió un productor X (rayos x)(véase Figura 4.10) encargado de publicar las mamografías, y 7 consumidores A,B,C,D,E y F los cuales descargarían las imágenes.

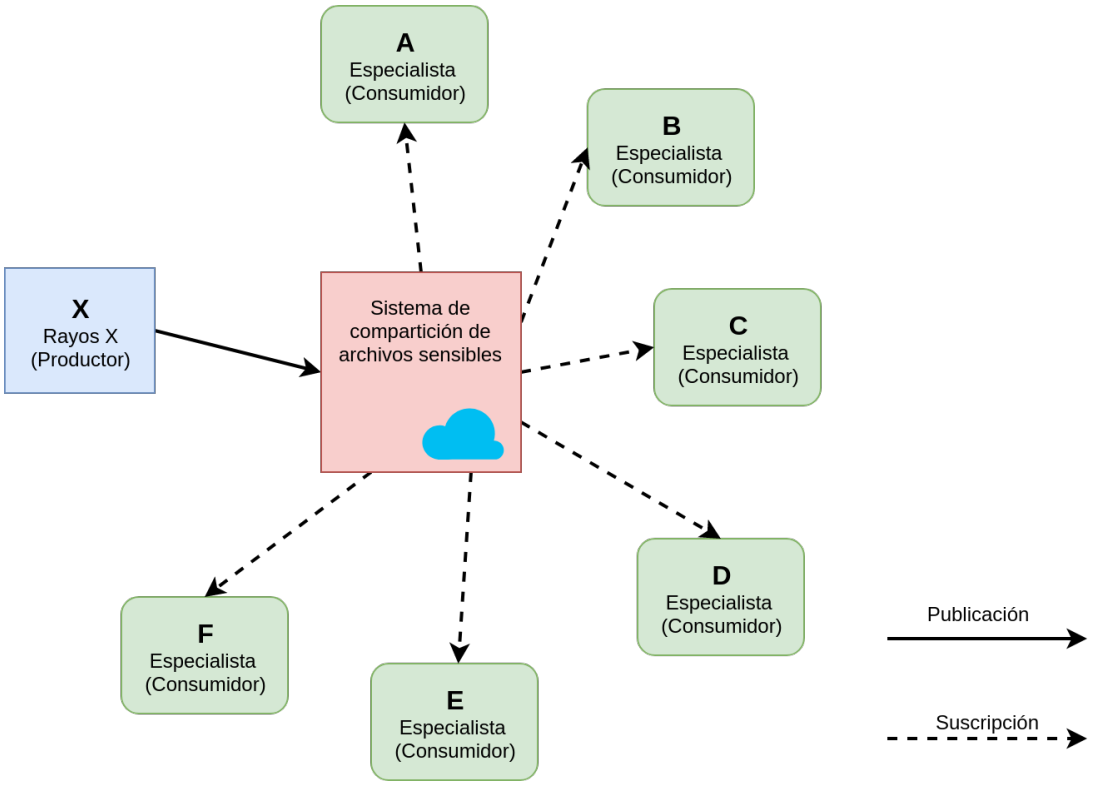

Figura 4.10: Diagrama Estudio de caso.

Para generar las políticas de seguridad se crearon diferentes grupos, de este forma un usuario consumidor puede pertenecer a varios grupos, en la Figura 4.12 se muestra el ordenamiento de los consumidores y grupos. Se tienen 9 grupos y a cada uno se asigno diferente configuración de seguridad, (tabla 4.2) variando el nivel de seguridad y las operaciones de cifrado y firma, pudiendo tener ambas o una de las 2, esto se realizó para poder probar con todas las configuraciones que soporta el mecanismo de seguridad.

Cada política de seguridad define a cada grupo, y los atributos son las letras A,B,C,D,E y F, por ejemplo el grupo 2 tiene la política $A, C, D1of3$ esto significa que solo los consumidores A, C o D podrán acceder a las imágenes médicas,

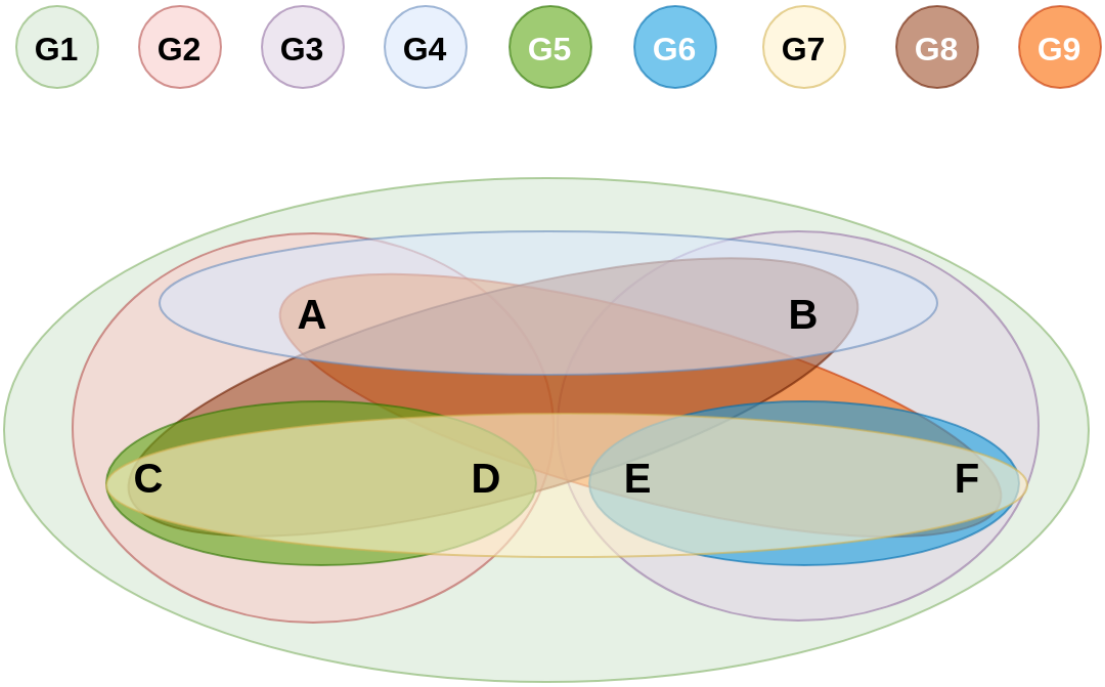

Figura 4.11: Grupos de cifrado.

Tabla 4.2: Configuraciones del nivel de seguridad, políticas de cifrado y operaciones de cifrado y firma.

Grupo	Política	Nivel Sec	Cifrado	Firma
1	A B C D E F 1of6	128	1	
2	A C D 1of3	128		1
3	B E F 1of3	128	1	1
4	A B 1of2	192	1	
5	C D 1of2	192		1
6	E F 1of2	192	1	1
7	C D E F 1of4	256	1	
8	B C D 1of3	256		1
9	A E F 1of3	256	1	1

La prueba consistió en asignarle a cada imagen un grupo elegido de forma pseudo-aleatoria y ejecutar los requerimientos de seguridad seleccionados para un grupo de seguridad dado, esto fue realizado por el productor, después los consumidores realizaron la adquisición de las imágenes de los grupos a los que pertenecen.

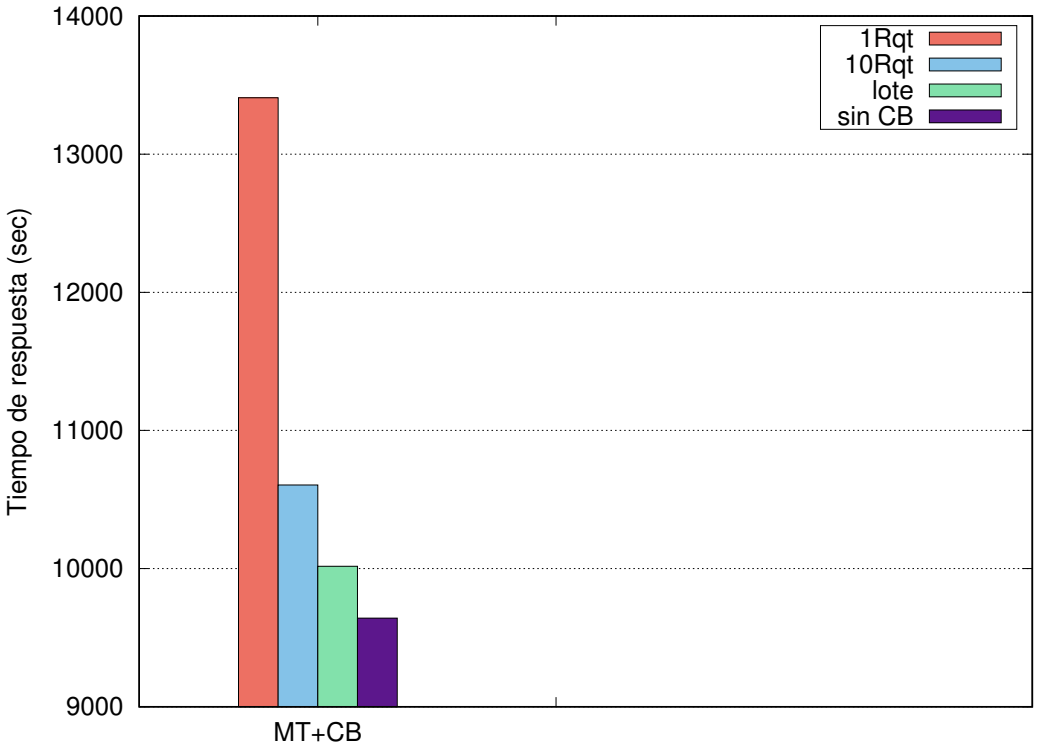

Figura 4.12: Sobre coste del registro de operaciones en la cadena de bloques (BC).

Este estudio de caso fue conducido para evaluar los sobre costes de registrar en la blockchain cada transacción realizada en cada intercambio de contenidos efectuado por las configuraciones estudiadas.

La Figura 4.12 muestra, en el eje vertical, el tiempo de respuesta producido por la configuración Maestro Trabajador ($MT + CB$) incluyendo el registro de transacciones en la blockchain. Tres configuraciones son mostradas en esta figura. En la primera (1RQT), la solución MT+CB contabiliza el costo de registrar una transacción realizada por cada operación ejecutada por los 10 diez

trabajadores del patrón evaluado en este experimento (que trabajaban en forma concurrente dentro del patrón). Esto quiere decir 1RQT consolida 10 registros realizados en forma concurrente. Esta solución produjo un sobre coste de 39.06 % con respecto a MT sin realizar registro alguno. Sin embargo es importante tomar en cuneta que el mejoramiento del rendimiento de los patrones de paralelismo permitirían absorber este sobre coste.

Las configuraciones 10RQT y lote no son realistas ya que postergan el registros de los eventos, lo que no sucede con 1 RQT.

La configuración 10 Rqt consolida 100 transacciones en cada registro, lo cual produce un sobre coste del 10 %, mientras que cuando solo se hace un registro del lote completo el sobre coste resulta ser insignificante.

Se detectó que el coste de realizar registros a la cadena de bloques (blockchain) es constante.

5

Conclusiones

En el presente trabajo de tesis se propuso un método de registro de transacciones para cadenas de valor en sistemas de compartición de archivos sensibles. La implementación del método propuesto brinda de forma eficiente los servicios de seguridad (confidencialidad, control de acceso, no repudio e integridad) a los archivos sensibles que viajan a través de cadenas de valor, así mismo permite la trazabilidad de los mismos durante su ciclo de vida. De esta forma este trabajo descansa sobre tres pilares principales: seguridad (criptografía), rendimiento (paralelismo) y trazabilidad (cadena de bloques o blockchain)

En la evaluación experimental se pudo observar de forma cualitativa lo siguiente:

- Flexibilidad: Las entidades pueden elegir entre diferentes configuraciones de seguridad y rendimiento, se pueden variar los niveles de seguridad, elegir entre cifrado (control de acceso y confidencialidad) o firma (integridad y no repudio) o ambos, los módulos que conforman este mecanismo puede ser intercambiables, solo es necesario definir entradas y salidas de cada uno. En la parte del rendimiento se puede elegir el número de trabajadores que realizarán el

procesamiento.

- Eficacia: El método mostró ser eficaz en realizar todas las tareas asignadas con cualquier configuración de parámetros aceptados.

- Factibilidad: Durante los estudios de caso se trabajó con elementos de incertidumbre que podrían aparecer en escenarios reales y el método funcionó eficaz y eficientemente en todos los casos.

De forma cuantitativa se observó lo siguiente:

El uso de patrones de paralelismo en el ambiente descrito en este trabajo de tesis pudo mejorar con 10 trabajadores hasta un 80 % del rendimiento comparándolo con la versión secuencial.

También se observó que para lotes de archivos grandes $> 100MB$ y niveles de seguridad 192 , 256 el enfoque solapado ofrece mejor rendimiento que la versión manejador-trabajador, caso contrario para archivos menores y niveles de 128 y 192 el enfoque manejador trabajador mostró un mejor rendimiento.

Al comparar el enfoque manejador-trabajador con tuberías creadas mediante una solución del estado del arte (Jenkins) se obtuvo una ganancia de hasta el 56 % del manejador-trabajador sobre Jenkins con niveles de 128, sin embargo utilizando el nivel de seguridad 256 se logró una ganancia del 25 %.

El mejoramiento del rendimiento por parte de los patrones de paralelismo permiten al método absorber los costes de registras transacciones en el la blockchain

5.1 Contribuciones

- Esquema de creación de sobres digitales mediante patrones de paralelismo basados en contenedores virtuales.

- Mecanismo de registro de seguimiento de transacciones en cadenas de valor mediante bloques encadenados.
- Mecanismo para fusionar esquemas de manejo de productos digitales en cadenas de valor con la creación de sobres digitales y la red de registro de transacciones.

5.2 Limitaciones

- El crecimiento del número de trabajadores aumentan el rendimiento, y el aumento del rendimiento esta limitado a los recursos físicos.
- El uso del método propuesto en sistemas de compartición de archivos sensibles podría enfrentar barreras impuestas por la asimilación de esta tecnología en el campo social y laboral asì como implicaciones legales en diferentes ámbitos.

5.3 Trabajo futuro

Como parte del trabajo futuro se mencionan las siguientes tareas:

- Tener un sistema de tolerancia a fallos en los patrones de paralelismo.
- Automatizar el despliegue de la red de la cadena de bloques.

Bibliografía

[1] Abdullah, M., Iqbal, W., and Bukhari, F. (2018). Containers vs virtual machines for auto-scaling multi-tier applications under dynamically increasing workloads.

[2] Aloraini, A. and Hammoudeh, M. (2017). A survey on data confidentiality and privacy in cloud computing. In *Proceedings of the International Conference on Future Networks and Distributed Systems*, ICFNDS '17, pages 10:1–10:7, New York, NY, USA. ACM.

[3] Appari, A. and Johnson, M. E. (2010). Information security and privacy in healthcare: current state of research. *International Journal of Internet and Enterprise Management*, 6:279–314.

[4] Azaria, A., Ekblaw, A., Vieira, T., and Lippman, A. (2016). Medrec: Using blockchain for medical data access and permission management. In *2016 2nd International Conference on Open and Big Data (OBD)*, pages 25–30.

[5] Benhamouda, F., Halevi, S., and Halevi, T. (2018). Supporting private data on hyperledger fabric with secure multiparty computation. In *2018 IEEE International Conference on Cloud Engineering (IC2E)*, pages 357–363.

[6] Bethencourt, J., Sahai, A., and Waters, B. (2007). Ciphertext-policy attribute-based encryption. In *2007 IEEE Symposium on Security and Privacy (SP '07)*, pages 321–334.

[7] Castro, M., Morán, G., Navarrete, D., Cruzatty, J., Anzúles, G., Mero, C., Quimiz, , and Merino, M. (2018). *Introducción a la seguridad informática y el análisis de vulnerabilidades*.

[8] Celesti, A., Mulfari, D., Galletta, A., Fazio, M., Carnevale, L., and Villari, M. (2019). A study on container virtualization for guarantee quality of service in cloud-of-things. *Future Generation Computer Systems*.

[9] Crosby, M., Nachiappan, Pattanayak, P., Verma, S., and Kalyanaraman., V. (2015). Blockchain technology beyond bitcoin. *Sutardja Center for Entrepreneurship and Technology Technical Report*.

[10] Crosby, M., P. P. V. S. and Kalyanaraman (2015). Blockchain technology beyond bitcoin.

[11] Dagher, G. G., Mohler, J., Milojkovic, M., and Marella, P. B. (2018). Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable Cities and Society*, 39:283 – 297.

[12] Dr. Tempich, C., Bodenbenner, P., and Feuerstein, L. (2011). Turning data into profit success factors in data-centric business models.

[13] Gillani, K. and Lee, J.-H. (2019). Comparison of linux virtual machines and containers for a service migration in 5g multi-access edge computing. *ICT Express*.

[14] Jakobsson, M. and Juels, A. (1999). *Proofs of Work and Bread Pudding Protocols(Extended Abstract)*, pages 258–272. Springer US, Boston, MA.

[15] Johnson, C. W. (2016). Why we cannot (yet) ensure the cybersecurity of safety-critical systems.

[16] Katz, J. and Lindell, Y. (2014). *Introduction to Modern Cryptography, Second Edition*. Chapman & Hall/CRC, 2nd edition.

[17] Kaur, H., Alam, A., Jameel, R., Mourya, A., and Chang, V. (2018). A proposed solution and future direction for blockchain-based heterogeneous medicare data in cloud environment. *Journal of Medical Systems*, 42:156.

[18] Li, H., Zhu, L., Shen, M., Gao, F., Tao, X., and Liu, S. (2018). Blockchain-based data preservation system for medical data. *J. Med. Syst.*, 42(8):1–13.

[19] McCool, M., Reinders, J., and Robison, A. (2012). *Structured Parallel Programming: Patterns for Efficient Computation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition.

[20] Mell, P. M. and Grance, T. (2011). Sp 800-145. the nist definition of cloud computing. Technical report, Gaithersburg, MD, United States.

[21] Menezes, A. J., Vanstone, S. A., and Oorschot, P. C. V. (1996). *Handbook of Applied Cryptography*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition.

[22] Morales-Sandoval, M. and Diaz-Perez, A. (2015). Det-abe: A java api for data confidentiality and fine-grained access control from attribute based encryption. pages 104–119.

[23] Morales-Sandoval, M., Gonzalez-Compean, J. L., Diaz-Perez, A., and Sosa-Sosa., V. J. (2017). A pairing-based cryptographic approach for data security in the cloud. *International Journal of Information Security*.

[24] Nakamoto, S. (2009). Bitcoin: A peer-to-peer electronic cash system.

[25] Nemati, H. (2007). Information security and ethics: Concepts, methodologies, tools, and applications.

[26] Okoye, H. (2015). Understanding healthcare's value chain.

[27] Paar, C. and Pelzl, J. (2009). *Understanding Cryptography: A Textbook for Students and Practitioners*. Springer Publishing Company, Incorporated, 1st edition.

[28] Pitta, D. and Laric, M. (2004). Value chains in health care. *Journal of Consumer Marketing*, 21:451–464.

[29] Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. Free Press, New York and London.

[30] Rodriguez-Henriquez, F. (2006). *Cryptographic Algorithms on Reconfigurable Hardware (Signals and Communication Technology)*. Springer.

[31] Sahai, A. and Waters, B. (2005). Fuzzy identity-based encryption. In Cramer, R., editor, *Advances in Cryptology – EUROCRYPT 2005*, pages 457–473, Berlin, Heidelberg. Springer Berlin Heidelberg.

[32] Salveggio, E. (2004). Your (un)reasonable expectations for privacy. *Ubiquity*, 2004(April):1–1.

[33] Sierra, J. Y. (2015). *Servicio seguro en un entorno de nube privada para almacenar, compartir y buscar documentos electrónicos*. Tesis de maestría del CINVESTAV Tamaulipas.

[34] Smart, J. F. (2011). *Jenkins: The Definitive Guide*. O'Reilly Media, Inc.

[35] Sukhodolskiy, I. and Zapechnikov, S. (2018). A blockchain-based access control system for cloud storage. In *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pages 1575–1578.

[36] Taylor, I. (2007). *Workflows for e-science : scientific workflows for grids*. Springer, London.

[37] vurukonda, N. and Rao, B. T. (2016). A study on data storage security issues in cloud computing. *Procedia Computer Science*, 92:128 – 135. 2nd International Conference on Intelligent Computing, Communication and Convergence, ICCC 2016, 24-25 January 2016, Bhubaneswar, Odisha, India.

[38] Wang, H. and Song, Y. (2018). Secure cloud-based ehr system using attribute-based cryptosystem and blockchain. *Journal of Medical Systems*, 42(8):152.

[39] Wang, V., Button, M., K, F., Motha, J, S., and Y, W. (2018). Cyber security breaches survey 2018.

[40] Whitman, M. E. and Mattord, H. J. (2011). *Principles of Information Security*. Course Technology Press, Boston, MA, United States, 4th edition.

[41] Yadav, A., Garg, M., and Mehra, R. (2019). *Docker Containers Versus Virtual Machine-Based Virtualization: Proceedings of IEMIS 2018, Volume 3*, pages 141–150.

[42] Yanez-Sierra, J., Diaz-Perez, A., Sosa-Sosa, V., and Gonzalez, J. L. (2015). A digital envelope scheme for document sharing in a private cloud storage. In *2015 12th International Conference Expo on Emerging Technologies for a Smarter World (CEWIT)*, pages 1–6.

[43] Yang, H. and Yang, B. (2016). A blockchain-based approach to the secure sharing of healthcare data. *Department of Information Security and Communication Technology, CCIS*.