Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ciencias Matemáticas
Unidad de Posgrado

"Simulación de un modelo de autómata celular para el tratamiento del problema del cifrado simétrico de la información"

TESIS

Para optar el Grado Académico de Magíster en Investigación de Operaciones y Sistemas

AUTOR

Pablo Edwin LÓPEZ VILLANUEVA

ASESOR

Mg. Inés GAMBINI LÓPEZ

Lima, Perú

2019

Referencia bibliográfica

López, P. (2019). *Simulación de un modelo de autómata celular para el tratamiento del problema del cifrado simétrico de la información*. Tesis para optar grado de Magíster en Investigación de Operaciones y Sistemas. Unidad de Posgrado, Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

HOJA DE METADATOS COMPLEMENTARIOS

CODIGO ORCID DEL AUTOR: 0000-0002-6198-9102

CODIGO ORCID DEL ASESOR: 0000-0003-4259-8808

DNI: 32920084

GRUPO DE INVESTIGACIÓN:

INSTITUCIÓN QUE FINANCIA PARCIAL O TOTALMENTE LA INVESTIGACIÓN:

UBICACIÓN GEOGRÁFICA DONDE SE DESARROLLÓ LA INVESTIGACIÓN. DEBE INCLUIR LOCALIDADES Y COORDENADAS GEOGRÁFICAS

Código Ubigeo (INEI)	:	150101
Departamento	:	LIMA
Provincia	:	LIMA
Distrito	:	LIMA
Longitud y Latitud	:	**-12.0431805, -77.0282364**

AÑO O RANGO DE AÑOS QUE LA INVESTIGACIÒN ABARCÓ:

2016 - 2018

ACTA DE SUSTENTACIÓN DE TESIS DE GRADO ACADÉMICO DE MAGÍSTER

Siendo las, 17:15 horas del día viernes veintidós de febrero del dos mil diecinueve, en el Auditorio de la Facultad de Ciencias Matemáticas, el Jurado Evaluador de Tesis, Presidido por la Mg. Carmela Catalina Velásquez Pino e integrado por los siguientes miembros, Mg. Daniel Quinto Pazce (Jurado Informante), Mg. Paulo César Olivares Taipe (Jurado Evaluador), Mg. Carlos Ortega Muñoz (Jurado Evaluador), y la Mg. Inés Gambini López como Miembro Asesor, se reunieron para la sustentación de la tesis titulada: «SIMULACIÓN DE UN MODELO DE AUTÓMATA CELULAR PARA EL TRATAMIENTO DEL PROBLEMA DEL CIFRADO SIMÉTRICO DE LA INFORMACIÓN» presentada por el Bachiller **Pablo Edwin López Villanueva** para optar el Grado Académico de Magíster en Investigación de Operaciones y Sistemas con Optimización de Sistemas de Mercadeo y Producción.

Luego de la exposición del graduando, los Miembros del Jurado hicieron las preguntas correspondientes, así como las observaciones e inquietudes acerca del trabajo de tesis, a las cuales el Bachiller Pablo Edwin López Villanueva respondió con acierto y solvencia, demostrando pleno conocimiento del tema.

A continuación se realizó la calificación correspondiente, según tabla adjunta, resultando el Bachiller Pablo Edwin López Villanueva aprobado con el calificativo de BUENO (16).

Habiendo sido aprobada la sustentación de la Tesis, el Jurado Evaluador recomienda para que el Consejo de Facultad apruebe el otorgamiento del Grado Académico de **Magíster en Investigación de Operaciones y Sistemas con mención en Optimización de Sistemas de Mercadeo y Producción** al **Bachiller Pablo Edwin López Villanueva**.

Siendo las 18:30 horas, se levantó la sesión, firmando para constancia la presente Acta.

Mg. Daniel Quinto Pazce
Miembro

Mg. Carmela Catalina Velásquez Pino
Presidenta

Mg. Carlos Ortega Muñoz
Miembro

Mg. Paulo César Olivares Taipe
Miembro

Mg. Inés Gambini López
Miembro Asesor

Dedicatoria

A mis padres Edilberto y Amalia:
Mis grandes maestros que me inculcaron enseñanzas y valores de honestidad, respeto y lealtad.

A mis hijos Sebastián y Carolina
Motivos de inspiración para conseguir objetivos en la vida.

Agradecimientos

A la profesora Mg. Inés Gambini López
Mi reconocimiento por su consejería y apoyo en la asesoría de la presente investigación.

A la Plana Docente de la Maestría en Investigación de Operaciones y Sistemas de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos.

Índice General

Lista de Figuras

Lista de Tablas

RESUMEN

Desde épocas remotas, el hombre ha desarrollado una gran variedad de técnicas y procedimientos orientados a la protección y seguridad de la información de modo que garanticen su privacidad, confidencialidad.

La información representa el insumo principal que genera conocimiento, base fundamental para la óptima toma de decisiones. De este modo la información adopta un rol preponderante constituyéndose en un factor crítico de éxito en personas, organizaciones y sociedades.

El tratamiento y solución del problema del cifrado de la información usando criptografía simétrica, será enfocado desde una perspectiva de la investigación de operaciones, cuyo campo de estudio principal es el modelado y simulación de sistemas dinámicos complejos.

Para tal efecto se va a desarrollar e implementar un criptosistema usando un autómata celular circular periódico como modelo representativo de un sistema dinámico complejo que evoluciona a pasos discretos en el tiempo a partir de reglas de transición de estados.

La información representada como archivos y contenida en los repositorios o unidades de almacenamiento externo en computadoras personales y organizacionales, servidores remotos en la nube, etc. representa el objeto de estudio de la investigación.

El modelado y simulación de un autómata celular asociado al desarrollo e implementación del sistema criptográfico adopta un contexto multidisciplinario e involucra el estudio y aplicación de importantes disciplinas tales como investigación de operaciones, teoría de sistemas, matemáticas, estadística, optimización, entre otras importantes áreas del conocimiento.

PALABRAS CLAVE: Sistema, modelo, simulación, autómata, vecindad, similaridad, grupo, byte.

ABSTRACT

Since remote times, man has developed a wide variety of techniques and procedures oriented at the protection and security of information in order to guarantee privacy and confidentiality.

The information represents the main input that generates knowledge, fundamental basis for optimal decision making. In this way, information plays a preponderant role, constituting a success critical factor in people, organizations and societies.

The treatment and solution of the problem of information encryption using symmetric cryptography, will be focused from a perspective of operations research, whose main field of study is the modeling and simulation of complex dynamic systems.

For this purpose, a cryptosystem will be developed and implemented using a periodic circular cellular automaton as a representative model of a complex dynamic system that evolves at discrete steps in time from state transition rules.

The information represented as files and contained in the repositories or external storage units in personal and organizational computers, remote servers in the cloud, etc. represents the object of study of the investigation.

The modeling and simulation of a cellular automaton associated with the development and implementation of the cryptographic system adopts a multidisciplinary context and involves the study and application of several important disciplines such as research operations, systems theory, mathematics, statistics, optimization, among other important areas of knowledge.

KEY WORDS: System, model, simulation, automaton, neighborhood, similarity, group, byte.

Capítulo 1

INTRODUCCIÓN

1.1 SITUACIÓN PROBLEMÁTICA

En las últimas décadas, las organizaciones y la sociedad en general, han visto afectadas su estructura y comportamiento debido al continuo desarrollo tecnológico, que ha generado impacto en sus componentes y procesos.

En las organizaciones se han evidenciado cambios profundos en las operaciones y sistemas que lo constituyen. Por otro lado, en la sociedad, estas transformaciones han propiciado el surgimiento de nuevas innovaciones tales como sociedades de la información y del conocimiento.

La información es un componente de vital importancia, puesto que describe la estructura y comportamiento de los sistemas en la información, permitiendo su funcionamiento y puesta en marcha[1]. En consecuencia, representa un recurso estratégico cuya disponibilidad oportuna permitirá la adecuada toma de decisiones.

En la actualidad un gran desafío para las organizaciones y la sociedad en general, es garantizar que la información sea confidencial y reservada[2].

La confidencialidad de la información consiste de manera estricta en su privacidad. Por otro lado, el carácter de reserva de la información se refiere a las restricciones establecidas para su utilización.

---

1 La puesta en marcha de una operación o sistema en la organización, es una fase que representa su implantación o puesta en producción.

2 La protección y seguridad de la información en algún contexto de uso y aplicación, es un derecho que se encuentra normado en la legislación de algunos países de modo que permita regular su utilización. En la legislación peruana la Ley Nº 29733 (Ley de Protección de Datos Personales) garantiza la protección de los datos personales.

En este trabajo de investigación se utiliza una metodología basada en el enfoque de la investigación de operaciones y teoría de sistemas, para resolver el problema del cifrado simétrico de la información mediante un modelo de autómata celular.

El ámbito o escenario de aplicación de este trabajo, corresponde a cualquier organización, tal como una universidad, entidad pública o empresa privada, en la cual la información se encuentra localizada en unidades externas de almacenamiento en sus computadores locales o servidores remotos.

Es conveniente señalar que la atención y solución del problema descrito en éste trabajo de investigación adopta niveles de gran importancia, debido al rol y carácter que representa la información como importante recurso estratégico para la oportuna toma de decisiones en la organización.

Así mismo, es importante indicar que este trabajo posee un alto grado de interés y gran novedad en la actualidad, debido al importante papel que desempeña la información en todos los ámbitos y contextos de aplicación.

Por otro lado, la solución del problema planteada en la presente investigación adopta múltiples criterios, entre los cuales vale mencionar prioridad, factibilidad y viabilidad.

El problema descrito en este documento, demanda una atención prioritaria, debido a que la garantía de la confidencialidad y reserva de la información incrementa la confianza de los usuarios, maximizando el rendimiento.

Por otro lado, la solución establecida a través de un sistema criptográfico o criptosistema, es factible y viable debido a que su desarrollo e implementación no involucra un costo[3].

3 En la sección Anexo 1 de este documento, se presenta una Propuesta Técnica-Económica correspondiente al desarrollo e implementación del Sistema de Confidencialidad y Reserva de la Información para la UNMSM.

1.2 FORMULACIÓN DEL PROBLEMA

1.2.1 PROBLEMA GENERAL

El problema general de la investigación, se formula del siguiente modo:

P_G ¿El problema del cifrado simétrico de la información puede ser tratado mediante la simulación de un modelo de autómata celular?

1.2.2 PROBLEMAS ESPECIFICOS

La formulación del problema general deriva en la formulación de los siguientes problemas específicos:

P_{E1} ¿Se puede diseñar el modelo del autómata celular para el problema del cifrado simétrico de la información?

P_{E2} ¿Se puede construir el modelo del autómata celular para el problema del cifrado simétrico de la información?

1.3 HIPÓTESIS Y VARIABLES

1.3.1 HIPÓTESIS GENERAL

La hipótesis general correspondiente al problema general P_G, permite realizar la confirmación o rechazo y es formulada del siguiente modo:

H_{G1} ¿El problema del cifrado simétrico de la información si puede ser tratado mediante la simulación de un modelo de autómata celular?

H_{G0} ¿El problema del cifrado simétrico de la información no puede ser tratado mediante la simulación de un modelo de autómata celular?

1.3.2 HIPÓTESIS ESPECÍFICAS

A partir de la hipótesis general, se han formulado las siguientes hipótesis específicas asociadas a los problemas específicos P_{E1} y P_{E2} respectivamente:

H_{E11} ¿Si se puede diseñar el modelo del autómata celular para el problema del cifrado simétrico de la información?

H_{E10} ¿No se puede diseñar el modelo del autómata celular para el problema del cifrado simétrico de la información?

H_{E21} ¿Si se puede construir el modelo del autómata celular para el problema del cifrado simétrico de la información?

H_{E20} ¿No se puede construir el modelo del autómata celular para el problema del cifrado simétrico de la información?

1.3.3 UNIDAD DE ANÁLISIS

La unidad de análisis de la investigación es determinada por el elemento u objeto de estudio que es representado por cada fichero[4] de disco almacenado en un repositorio o unidad externa en computadores organizacionales o servidores remotos.

1.3.4 IDENTIFICACIÓN DE VARIABLES

Las variables asociadas a la investigación, pueden ser clasificadas como independientes y dependientes.

Las variables independientes son aquellas que representan la entrada, tales como:

- *Tamaño del Fichero Inicial*, que constituye el tamaño en bytes del fichero de entrada a codificar.
- *Longitud de la Clave*, que representa el tamaño en bytes de la clave utilizada para realizar el cifrado simétrico.
- *Longitud del Autómata*, que hace referencia al tamaño en bytes de la estructura asociada al autómata celular.

El criptosistema genera el fichero codificado que será decodificado con la misma clave simétrica.

Por otro lado, las variables dependientes o de salida son aquellas que son el resultado del procesamiento. Entre ellas, se tiene

- *Tamaño del Fichero Final*, que constituye el tamaño en bytes del fichero decodificado.
- *Indicador de Equivalencia,* valor lógico que verifica la igualdad o desigualdad de los ficheros inicial y final.

4 El termino *archivo* denominado también *fichero* es la unidad de información digital que contiene nombre, tamaño, fecha de creación, atributos etc. y es gestionado por el sistema operacional subyacente.

1.3.5 OPERACIONALIZACIÓN DE VARIABLES

Los aspectos que definen la naturaleza de las variables y los mecanismos correspondientes a su tratamiento y procedimientos de obtención se determinan en esta sección:

De este modo, las variables especificadas en el punto anterior, van a adoptar sus valores según se indica:

A. Variables Independientes

- *Tamaño del Fichero Inicial*, es el tamaño en bytes del fichero de entrada a codificar. Representa el tamaño físico del fichero y es proporcionado por el sistema operativo subyacente.

- *Longitud de la Clave*, representa el tamaño en bytes de la clave que va a ser utilizada para el cifrado simétrico. La clave es generada mediante un proceso de simulación que maximiza la varianza.

- *Longitud del Autómata*, que hace referencia al tamaño de la estructura asociada al autómata celular. Este valor es proporcional al tamaño en bytes del fichero de entrada y es generado en tiempo de ejecución.

B. Variables Dependientes

Estas variables son el resultado de un proceso estrictamente computacional.

- *Tamaño del Fichero Final*, es el tamaño en bytes del fichero decodificado y es generado por el criptosistema.

- *Indicador de Equivalencia,* valor lógico que compara los ficheros inicial y final para determinar su igualdad o desigualdad. Es obtenido por un proceso de validación en el criptosistema.

1.4 JUSTIFICACIÓN TEÓRICA

El contexto tecnológico en la actualidad, se caracteriza por el flujo y tratamiento de grandes cantidades de información, generando en las organizaciones la necesidad de disponer información confidencial y reservada.

El problema formulado en el presente trabajo de investigación adopta un carácter multidisciplinario[5], con principal énfasis en la investigación de operaciones y teoría de sistemas.

Estas importantes disciplinas se han visto fortalecidas con el advenimiento y continuo desarrollo de la computación electrónica, y tienen como campo principal de estudio a los sistemas dinámicos complejos[6] que involucra a numerosos fenómenos del mundo real.

En el contexto de la investigación de operaciones y teoría de sistemas, el análisis, modelado y simulación representan las principales fases[7] para solucionar un problema enfocado como sistema dinámico complejo y consecuentemente se encuentran estrechamente relacionados.

El análisis de un sistema consiste en describir su estructura y comportamiento, identificando sus propiedades, las interrelaciones existentes y cuantificando los mecanismos o procedimientos conducentes a su tratamiento y su resolución.

El modelado permite representar al sistema real mediante un prototipo, adoptando diferentes configuraciones o interpretaciones tales como modelado lógico y modelado físico. En este documento el modelado lógico es expresado como diseño del modelo, y el modelado físico es referido como construcción del modelo.

---

5 Un problema es multidisciplinario, cuando su tratamiento y resolución involucra la participación y aplicación de diferentes disciplinas o áreas del conocimiento. Diferentes problemas de Investigación de Operaciones y Teoría de Sistemas adoptan esta característica.

6 En el ámbito de la computación científica y las ciencias exactas tales como el análisis numérico o física teórica, los sistemas dinámicos permiten modelar y analizar ciertos fenómenos del mundo real. No obstante, en este trabajo se adopta la expresión sistema dinámico complejo para hacer referencia a los sistemas complejos que de modo implícito adoptan naturaleza dinámica en el tiempo.

7 En el capítulo 3 se expone la metodología de desarrollo e implementación de la solución del problema desde una perspectiva de la investigación de operaciones y teoría de sistemas.

A través de la simulación es posible aproximar el modelo construido al sistema real, permitiendo su adaptación a diferentes contextos mediante las alteraciones de sus reglas lógicas y los valores en sus parámetros.

En este trabajo, el problema del cifrado simétrico de la información es abordado desde un enfoque de la investigación de operaciones y teoría de sistemas, a través de la utilización de un modelo de autómata celular asociado a un sistema criptográfico.

Un autómata celular es un modelo matemático y computacional cuya simulación y funcionalidad describe a un sistema dinámico complejo con evolución en el tiempo.

Una importante aplicación de los autómatas celulares son los sistemas criptográficos simétricos, poderosas herramientas computacionales que permiten mantener la confidencialidad y reserva de la información.

Von Neumann (1966), uno de los grandes pioneros y artífices de la computación, en su publicación, describe el uso de modelos de autómatas celulares como una gran alternativa a la aplicación de métodos analíticos tales como modelos de ecuaciones diferenciales cuyo rigor no hacia posible la construcción de sistemas auto-reproductivos.

Un modelo de autómata celular es diseñado a partir de un conjunto de reglas simples que definen la transición de estados generando su evolución en el tiempo y facilitando su implementación computacional.

La solución del problema tiene como soporte una base teórica y su desarrollo ha sido realizado mediante una metodología basada en un enfoque de la investigación de operaciones y teoría de sistemas.

Así mismo, es importante indicar que esta investigación, representa a nivel nacional el primer trabajo en su categoría, sobre el tratamiento del problema de cifrado simétrico de la información utilizando modelos de autómatas celulares.

1.5 JUSTIFICACIÓN PRÁCTICA

En las organizaciones, un importante recurso estratégico es la información, que además de ser un factor crítico que crea valor, permite la oportuna y adecuada toma de decisiones. Por lo tanto, su gestión y tratamiento requiere la incorporación de técnicas o procedimientos que permitan establecer su confidencialidad y reserva mediante un sistema criptográfico.

Los autómatas celulares son modelos matemáticos y computacionales que evolucionan en el tiempo a partir de ciertas reglas de transición pre-establecidas.

Para cada regla de transición del autómata celular, es posible desarrollar un procedimiento que permita adecuar la técnica de cifrado y descifrado de la información personalizando de algún modo la solución.

La mayoría de las aplicaciones existentes de software, tales como los sistemas de gestión y manejo de bases de datos DBMS[8], realizan la gestión y tratamiento de la información en modo cifrado como un mecanismo de protección y seguridad de la información, los cuales son totalmente desconocidos por los usuarios.

No obstante, el mecanismo de cifrado y descifrado de la información utilizando modelos de autómatas celulares, constituye una alternativa de solución viable y a bajo costo.

Por otro lado, vale mencionar que la implementación de cada una de las fases se ha realizado en el lenguaje Java. Los algoritmos correspondientes más representativos han sido incorporados a este trabajo de investigación.

8 DBMS, acrónimo de Database Management System, *Sistema de Gestión de Base de Datos*.

1.6 OBJETIVOS

1.6.1 OBJETIVO GENERAL

Efectuar el tratamiento del problema del cifrado simétrico de la información mediante la simulación de un modelo de autómata celular.

1.6.2 OBJETIVOS ESPECÍFICOS

Se han considerado los siguientes:

- Efectuar el diseño del modelo usando un autómata celular.
- Efectuar la construcción del modelo basado en autómata celular.

1.7 ALCANCE

Dada la naturaleza de la investigación orientada a garantizar la confidencialidad y reserva de la información, el alcance de la investigación puede ser extendida a cualquier organización, tal como una entidad pública, empresa privada o universidad con generalización inclusive hacia la sociedad en general.

1.8 LIMITACIONES

La investigación contiene una serie de restricciones tales como:

- ✓ El código fuente en su totalidad no ha sido incorporado al documento correspondiente a esta investigación[9]. No obstante, se hace mención a ciertas unidades funcionales denominados *instrumentos*[10] conformados por procesos automatizados.

- ✓ No se ha considerado algún mecanismo de compactación o compresión de la información. Para futuras versiones, se recomienda su implementación.

- ✓ El muestreo adoptado es múltiple. En su fase preliminar es por conveniencia, y en la siguiente fase es probabilístico.

9 Por criterios de protección a la propiedad del autor para una futura patente.
10 Todos los instrumentos indicados han sido elaborados por el autor.

Capítulo 2

MARCO TEÓRICO

2.1 MARCO EPISTEMOLÓGICO

En esta sección se expone una perspectiva epistemológica sustentada en el método científico aplicado a este trabajo de investigación.

La presente investigación se caracteriza por su evidente naturaleza científica y tecnológica, dada su orientación a la protección y seguridad de la información.

2.1.1 LA RELACIÓN ENTRE CIENCIA Y FILOSOFÍA

En su publicación, Alex Rosenberg (2005) establece que la ciencia, comenzó desde la antigua Grecia. Desde aquella época, la ciencia y la filosofía estaban estrechamente relacionadas y compartidas. En la actualidad la ciencia está separada de la filosofía, surgiendo como una disciplina separada.

Según lo indicado por Gracia (2005), “las relaciones entre ciencia y filosofía han sido siempre difíciles. Durante toda la época clásica y medieval, la filosofía anuló a la ciencia como actividad autónoma, al definirla como episteme o saber apodíctico y deductivo. Ciencia venía a confundirse, de ese modo, con filosofía. Ello impidió el desarrollo de la ciencia experimental hasta bien entrado el mundo moderno. Cuando, a partir del siglo XVII, irrumpe con fuerza la ciencia moderna, se inicia un nuevo proceso que culminaría en el siglo XIX, al convertirse la filosofía en teoría de la ciencia. Ahora es la ciencia la que anula a la filosofía. Tal fue la obra del positivismo”.

De acuerdo a lo señalado por los autores, es conveniente indicar que a lo largo del espacio y del tiempo, la ciencia y filosofía han desempeñado un importante papel en la búsqueda del conocimiento siendo materia de estudio, investigación y discusión.

2.1.2 CIENCIA Y TECNOLOGIA

Desde su existencia, el hombre ha desarrollado y perfeccionado diferentes técnicas con la finalidad de establecer y afianzar su dominio en la naturaleza.

En su publicación, Uribe (2007) establece que “la primera revolución científica de la era moderna se dio en Europa en los siglos XVI y XVII”. Muchos científicos de aquella época medieval tales como Copérnico, Galileo, Kepler, Newton y Leibniz quebraron de manera radical el modelo aristotélico imperante acerca de la concepción del mundo, propiciando el desarrollo de la astronomía, física y matemáticas, que fueron fortalecidas en el siglo XVIII con la Revolución Industrial que tuvo su origen en Gran Bretaña.

La aplicación de las matemáticas al conocimiento y la inclusión de la metodología experimental afectaron a las otras ciencias tales como biología, química, geología y las ciencias sociales.

La tecnología adoptó un rol preponderante en esta revolución; sin embargo, la ciencia tuvo un papel con influencia indirecta. A partir de estas transformaciones, el siglo XX quedó marcado por una revolución que cambió de manera radical la forma de hacer ciencia y tecnología: la revolución tecno científica”.

El autor señala 3 fases que tuvieron incidencia:

- Macro ciencia o Big Science, que sucedió durante los años 30 hasta finales de los años 50
- Período de crisis, comprendido entre los años 60 hasta mediados de los años 70
- Revolución tecno científica, el cual corresponde desde los años 70 extendiéndose hasta la actualidad.

La tecnología para su desarrollo y perfeccionamiento, siempre ha requerido la aplicación de los fundamentos del conocimiento científico, y su evolución durante el transcurso del tiempo ha sido de lento y gradual con gran impacto en la transformación de la sociedad.

2.1.3 INFORMACIÓN, CONOCIMIENTO E INTELIGENCIA

Las bases del desarrollo y evolución del conocimiento tienen como soporte a la información. La aplicación del conocimiento propició la adquisición de la inteligencia, considerada como una capacidad perfeccionada para la resolución de problemas y conflictos, propiciando de este modo la transformación de las organizaciones y sociedades.

Expresado en otras palabras, existe una clara y evidente correspondencia entre estos conceptos. La información permite la obtención de conocimiento, y éste a su vez propicia la adquisición de inteligencia.

En su publicación, Ladyman (2002), considera desde una perspectiva filosófica que la ruptura y consecuente alejamiento con las teorías de Aristóteles (384-322 AC) fue el acontecimiento que cobró mayor importancia durante la revolución científica. Con el surgimiento y propuestas de nuevas ideas, los pensadores de la época iniciaron la búsqueda nuevos métodos para generar y descubrir conocimiento.

2.1.4 ENFOQUE MULTIDISCIPLINARIO

A largo de su existencia la ciencia y tecnología ha involucrado en su desarrollo diferentes áreas del conocimiento para el tratamiento de múltiples problemas que adoptan en su resolución criterios de complejidad y un alto grado de dificultad.

Esta característica señalada no es ajena al problema descrito en el presente trabajo. En consecuencia, constituye el enfoque o perspectiva multidisciplinaria.

El enfoque o visión multidisciplinaria aplicada al tratamiento de un problema consiste fundamentalmente en la aplicación de los fundamentos, aplicaciones y experiencias provenientes del saber de diferentes áreas de la ciencia y tecnología, de modo que permitan su entendimiento, análisis y resolución.

El presente trabajo de investigación adopta en su desarrollo un enfoque multidisciplinario, sustentado en el requerimiento de diferentes disciplinas, tales como optimización, estadística, computación, entre otros importantes campos del conocimiento.

2.2 ANTECEDENTES DE INVESTIGACIÓN

El surgimiento y evolución de la computación electrónica como poderosa herramienta tecnológica para la simulación de problemas han propiciado el desarrollo de numerosas técnicas de cifrado de la información, las cuales se encuentran publicados en numerosos artículos de investigación a nivel mundial.

De este modo, los trabajos previos relacionados al contexto de la investigación se pueden categorizar en el ámbito nacional y también a nivel internacional.

2.2.1 ANTECEDENTES NACIONALES

Cabe mencionar que, a nivel nacional, no se han identificado trabajos previos consistentes en artículos, tesis, libros, conferencias o publicaciones, que estén relacionados al problema general de la investigación.

Por lo tanto, es importante señalar que, en el ámbito nacional, esta investigación, constituye el primer trabajo en su categoría, esto es, el primer trabajo que desarrolla e implementa un criptosistema simétrico mediante la utilización de un modelo basado en autómata celular.

2.2.2 ANTECEDENTES INTERNACIONALES

En el ámbito internacional se han identificado trabajos previos en relación estrecha con el tema de la investigación. No obstante, las publicaciones asociadas establecen un enfoque o perspectiva para el tratamiento del problema descrito.

En las siguientes secciones, se describen las publicaciones de trabajos previos asociadas al tema de la investigación que están referenciados en la bibliografía correspondiente a este documento.

Su revisión y análisis han contribuido al desarrollo de la presente investigación.

2.2.2.1 Criptografía de Clave Secreta con Autómata Celular

En su publicación, Seredynski et al. (2003), describen un modelo de autómata celular unidimensional no uniforme que generan secuencias de números pseudoaleatorios (*Pseudorandom Number Sequences* - PNS), a partir de un conjunto de reglas utilizando una técnica de computación evolutiva denominada programación celular.

En la siguiente figura, se tiene N reglas que corresponden a una sola celda del CA que produce una secuencia PNS.

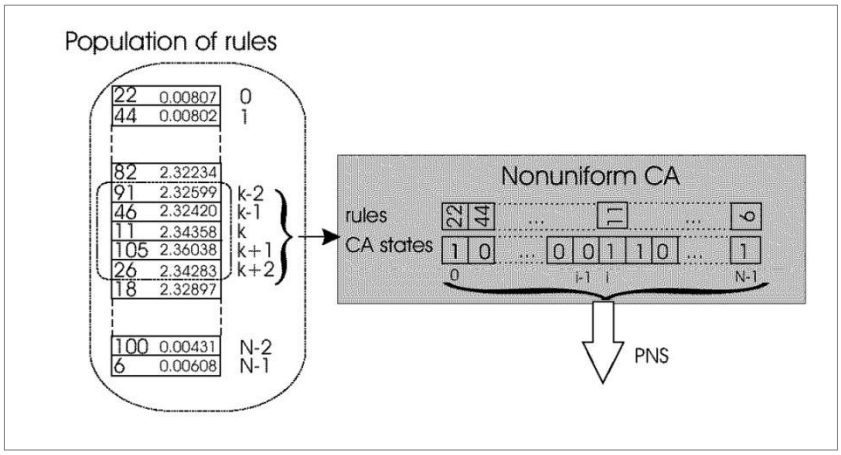

Figura 2.1 Entorno de Programación Celular para la evolución de reglas de Autómatas Celulares no uniformes. Fuente: Seredynski et al. (2003).

La calidad estadística de cada PNS es obtenida por la entropía E_h, dividiendo cada PNS en subsecuencias de tamaño $h = 4$

$$E_h = -\sum_{j=1}^{k^h} p_{h_j} \, log_2 p_{h_j}$$

$k = 2$ es el número de valores que toma cada elemento de una secuencia y p_{h_j} es la probabilidad de ocurrencia de una secuencia h_j en una PNS. La entropía E_h alcanza su máximo valor en $E_h = h$

CONCLUSIÓN

El modelo de autómata celular utilizado basado en la generación de números pseudoaleatorios, será incorporado a la investigación. No obstante, la entropía importante concepto en teoría de la información, que determina la incertidumbre, no será incluido en la investigación.

2.2.2.2 Autómatas celulares elementales para la encriptación de datos

En su publicación Villarreal et al. (2011), establecen un procedimiento de obtención de números pseudoaleatorios como mecanismo para generar una clave pseudoaleatoria.

Los autores definen un modelo de autómata celular denominado elemental caracterizado por ser lineal, adoptar una vecindad de tamaño 3 y asumir estados de transición binarios.

De este modo, para una vecindad con un esquema de 3 celdas binarias, se tienen 2^3=8 configuraciones de vecindad, a partir de las cuales se pueden establecer 2^8=256 reglas de transición, las cuales van a determinar la evolución del autómata celular.

Una de las reglas es la Regla 30, denominada así por la equivalencia de la secuencia 00011110 con el valor 30 decimal. El siguiente ejemplo ilustra esta regla.

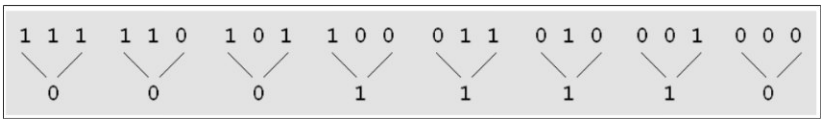

Figura 2.1 Regla 30. Fuente: Villarreal et al. (2011)

A partir de las 256 reglas obtenidas y considerando las transiciones de estado, a partir de una condición aleatoria, se establecieron las siguientes clases:

- *Clase 1*, denominada de tipo fijo, cuyas transiciones evolucionan hacia un estado homogéneo y estable en la cual todas celdas adoptan el mismo valor.
- *Clase 2*, denominada de tipo periódico, caracterizada por repetir el mismo patrón de transición en un bucle.
- *Clase 3*, conocida como pseudoaleatoria debido a que evoluciona hacia un patrón caótico.
- *Clase 4*, que adopta características de complejidad, dado que incorpora comportamientos de las clases 2 y 3.

Tomando en consideración los patrones pseudoaleatoridad, fueron seleccionadas las reglas 30, 54, 73 y 110.

Estas reglas fueron sometidas a una serie de pruebas o test de aleatoriedad de secuencias de bits, según se ilustra en la siguiente tabla.

Tabla 2.1 Pruebas de aleatoriedad para las reglas 30, 54, 73 y 110 (A: Aprobada R: Reprobada). Fuente: Villarreal et al. (2011).

Prueba	R30	R54	R73	R110
Frecuencia (Monobit)	A	R	R	R
Frecuencia dentro de un bloque	A	R	R	R
Corridas	R	R	R	R
Más larga corrida de unos en un bloque	A	R	R	R
Rango de la matriz binaria	A	R	A	R
Transformada discreta de Fourier (Espectral)	R	R	R	R
No acumulación de coincidencia de plantilla	A	A	A	A
Acumulación de coincidencia de plantilla	A	A	R	R
Estadística Universal de Maurer	A	A	R	R
Complejidad lineal	A	R	A	R
Serie	A	R	R	R
Entropía aproximada	A	R	R	R
Sumas acumulativas	A	R	R	R
Excursiones aleatorias	A	R	R	R
Variante de excursiones aleatorias	A	R	R	R

Luego, de acuerdo a los valores expresados en la tabla se concluye que la regla 30 es la que proporciona mayores características de pseudoaleatoriedad.

CONCLUSIÓN

De acuerdo a lo establecido en dicha publicación, vale señalar que el modelo de autómata celular elemental descrito, va a ser incorporado en el desarrollo de la investigación.

A este modelo de autómata celular lineal, se le va a incorporar una característica circular, de modo que todas las vecindades, incluyendo las extremas adopten la misma longitud de vecindad.

2.2.2.3 Cifrado de Imágenes usando Autómatas Celulares con Memoria

Hernández Encinas L. et al. (2004) realizan el cifrado de imágenes usando autómatas celulares reversibles con memoria definido

$$A = (C, S, V, f),$$

donde

- C es el espacio celular formado por celdas $< i >, 0 \leq i \leq n-1$
- S es el conjunto de estados de cada célula. $|S| = k$, y $S = \mathbb{Z}_k$
- V es el conjunto de índices de C con vecindad V_i dado por:

  $$V_i = \{< i + \alpha_1 >, \dots, < i + \alpha_m >\}\ \alpha_i \in V,\ V \subset \mathbb{Z} \text{ y } |V| = m,$$

- $f: S^m \to S$ es la función de transición del AC.

  Dado un instante t, un estado $a_i^{(t)}$ y un conjunto de estados $V_i^{(t)}$, el siguiente estado de la célula viene dado por:

  $$a_i^{(t+1)} = f\left(V_i^{(t)}\right) = f(a_{i+\alpha_1}^{(t)}, \dots, a_{i+\alpha_m}^{(t)})$$

La condición de contorno que determina la evolución del autómata es

$$Si\ i \equiv j\ (mod\ n)\ entonces\ a_i^{(t)} \equiv a_j^{(t)} (mod\ n)$$

Si C representa todas las configuraciones posibles, entonces la función de transición global es una transformación lineal

$$\Phi = C \to C, C^{(t)} \mapsto C^{(t+1)}$$

Si Φ es biyectiva entonces existe otro autómata celular, denominado inverso, con función global Φ^{-1}.

La función de transición del AC con memoria en el tiempo $t+1$ depende del estado de las células en instantes: $t,\ t-1,\ t-2$, etc.

$$a_i^{(t+1)} = \sum_{h=0}^{k} f^{(t-h)}(V_i^{(t-h)})$$

donde $f^{(t-h)}$ representa una función de transición local especifica.

El criptosistema propuesto por los autores consta de tres fases:

- *Inicialización*, que utiliza una semilla K de 1024 bits para obtener una secuencia pseudoaleatoria de $2n + 2$ bits.
- *Cifrado*, que define las primeras configuraciones: $C^0 = P$, $C^1 = Z$, y aplica la función de transición dada por

$$a_i^{(t+1)} = \sum_{j=-n/2}^{n/2} b_{n/2+j}^{(t \bmod 2))} a_{i+j}^{(t)} + a_i^{(t-1)} \pmod{c}, \quad 0 \le i \le n-1$$

  calculando las configuraciones $C^{(2)}$ y $C^{(3)}$. La imagen cifrada J, es la concatenación de J_1 y J_2 , cuyas representaciones secuenciales son las configuraciones $C^{(2)}$ y $C^{(3)}$.
- *Descifrado*, que calcula las dos configuraciones $\bar{C}^{(0)} = C^{(3)}$ y $\bar{C}^{(1)} = C^{(2)}$, y aplica el AC inverso definido por

$$a_i^{(t+1)} = -\sum_{j=-n/2}^{n/2} b_{n/2+j}^{(t \bmod 2))} a_{i+j}^{(t)} + a_i^{(t-1)} \pmod{c}, \quad 0 \le i \le n-1$$

  obteniendo $\bar{C}^{(3)} = P$ como representación de la imagen original.

El siguiente ejemplo, muestra una imagen en tono gris con tamaño de resolución de $512x512$ pixels y su imagen cifrada de $512x1024$ pixels.

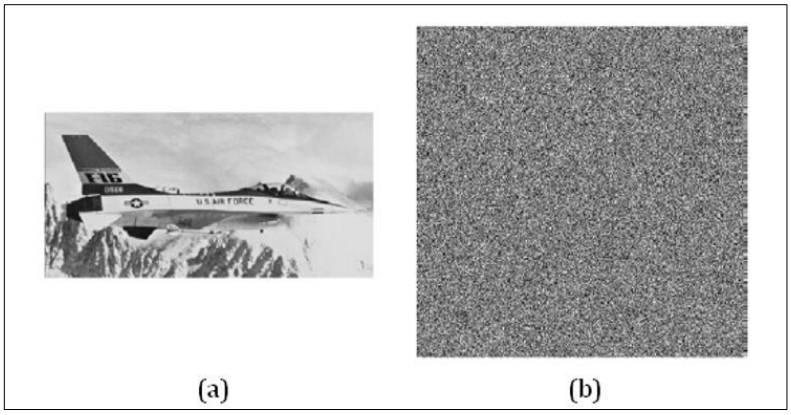

Figura 2.3 Cifrado de Imagen (a) original (b) cifrada Fuente: Hernández et al. (2004)

CONCLUSIÓN

Se adoptará el criterio de reversibilidad para el criptoanálisis como mecanismo de recuperación de la información. No obstante, el procedimiento utilizado en el artículo duplica el tamaño de la imagen original en la imagen cifrada constituyendo una desventaja.

2.3 BASES TEÓRICAS

La perspectiva multidisciplinaria aplicada al presente trabajo de investigación, requiere la incorporación de los fundamentos teóricos de algunas disciplinas, que son descritas en las siguientes secciones.

En la sección 2.3.1 se exponen una introducción a los sistemas, modelos y simulación. Se expone las nociones elementales de la teoría de sistemas que incluye conceptos básicos de su organización y propiedades. Así mismo, se proporciona los aspectos básicos de los modelos como representaciones de sistemas, su clasificación y los criterios a considerar para su diseño. Al final se efectúa una breve descripción de la simulación como técnica experimental, las condiciones y riesgos que involucra.

Por otro lado, la sección 2.3.2 establece una introducción a los autómatas celulares como modelos representativos de sistemas dinámicos complejos de tiempo discreto, que incluye la descripción de sus reglas de transición o funciones de evolución, tipos y modelos con aplicaciones en diferentes contextos y ámbitos del conocimiento.

Finalmente, en la sección 2.3.3 se hace una breve exposición de las técnicas y procedimientos de la criptografía que garantizan la codificación y protección de la información. Se estudian los tipos y modalidades, sus características y aplicaciones en los diferentes ámbitos del conocimiento.

2.3.1 SISTEMAS, MODELOS Y SIMULACIÓN

Los sistemas dinámicos complejos están asociados a los fenómenos del universo y del mundo real, con múltiples contextos de aplicación en diferentes ámbitos del conocimiento.

2.3.1.1 FUNDAMENTOS DE LA TEORIA DE SISTEMAS

La Teoría General de Sistemas, introducida por Von Bertalanffy (1989) a mediados del siglo XX, estudia a los sistemas desde una perspectiva multidisciplinaria que incluye el análisis de su estructura y comportamiento, los principios y las leyes que la rigen.

- **El Enfoque de Sistemas**

  Johansen (1993), establece el enfoque reduccionista que consiste en que un fenómeno complejo es estudiado a partir del análisis de sus componentes. No obstante, a medida que los sistemas se hacen más complejos, su estudio se hace más difícil dado que involucran más variables tales como el entorno, las interacciones entre sus componentes, entre otros.

- **Conceptos de Sistema**

  En Teoría de Sistemas, el concepto de sistema constituye la base fundamental para el estudio e investigación de diversos fenómenos del mundo real en diferentes contextos y aplicaciones. Existen diferentes percepciones que pretenden definirlo:

  - ✓ Johansen (1993), define a los sistemas como un "conjunto de partes coordinadas y en interacción para alcanzar un conjunto de objetivos".
  - ✓ Meadows (2008), establece que un sistema es un conjunto interconectado de elementos que está organizado coherentemente de una manera que logra alguna cosa.

  No obstante, en la presente investigación se considera a un sistema como "*un conjunto de elementos interrelacionados con estructura y parámetros definidos que se encuentran en interacción y orientados hacia la obtención de un objetivo determinado*".

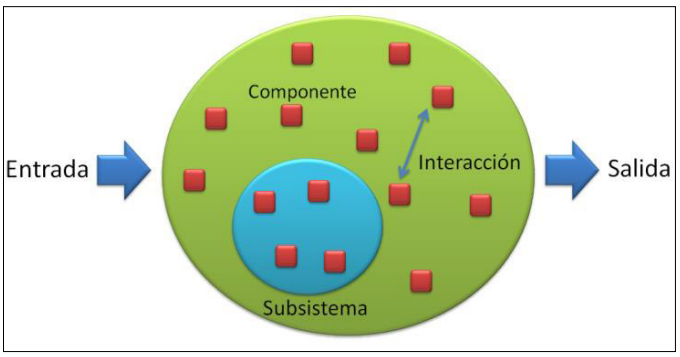

Figura 2.4 Representación básica de un Sistema. Fuente: Elaboración Propia.

- **Parámetros o Variables de Estado**

  Los parámetros del sistema son variables de estado que determinan el comportamiento del sistema en un instante determinado, esto es, la instancia de un sistema representa su estado en un instante determinado y es descrito a partir de los valores que adoptan los parámetros.

- **Procesos Internos y Externos**

  Los sistemas contienen procesos que generan alteraciones en sus parámetros o variables de estado. Se agrupan en procesos internos o endógenos cuando se realizan dentro del sistema y procesos externos o exógenos cuando provienen del medio externo.

- **Subsistema y Supersistemas**

  Un subsistema es un subconjunto de componentes del sistema que poseen estructura, naturaleza y comportamiento determinado, constituyéndose como un sistema más especializado.

  De igual modo, un supersistema es un superconjunto que contiene al sistema en referencia y se constituye como un sistema más general.

- **Clasificación de los Sistemas**

  Los sistemas se clasifican de varias formas de acuerdo a determinadas categorías. Se considera la siguiente clasificación:

  - ✓ *Simples y Complejos*. Los sistemas son simples, cuando su representación y tratamiento no es complicado, tal como el organigrama de una universidad o empresa. Así mismo, son complejos, cuando su modelado y simulación involucra cierto grado de dificultad.
  - ✓ *Estáticos y Dinámicos*. Los sistemas son estáticos cuando su estructura es conocida y su comportamiento es invariable en un determinado periodo de tiempo. Del mismo modo son dinámicos,

cuando admite cierto grado de variación en el tiempo. Tales sistemas pueden ser físicos, biológicos, sociológicos, etc.

- ✓ *Determinísticos y Probabilísticos*. Los sistemas son determinísticos, cuando los parámetros se especifican de modo preciso. Así mismo, son probabilísticos o estocásticos, cuando las condiciones de entrada generan incertidumbre y adoptan naturaleza aleatoria.
- ✓ *Discretos y Continuos*. Los sistemas son de tiempo discreto, cuando sus parámetros pertenecen a un conjunto numerable. Del mismo modo, son de tiempo continuo, cuando sus variables varían de modo continuo en el tiempo.

- **Propiedades de los Sistemas**

Entre las propiedades que caracterizan a los sistemas se tiene:

- ✓ *Entropía*, se refiere al colapso y degeneración progresiva del sistema. La entropía es directamente proporcional al tiempo e inversamente proporcional a la información.
- ✓ *Sinergia*, asociada a las interacciones entre los partes o componentes, cuyo funcionamiento contribuye al logro del fin u objetivo del sistema.
- ✓ *Homeostasis*, típico de los sistemas adaptables, es la tendencia a conservar el equilibrio y la estabilidad a través de sus procesos que responden a las variaciones del entorno del sistema.
- ✓ *Equifinalidad*, consiste en alcanzar un objetivo o estado final a partir de diferentes rutas o distintas condiciones iniciales
- ✓ *Emergencia*, señala que, si un sistema funciona como un todo, entonces tiene propiedades distintas a las de sus partes que lo componen y que emergen de él cuándo está en acción.
- ✓ *Control*, se sustenta en los componentes del sistema sus interrelaciones en un instante determinado.

- **Sistemas Dinámicos Complejos**

Los fenómenos del universo y del mundo real pueden ser explicados y analizados bajo la perspectiva de los sistemas dinámicos complejos. El modelado y simulación permite su representación y el análisis de su comportamiento, adoptando ciertos niveles de complejidad y dificultad de acuerdo a la naturaleza de los variables o parámetros que lo describen.

Numerosas percepciones y puntos de vista establecidos por diferentes autores describen a los sistemas complejos.

Ladyman et al. (2012) establece que, a menudo los sistemas complejos se caracterizan en términos de la teoría de la información, basado en la idea de que el orden de un sistema complejo puede entenderse como mantenido por el procesamiento interno de la información.

Por otro lado, Boccara (2010), establece como ejemplo de sistema complejo a una colonia de hormigas, conformada por grandes poblaciones de agentes conectados que adoptan una dinámica global emergente que es el resultado del comportamiento cooperativo e interacciones locales de sus partes en lugar de ser impuesta por un controlador central.

Los sistemas complejos están asociadas a propiedades específicas que determinaran su comportamiento y consecuentemente van a determinar su posterior formulación analítica o simulación computacional.

Ladyman et al (2012) establecen varias propiedades asociadas a un sistema complejo, tales como, no-linealidad, retroalimentación, orden espontáneo, robustez y falta de control central, emergencia, organización jerárquica y numerosidad.

- **Caos, complexidad y entropía**

La teoría de la complejidad abarca numerosas áreas del conocimiento entre las cuales se pueden citar a la dinámica no lineal, los sistemas dinámicos, la teoría del caos entre otras importantes líneas del conocimiento.

Un sistema complejo, tal como un autómata celular o un fractal geométrico, puede ser descrito o modelado a partir de reglas simples.

Baranger (2001) señala que el siglo XXI ha comenzado con una gran explosión. Un aspecto para los científicos, es la revolución de la complejidad, que en todas las disciplinas científicas se está cambiando el enfoque de la investigación.

En la referida publicación, el autor especifica al espacio y el tiempo como las dimensiones más importantes, y presenta al triangulo de Sierpinski como un ejemplo de un objeto caótico en el espacio denominado fractal.

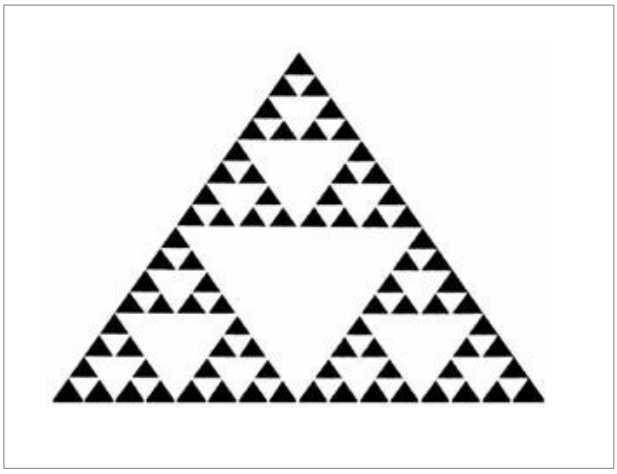

Figura 2.5 Triángulo de Sierpinski. Fuente: Baranger (2001).

2.3.1.2 LOS MODELOS COMO REPRESENTACIONES DE SISTEMAS

De modo general, los modelos son representaciones simplificadas de un sistema real, que hacen posible su aproximación a la realidad, tal como se ilustra en la siguiente figura.

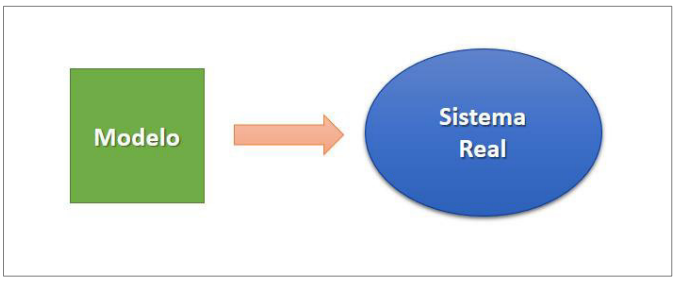

Figura 2.6 Modelo como representación de un Sistema. Fuente: Elaboración Propia.

Un modelo es la representación abstracta, análoga o idealizada de un determinado sistema real.

Winston (2005), señala que "*en el enfoque científico de toma de decisiones, se requiere el uso de uno o más modelos matemáticos*", y los define como "*representaciones matemáticas de situaciones reales que se podrían usar para tomar mejores decisiones o bien simplemente para interpretar mejor la situación real*".

El diseño del modelo representativo de un sistema, debe tomar en consideración los elementos que caracterizan al sistema, tales como su composición o estructura, componentes, parámetros o variables de estado.

Mediante el establecimiento de relaciones matemáticas o lógicas entre sus componentes, el modelo busca aproximarse al sistema real.

A través de los modelos se busca conocer y entender los fenómenos o procesos asociados, mediante la interacción de los componentes de un sistema.

- **CLASES DE MODELOS**

Existen diferentes modalidades para clasificar a los modelos. Prawda (2004), establece que "*en la Investigación de Operaciones existen tres clases de modelos: icónicos, analógicos y simbólicos*".

Dado que los modelos describen un sistema real, su clasificación está asociada al sistema que representan. Por lo tanto, se tiene:

- ✓ Modelos de Simulación Tiempo Discreto y Continuo.
- ✓ Modelos de Simulación Estática y Dinámica
- ✓ Modelos de Simulación Determinista y Estocástica

- **CRITERIOS PARA DISEÑAR UN MODELO**

El diseño y construcción de un modelo asociado a un problema involucra establecer los siguientes criterios:

- ✓ Proceso de abstracción consistente básicamente en:
  - Dado el problema, identificar los elementos que tienen un rol más determinante y descartar aquellos que no tienen incidencia en el mismo.
  - Determinar las relaciones o reglas de interacción entre los componentes del sistema.
- ✓ Diseño y construcción del modelo, considerando los parámetros que determinarán su estado.
- ✓ Resolución del Problema, mediante la manipulación de sus parámetros o variables de entrada:
  - Si el modelo no admite solución, entonces no es factible su implementación y simulación computacional
  - Si el modelo admite solución, entonces es conveniente efectuar su implementación y simulación computacional.
- ✓ Aplicación del modelo a la realidad, cuando la resolución del problema es posible.

2.3.1.3 LA SIMULACIÓN COMO APROXIMACIÓN A LA REALIDAD

En ciencias e ingeniería, la simulación permite imitar el funcionamiento y operación de los sistemas en el mundo real, cuando evoluciona dinámicamente en el tiempo. La simulación es una poderosa técnica utilizada para analizar y evaluar el comportamiento de un sistema.

La representación de un sistema a través de un modelo matemático, tiene por finalidad encontrar soluciones analíticas al problema, realizar predicciones al comportamiento del sistema a través de un conjunto de parámetros y condiciones iniciales.

Si el sistema adopta una naturaleza dinámica y compleja, su modelado matemático usando métodos y procedimientos analíticos es extraordinariamente complicada dificultando en gran medida su implementación debido al elevado costo computacional. En este caso, la aplicación del enfoque heurístico en el tratamiento y simulación del problema representa una gran alternativa adoptando un rol de preponderancia en la solución.

- **CONCEPTOS DE SIMULACIÓN**

Dada su importancia, existen múltiples percepciones acerca de la simulación que permiten su definición:

- ✓ Taha (2012) resalta al experimento de Montecarlo como “un esquema de modelado para estimar parámetros estocásticos o determinísticos con base a un muestreo aleatorio”.
- ✓ De acuerdo a Hillier et al. (2010), señalan a la simulación como “una técnica que involucra el uso de una computadora para imitar la operación de un proceso o sistema completo”.
- ✓ Por otro lado, Ortiz M. & Olivares P. (2018), definen a la simulación como una técnica de la investigación de operaciones que consiste en describir el comportamiento de los procesos o actividades relevantes de un sistema mediante modelos o prototipos lógico matemáticos.

- **EL PROCESO DE SIMULACIÓN**

El proceso de simulación consiste en un conjunto organizado de fases con el propósito de realizar experimentos muestrales con el modelo. Luego, los resultados a obtener consisten en las observaciones en el conjunto de salida.

Ortiz M. & Olivares P. (2018), establecen las siguientes fases de un proyecto de simulación:

- ✓ Definición del sistema, problema, objetivos, requerimientos, etc.
- ✓ Diseño del modelo conceptual
- ✓ Obtención y análisis de datos para el modelo
- ✓ Generación del modelo de simulación preliminar
- ✓ Verificación del modelo
- ✓ Validación del modelo
- ✓ Generación del modelo final y de ejecución
- ✓ Documentación del modelo y presentación de los resultados
- ✓ Análisis de sensibilidad y conclusiones

No obstante, el proceso de simulación adoptado en la presente investigación, tiene relación directa con la metodología propuesta en la sección 3.2 conducente a la resolución de un problema bajo el enfoque de la investigación de operaciones y teoría de sistemas.

- **CONDICIONES PARA USAR SIMULACIÓN**

La simulación tiene por finalidad la puesta del modelo a la realidad, esto es, debe garantizar la implantación y operación del sistema real. En consecuencia, su aplicación es conveniente cuando se presentan algunos de los siguientes escenarios:

- ✓ Los métodos analíticos o numéricos son demasiado complejos y difíciles de formular.

- ✓ No se ha completado la formulación matemática del modelo correspondiente al problema.
- ✓ Las soluciones matemáticas existen y son posibles, pero adoptan mucho rigor y complejidad en su implementación.
- ✓ Los costos de testing y pruebas en el escenario real son elevados

- **RIESGOS EN LA SIMULACIÓN**

Se han identificado algunos factores que generan riesgos en el proceso de simulación.

- ✓ Cuando el modelo matemático es demasiado complejo, su implementación computacional, tal como el diseño de algoritmos y codificación de programas se hace complicada. El uso de paquetes de software especializados representa en la mayoría de los casos una solución a este problema.
- ✓ Cuando el procesamiento tiene un elevado costo computacional, el modelo ya no es factible. No obstante, haciendo uso de procedimientos heurísticos es posible minimizar la complejidad computacional.

2.3.2 INTRODUCCIÓN A LOS AUTÓMATAS CELULARES

Las siguientes secciones exponen las nociones elementales de los autómatas celulares de modo que permitan su entendimiento.

2.3.2.1 NOCIONES DE AUTÓMATAS CELULARES

- **Máquina de Estados**

Un autómata denominado también máquina de transición de estados, es un modelo matemático correspondiente a una máquina abstracta definida de modo formal sobre un conjunto de estados y una serie de reglas las transiciones entre estados que establecen el comportamiento dinámico del autómata.

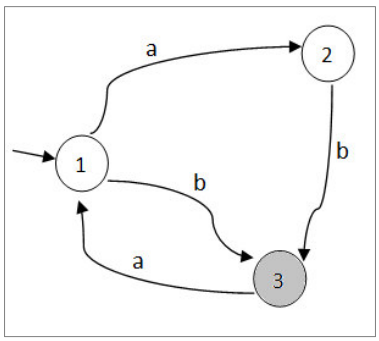

Figura 2.7 Autómata que representa una Maquina de Estados.
Fuente: Elaboración Propia.

- **Autómata Celular**

Un autómata celular es un modelo matemático y computacional que representa a un sistema dinámico complejo que adopta naturaleza discreta en sus parámetros o variables de estado.

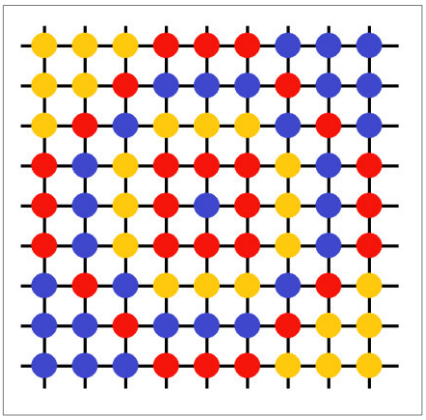

Figura 2.8 Ejemplo de Autómata Celular. Fuente: Elaboración Propia.

Los componentes de un autómata celular se denominan celdas, nodos o células que evolucionan en el tiempo a partir de la aplicación de un conjunto de interacciones locales o reglas simples de transición o funciones de evolución.

- **Estructura de datos**

Los autómatas celulares se implementan sobre estructuras de datos n-dimensionales donde $n \epsilon \mathbb{Z}^+$.

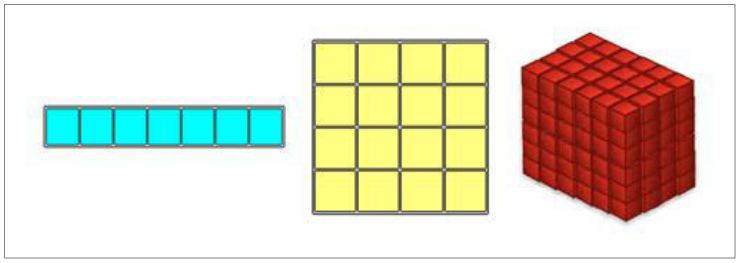

Figura 2.9 Estructuras n-dimensionales para un autómata celular. Fuente: Elaboración Propia.

- **Vecindad o entorno**

La vecindad o entorno de una célula de autómata celular es el conjunto de células que son adyacentes o contiguas a ella.

Los estados correspondientes al entorno de una célula, determinarán el estado siguiente de la misma.

La siguiente figura describe los entornos celulares más notables que corresponden a un autómata celular plano o bidimensional:

- ✓ *Vecindad de Von Newman*, conformada por 4 células vecinas que se encuentran a cada lado de la célula en referencia
- ✓ *Vecindad de Moore*, constituida por 6 células que contienen a la célula referenciada.

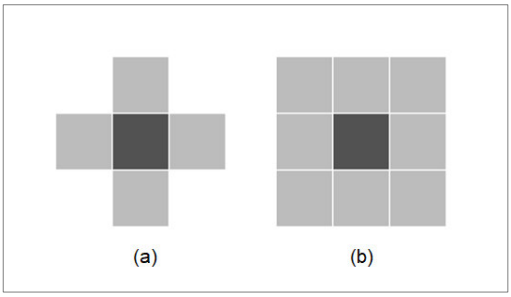

Figura 2.10 (a) Tipos de Vecindades: (a) Von Neumann y (b) Moore. Fuente: Elaboración Propia.

- **Conjunto de Estados**

  Las células de un autómata celular adoptan un conjunto finito de estados que determinaran el estado del sistema en la próxima generación. El estado siguiente de una célula determinada del autómata, depende de su estado actual y el de sus vecinos.

  En una generación determinada, el estado de una célula depende de manera única y exclusiva de su propio estado y de los estados de las células vecinas en la generación anterior. Esto les permite tener la capacidad de representar comportamientos complejos de diferentes fenómenos del mundo real.

- **Reglas de Transición**

  Los autómatas celulares adoptan un comportamiento complejo a partir de ciertas funciones de transición o reglas de evolución temporal aplicada a cada célula, que determina los estados siguientes.

  Esta función discreta y temporal, se aplica a todas las células del autómata y determina el siguiente estado de una célula basada en los estados actuales de sí misma y de sus vecinas.

2.3.2.2 AUTÓMATA CELULAR SIMPLE O ELEMENTAL

Los autómatas celulares están conformados por una red de células conectadas de modo local.

Cada célula representa un autómata simple y genera una salida a partir de varias entradas, modificando su parámetro de estado durante el proceso, de acuerdo a la función de transición.

Un autómata celular simple o elemental tiene las siguientes características:

- ✓ La estructura de datos correspondiente a su implementación es un arreglo unidimensional o vector.
- ✓ Cada celda adopta únicamente dos estados posibles: 0: *Inactivo* y 1: *Activo* las cuales se asociarán respectivamente a los colores blanco y negro.

- ✓ El proceso de evolución del autómata celular se inicia a partir de un estado inicial (T=0) en la cual se activa una de las células.

Se presenta como modelo de ejemplo, el autómata celular elemental que satisface las siguientes reglas:

- ✓ Si una célula está *Activa* entonces en el siguiente paso se mantiene *Activa*.
- ✓ Si una célula está *Activa* entonces en el siguiente paso las células vecinas se encuentran *Activas*.

La consecuencia de esta definición, muestra la evolución del autómata celular en 4 pasos, etapas o generaciones visualizado en la siguiente figura.

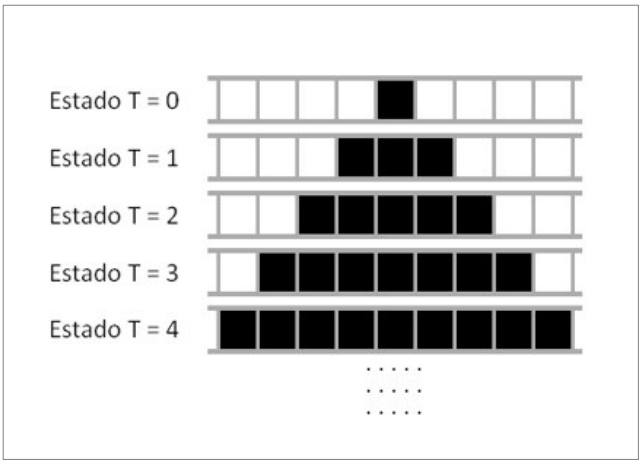

Figura 2.11 Ejemplo de Autómata Celular Elemental. Fuente: Elaboración Propia.

2.3.2.3 APLICACIONES DE LOS AUTÓMATAS CELULARES

La aplicación que hace referencia la presente investigación corresponde a la simulación de un sistema criptográfico haciendo uso de un modelo de autómata celular elemental y circular.

No obstante, son múltiples los contextos de aplicación de los autómatas celulares como modelos representativos de sistemas dinámicos complejos. Se describen las principales aplicaciones de autómatas celulares a problemas de teoría de números, procesamiento de imágenes, ciencia de los materiales, tráfico vehicular, prevención y control de incendios, economía y finanzas, entre otros.

- **TEORÍA DE NÚMEROS**

Muthuswamy et al. (2008), indica el uso de un autómata celular para realizar la factorización entera de los números de Fermat, que tienen aplicaciones en procesamiento de señales y criptografía y tienen la forma

$$\mathrm{F_k} = 2^{2^{\mathrm{k}}} + 1, \mathrm{k} = 0,1,2, \ldots$$

La factorización de los números de Fermat es un problema muy complicado dado que no existe un algoritmo general que realice el procesamiento.

El autómata celular asociado al problema, adopta células con estado binario 0 o 1, interactúa únicamente con sus 2 vecinos a la izquierda y derecha, y adopta como función de transición la regla 46. Contiene $\mathrm{L} = \mathrm{n} + 1$ células con condiciones de frontera periodicas.

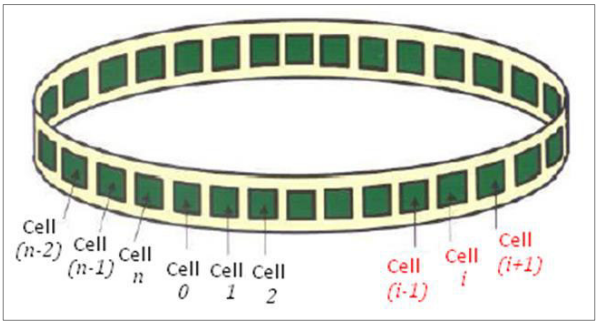

Figura 2.12 Estructura del autómata celular. Fuente: Muthuswamy et al. (2008).

- **DETECCIÓN DE BORDES EN PROCESAMIENTO DE IMÁGENES**

En el ámbito del procesamiento de imágenes y visión artificial, la segmentación de imágenes representa una de las etapas más importantes.

De acuerdo a lo indicado por Parra et al. (2017), la cooperación entre las células del autómata celular con sus vecinas, así como las variaciones en los aspectos de color basados en componentes RGB (Red Green Blue) de la imagen con respecto a un determinado umbral de reconocimiento, permiten realizar la detección de bordes en las imágenes.

La segmentación de imágenes consiste en minimizar la cantidad de datos que corresponden a la imagen de modo que permita reducir la complejidad y el costo computacional de procesamiento.

La detección de los bordes de una imagen permite obtener información sobre ella, tal como su composición. De este modo, se puede reducir de manera significativa la cantidad de datos utilizada para su representación, realizar el filtrado de información no necesaria conservando sus principales características. Esto va a permitir acelerar el procesamiento.

La siguiente figura ilustra el resultado del proceso de detección de bordes.

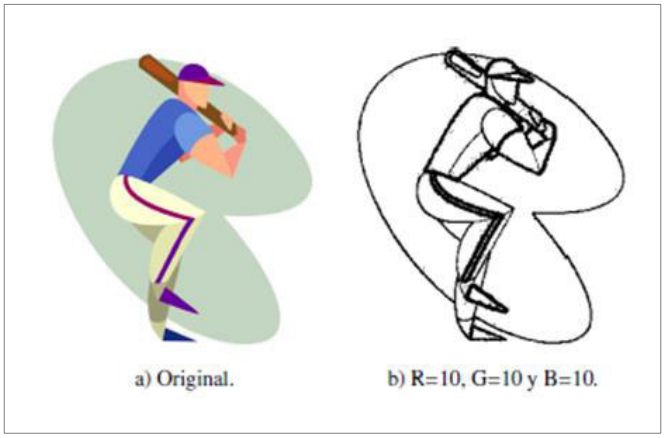

Figura 2.13 Ejemplo de Detección de Bordes. Fuente: Parra et al. (2017).

- **FENÓMENOS EN CIENCIA DE LOS MATERIALES**

Otra importante utilización de los autómatas celulares es establecida por Kohutek et al. (2009), quienes describen la simulación de fenómenos en el campo de ciencia de los materiales, aplicado a los procesos de recristalización y de crecimiento de grano.

Cuando un material metálico es sometido a temperaturas bajas respecto a su punto de fusión, sufren deformaciones que modifican sus microestructuras y propiedades. Sin embargo, las propiedades y estructura original pueden restablecerse mediante la realización de procedimientos térmicos apropiados a temperaturas elevadas, debido a la ocurrencia de fenómenos de recuperación y recristalización, que pueden ser analizadas por el crecimiento de granos.

Los autores han considerado, para la simulación de los procesos de recristalización y crecimiento de granos, un autómata celular representado por matriz cuadrada constituida por 40000 células.

La siguiente figura ilustra la evolución en el tiempo de la microestructura simulada durante la secuencia de los procesos de recristalización y crecimiento de grano.

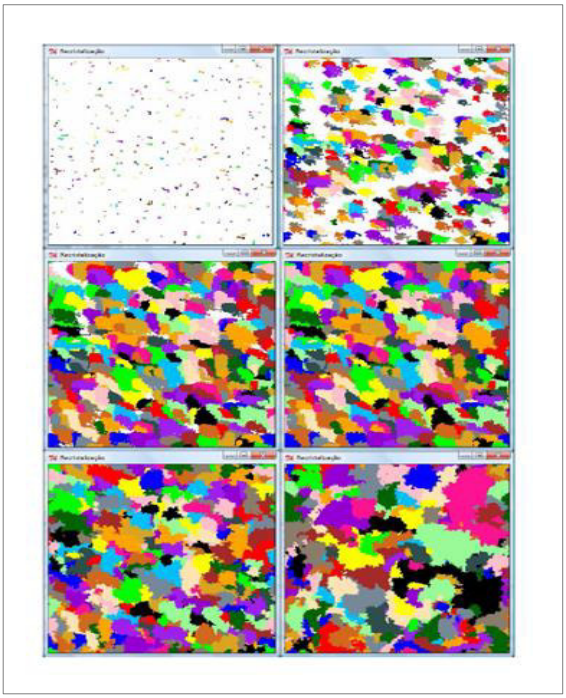

Figura 2.14 Simulación de la recristalización y del crecimiento de grano. Fuente: Kohutek et al. (2009).

- **Predicción de Incendios Forestales**

En su publicación, Karafyllidis et al. (1996), realizan la predicción de la propagación del fuego y su evolución en el tiempo en bosques, bajo diferentes factores tales como la topografía del terreno, la velocidad y dirección del viento, entre otras condiciones climáticas. Estos factores determinan la tasa de propagación del fuego, es decir, la velocidad del frente de fuego que es la frontera divisoria entre las regiones quemadas y no quemadas en un bosque.

Debido a su naturaleza discreta y fácil implementación computacional, los autómatas celulares son apropiados modelar y simular la propagación de incendios forestales, puesto que permiten analizar su comportamiento, en cada punto del bosque.

El bosque se asocia a un autómata celular y representa a una matriz de celdas cuadradas idénticas con una longitud lateral a. Cada celda del bosque corresponde a una célula del autómata. Los anchos de los dos lados del autómata se consideran iguales, es decir

$$\mathrm{w}_1 = \mathrm{w}_2$$

Luego, cada célula (i, j) del autómata posee 8 células vecinas (4 adyacentes y 4 diagonales), según se muestra en la figura

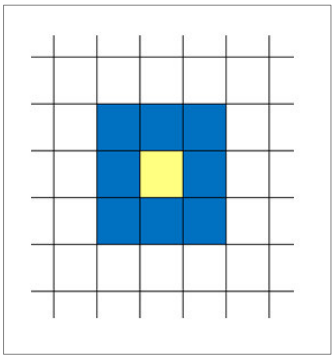

Figura 2.15 Vecindad con ocho celdas para la célula (i, j).Fuente: Karafyllidis et al. (1996).

El estado local $\mathrm{S}_{\mathrm{i,j}}^{\mathrm{t}}$de la célula (i, j) en el tiempo t es la tasa entre el área de la celda quemadaA_b y el área total de la celda$\mathrm{A_t}$:

$$\mathrm{S}_{\mathrm{i,j}}^{\mathrm{t}} = \frac{\mathrm{A_b}}{\mathrm{A_t}}$$

- **La Predicción del Movimiento de Precios de Bienes Raíces**

Una importante aplicación de los autómatas celulares es descrita por Cavada (2007), que describe su utilización para hacer estimaciones del movimiento del costo promedio por metro cuadrado de departamentos en algunas comunas.

Para tal efecto, se utilizó información mensual del costo promedio de oferta del metro cuadrado de departamentos de ciertas comunas, durante un periodo aproximado de 3 años.

La oferta y la demanda en el mercado de bienes raíces, está condicionado por las decisiones de los inversionistas, quienes disponen de información actualizada y son considerados como sistemas complejos adaptativos.

Cada comuna representa una célula del autómata, cuya vecindad está determinada por el conjunto de las otras comunas, esto es, todas son vecinas entre sí. En consecuencia, cada comuna tiene influencia sobre sus demás vecinas, a través de las correlaciones existentes.

La función de transición o influencia obtiene la predicción del estado siguiente para cada comuna tomando como base los estados de cada célula. La influencia de las células C_i y C_j está dada por:

$$I_{ij} = e^{-|1-\rho_{ij}|}.S_j$$

donde

ρ_{ij} : correlación entre C_i y C_j

S_j : estado actual de C_j

Las predicciones obtenidas se contrastan con las variaciones reales y se calcula el porcentaje de predicción de signo (PPS) para cada comuna. PPS es el porcentaje de aciertos en la dirección del movimiento de los precios que suceden durante el periodo. Se adoptó la tasa del PPS igual o superior al 60%.

2.3.3 INTRODUCCIÓN A LA CRIPTOGRAFÍA DE LA INFORMACIÓN

Desde épocas remotas, la codificación y cifrado de la información constituye uno de los problemas más significativos que involucra a personas, organizaciones y sociedad en general. La confidencialidad y reserva de la información es el factor crítico de éxito que garantiza la protección y seguridad de la información.

El concepto de protección y seguridad de la información tiene un alcance demasiado amplio que va mucho más allá del objetivo de la presente investigación e involucra tecnologías, protocolos y normas internacionales entre otros aspectos, generando muchas disciplinas que representan importantes campos de especialización académica y profesional.

La criptografía hace referencia al desarrollo e implementación de un sistema criptográfico mediante el uso y aplicación de métodos matemáticos, técnicas de optimización y algoritmos computacionales para codificar y ocultar la información.

2.3.3.1 ORGANIZACIÓN Y ESTRUCTURA DE LA INFORMACIÓN

Las nociones básicas respecto al significado de datos e información son requeridas y son tratados de forma elemental en esta sección, de modo que permitan su comprensión y entendimiento.

El marco conceptual asociado a los términos conocimiento e inteligencia no es parte de la presente investigación, pero es necesario proporcionar una breve descripción.

- **DATOS E INFORMACIÓN**

Usualmente los términos dato e información están estrechamente relacionados y suelen ser expresados como sinónimos propiciando su uso de manera indistinta y generando ambigüedad en su utilización. Sin embargo, en el contexto de la presente investigación cada uno adoptara su propio rol.

Los datos se encuentran representados como secuencias binarias de bytes[11] y están localizados en repositorios o volúmenes magnéticos o electrónicos de almacenamiento tales como discos, cintas u otras memorias externas, en equipos de cómputo de la organización constituyendo la base de datos organizacional.

La información es el resultado del procesamiento y transformación de los datos tal como tal como una planilla, reporte o consulta determinada, y es obtenida al agregarle ciertos aspectos tales como cálculo, depuración, clasificación, validación y consistencia entre otros importantes criterios de cantidad y calidad.

Conceptos importantes derivados de los datos y la información lo constituyen el conocimiento y la inteligencia.

- **CONOCIMIENTO E INTELIGENCIA**

El conocimiento es el resultado de un complejo proceso de transformación de la información y adopta una naturaleza dinámica, que permite la toma de decisiones.

En las organizaciones el conocimiento se encuentra en las normas directivas, en las prácticas rutinarias del especialista, permitiendo su utilidad para la acción.

Por otro lado, la inteligencia representa el fin que las organizaciones están en permanente búsqueda. Se sustenta en el aprendizaje continuo y adaptativo.

En su publicación, Senge (2005) establece que una organización inteligente aprende y continuamente expande su capacidad para crear su futuro y señala también que el aprendizaje adaptativo generalmente denominado aprendizaje para la supervivencia, es necesario e importante.

11 Un byte u octeto es la unidad de almacenamiento de datos. Está constituida por ocho dígitos binarios o bits.

Resumiendo lo descrito en las secciones anteriores, los datos permiten obtener información que se transforma en conocimiento y permite la generación de inteligencia, según se ilustra en la figura.

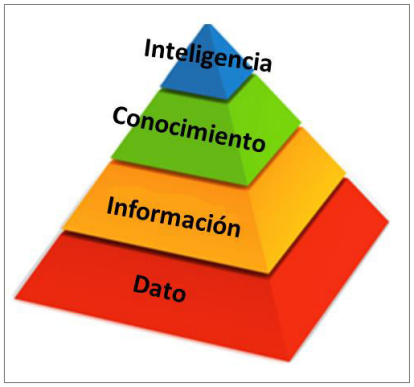

Figura 2.16 Representación Jerárquica de dato, información, conocimiento e inteligencia. Fuente: Elaboración Propia.

- **ORGANIZACIÓN Y REPRESENTACIÓN**

Los datos se representan como una secuencia de códigos binarios contenidos en los medios de almacenamiento.

La unidad básica de almacenamiento de datos es el byte constituido por ocho dígitos binarios.

La siguiente figura proporciona un ejemplo de byte que contienen la secuencia binaria *10101010*

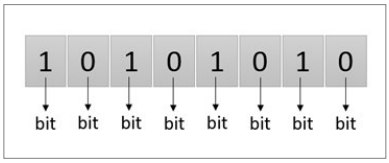

Figura 2.17 Byte u Octeto. Fuente: Elaboración Propia.

- **CLASIFICACIÓN Y TIPOLOGÍA DE LA INFORMACIÓN**

Tomando como referencia los formatos de almacenamiento de los datos, la información puede ser:

- Texto
- Imagen
- Audio
- Video
- Otros Formatos: Documento, Hoja de Cálculo, Presentaciones, etc.

2.3.3.2 CRIPTOGRAFÍA PARA EL CIFRADO DE LA INFORMACIÓN

La cristología (del griego *criptos*=oculto y *logos*=tratado) es la ciencia que estudia los criptosistemas permitiendo la codificación y decodificación de mensajes.

Tilborg (2000), señala que la criptología es el estudio de criptosistemas y puede ser subdividida en dos disciplinas. La *criptografía***,** que se refiere al diseño de criptosistemas, y el *criptoanálisis* que estudia la ruptura de tales criptosistemas.

La *criptografía* está orientado a la codificación y protección de la información y representa la base de la presente investigación. Se hará una breve descripción teórica en la siguiente sección.

El *criptoanálisis* permite la decodificación y recuperación de la información. Es el sentido opuesto de la criptografía y consiste en el estudio de métodos y procedimientos conducentes a la obtención de las claves y técnicas de los algoritmos asociados al criptosistema que permitan descifrar la información cifrada.

El criptoanálisis, no forma parte ni constituye tema de estudio de la presente investigación.

- **Fundamentos de Criptografía**

Ferguson et al. (2010), indican que "*la criptografía es el arte y la ciencia del cifrado*".

La criptografía tiene como campo de estudio el desarrollo de procedimientos, técnicas y algoritmos para el cifrado o codificación de la información de modo que hagan imposible su comprensión y entendimiento, mediante la construcción de sofisticados y complejos sistemas criptográficos denominados criptosistemas.

La criptografía es una importante área de estudio multidisciplinar con especial aplicación a la seguridad y protección de la información.

- **Clases de Algoritmos Criptográficos**

Existen varias modalidades o tipos de sistemas criptográficos o criptosistemas, de acuerdo a las técnicas o métodos de cifrado, tales como: simétricos, asimétricos y de resumen de mensajes (funciones de dispersión).

La *criptografía simétrica*, requiere el uso de una clave simétrica la cual es usada para realizar los procesos de codificación y decodificación de la información. Tanto el emisor como el receptor deben compartir la misma clave simétrica en sus procesos. Este tipo de criptografía será citado en la siguiente sección.

La *criptografía asimétrica*, denominada también criptografía de *clave pública*, utiliza una clave pública y otra clave privada para realizar el cifrado y descifrado de la información. Se diseñó como una solución al problema del intercambio de claves típico de los criptosistemas de clave simétrica. Este tipo de criptografía tiene un elevado costo de diseño e implementación. Ejemplos y aplicaciones de los criptosistemas asimétricos lo constituyen el cifrado RSA, los certificados y firmas digitales.

La criptografía asimétrica no es tema de estudio ni forma parte del contexto de la presente investigación.

- **Modelo de Cifrado Simétrico**

En la criptografía simétrica, denominada también de clave secreta, tanto el emisor como el receptor deben de ponerse de acuerdo sobre la clave a utilizar.

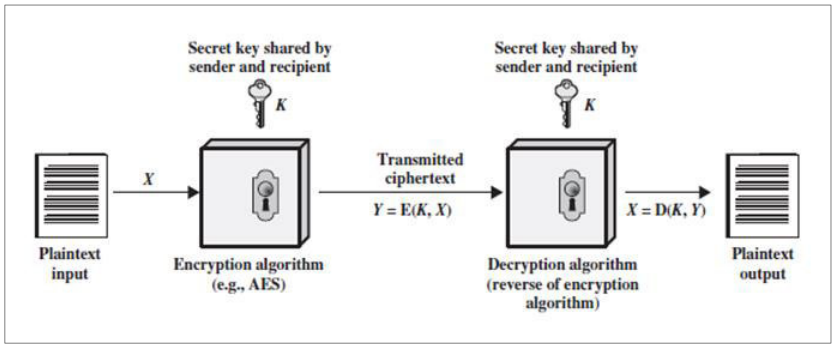

Figura 2.18 Modelo Simplificado de Cifrado Simétrico. Fuente: Stallings (2014).

Como ejemplos de cifrado simétrico se tiene a DES (*Data Encryption Standard*) y AES (*Advanced Encryption Standard*).

Según Stallings (2014), un esquema de cifrado simétrico tiene cinco componentes:

- *Texto plano o sin formato*, representa la información de entrada para el algoritmo.
- *Algoritmo de encriptación*, el algoritmo de cifrado realiza transformaciones sobre el texto de entrada.
- *Clave secreta*, es un valor fijo e independiente de la entrada y del algoritmo de cifrado, que genera una salida específica para el valor que adopta en ese momento. El proceso de cifrado depende totalmente de la clave utilizada.
- *Texto cifrado*, el texto cifrado o encriptado es el resultado del proceso aplicado sobre el texto plano de entrada y de la clave simétrica adoptada. Valores diferentes de la clave generan textos cifrados diferentes.
- *Algoritmo de descifrado*, es un proceso reverso al algoritmo de encriptación. La entrada es el texto cifrado y la clave secreta y la salida debe ser equivalente al texto original.

La referencia establece el siguiente modelo de criptosistema simétrico.

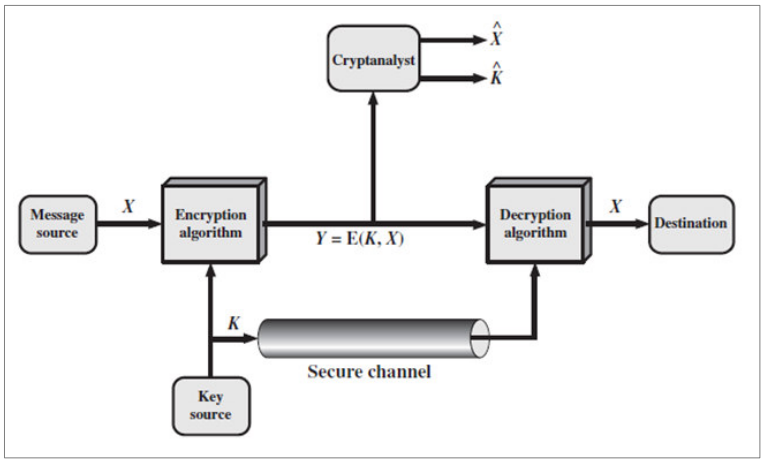

Figura 2.19 Modelo de Criptosistema Simétrico. Fuente: Stallings (2014).

donde

$$Y = E(K, X)$$

indica que el texto cifrado Y se obtiene como resultado de la aplicación del algoritmo de cifrado E utilizando como información de entrada a X y usando la clave K

El receptor que posee la clave K puede invertir el proceso mediante

$$X = D(K, Y)$$

en la cual X es el texto original resultado de la aplicación del algoritmo de desencriptado D usando como parámetros el texto cifrado X y la clave K

Las notaciones $\widehat{K}$ y $\widehat{X}$ representan estimadores o valores supuestos que posee cualquier criptoanalista que desea obtener respectivamente la clave K, el texto original X, o ambos.

Capítulo 3

METODOLOGÍA

En este capítulo se realizará una exposición detallada, de los elementos conducentes al desarrollo de la investigación, consistente en realizar la simulación de un modelo de autómata celular para el tratamiento del problema del cifrado simétrico de la información.

3.1 DISEÑO DE LA INVESTIGACIÓN

Es importante indicar que la metodología utilizada es propia y adopta una perspectiva de la investigación de operaciones y sistemas, según se describe en la sección 3.2.

En esta sección se exponen los elementos básicos que describen de manera general el presente trabajo de investigación.

3.1.1 TIPO DE INVESTIGACION

La investigación adopta diferentes categorías o tipos, entre las cuales vale citar:

- *Descriptiva*, porque describe la organización, elementos y comportamiento de un problema, más allá de un simple acercamiento al mismo.
- *No Experimental*, porque las variables asociadas al problema, descritas en la sección 1.3.4, no tienen manipulación directa, sino representan entradas y salidas del proceso de simulación del modelo.
- *Cuantitativa*, porque las variables de estudio utilizan información numérica, expresada en tamaño en bytes, tiempo de ejecución, etc.

3.1.2 COMPONENTES DE LA INVESTIGACION

Los siguientes componentes

- Unidad de análisis
- Población de estudio
- Muestra representativa
- Procedimientos de recolección de datos

describen la presente investigación y serán descritos al detalle en las fases

- Planteamiento del Problema
- Análisis del Sistema
- Diseño del Modelo

de la metodología propuesta indicada en la sección 3.2.

3.2 ENFOQUE DE LA INVESTIGACIÓN DE OPERACIONES Y TEORÍA DE SISTEMAS

Esta sección representa la metodología adoptada para el desarrollo de la investigación, y se sustenta en la investigación de operaciones y la teoría de sistemas, importantes áreas asociadas e interrelacionadas entre si y que tienen vinculación estrecha al problema de investigación.

Un problema específico que requiere solución óptima, puede ser asociado a un sistema, adoptando, por consiguiente, sus propiedades, comportamiento y consecuentemente puede ser modelado y simulado.

Las resoluciones de problemas basados en este enfoque permiten formular y simular el modelo representativo que describe al sistema.

Para resolver un problema específico basado en este enfoque, se plantea una secuencia jerárquica organizada en fases o etapas las cuales a su vez se descomponen funcionalmente en un conjunto de actividades o tareas determinadas.

Dado que la secuencia de fases representa la metodología propuesta, su tratamiento será realizada con detalle en las secciones posteriores.

La siguiente figura visualiza las todas fases que componen la resolución de un problema bajo el enfoque de la investigación de operaciones y teoría de sistemas y ha sido adoptada como metodología para el desarrollo de la investigación.

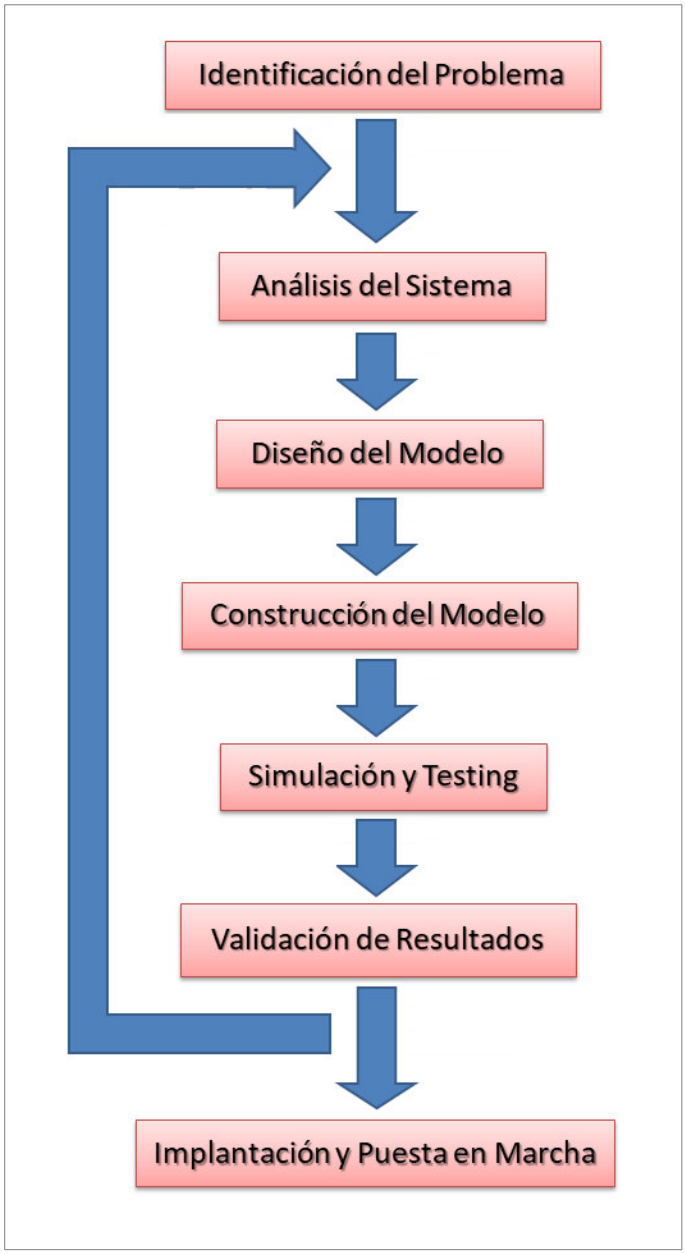

Figura 3.1 Fases de solución de un problema con enfoque de Investigación de Operaciones y Teoría de Sistemas. Fuente: Elaboración Propia.,

La secuencia de fases propuesta, incorpora un comportamiento iterativo para la solución de un problema.

Es posible volver a plantear, formular y diseñar la solución cuando no se han obtenido los resultados esperados después de la fase de simulación.

3.3 FASES DE LA INVESTIGACION

La metodología propuesta está organizada y estructurada a través de una secuencia ordenada de fases o etapas definidas, las cuales contienen los elementos necesarios y suficientes que van a permitir la descripción de la solución del problema.

A continuación, se describen las fases correspondientes a la investigación.

3.3.1 IDENTIFICACIÓN DEL PROBLEMA

Esta fase se encuentra asociada al problema de la investigación descrita en la sección 1.2 de este documento que plantea la simulación de un modelo de autómata celular para el tratamiento del problema del cifrado simétrico de la información.

El carácter de preliminar esta fase, es de vital importancia puesto que permite tener una idea clara y exacta del problema a resolver constituyendo una garantía para el desarrollo de las posteriores fases.

El problema de la investigación surge como una necesidad de respuesta a una situación específica la cual fue descrita en la sección *Planteamiento del Problema* de este documento.

Por tanto, se han considerado los siguientes criterios para realización:

- *DETERMINAR LA NECESIDAD O REQUERIMIENTO*

La necesidad es garantizar la confidencialidad y reserva de la información como recurso estratégico en la organización para su afianzamiento y la oportuna toma de decisiones.

- *ESPECIFICAR LAS LIMITACIONES DEL PROBLEMA*

El problema de la investigación está limitado únicamente a garantizar la confidencialidad y seguridad de la información, mediante criptografía simétrica y utilizando autómatas celulares.

- *DETERMINAR EL ÁMBITO ESPACIAL DEL PROBLEMA*

  Representa el escenario de su planteamiento y resolución. En este caso, lo constituyen las unidades de almacenamiento de la organización que aplicará esta investigación.

- *IDENTIFICAR LA SOLUCIÓN AL PROBLEMA*

  La solución al problema se sustenta a través del desarrollo e implementación de un criptosistema simétrico basado en un modelo de autómata celular circular reversible.

- *ESPECIFICAR EL ALCANCE DE LA SOLUCIÓN*

  La repercusión, influencia o ámbito de aplicación de la solución, es cualquier organización tal como una empresa o universidad con repositorios de información de cualquier tipo o formato de fichero en unidades externas de computadores.

- *DETERMINAR EL HORIZONTE TEMPORAL DE LA SOLUCIÓN*

  Representa el periodo del planteamiento y resolución del problema. En el caso de la presente investigación corresponde al periodo de desarrollo e implementación, la cual se ilustra en la sección *Cronograma de Actividades* de este documento.

- *ESPECIFICAR LAS RESTRICCIONES DE LA SOLUCIÓN*

  La restricción principal de la solución al problema es la no realización de la compresión de la información. Sin embargo, en la sección *Recomendaciones* se sugiere incorporar mecanismos de compactación de la información para futuras versiones de este trabajo.

- *IDENTIFICAR VARIABLES O PARÁMETROS*

El problema contiene un conjunto de variables o parámetros que lo tipifican y determinan su comportamiento, y fueron previamente descritas en la sección *1.3.4 Identificación de Variables*.

Entre las variables cabe mencionar:

- Tamaño de fichero inicial
- Tamaño de fichero final
- Longitud de la clave simétrica
- Longitud del autómata
- Indicador de equivalencia

La descripción y operacionalidad de las variables fueron previamente descritas en las secciones *1.3.4 Identificación de Variables* y *1.3.5 Operacionalidad de las Variables* respectivamente.

- *IDENTIFICAR MODELOS DE DECISIÓN Y OPTIMIZACIÓN*

El problema contiene una metodología basada en una secuencia organizada y estructurada de fases en la cual cada subproblema adopta su propio criterio.

El criptosistema contiene una serie de subproblemas basados en modelos de decisión y optimización.

- *DETERMINAR FUENTE Y TÉCNICA DE RECOLECCIÓN DE DATOS*

La fuente u origen de datos es establecido por la muestra representativa de la población.

La técnica de recolección de datos es por muestreo sobre una población representado por los ficheros contenidos en unidades de almacenamiento externo de la organización que serán procesados por el criptosistema.

Considerando la metodología propuesta y adoptada, los criterios establecidos serán expandidos más adelante en las fases posteriores.

3.3.2 ANÁLISIS DEL SISTEMA

Los problemas adoptan parámetros, fronteras y restricciones desarrollándose sobre un determinado ámbito espacial y temporal.

Desde una perspectiva de la investigación de operaciones y teoría de sistemas un problema puede ser asociado a un sistema. Luego, el problema de la simulación de un modelo de autómata celular para el tratamiento del problema del cifrado simétrico de la información, es asociado al desarrollo de un sistema criptográfico o criptosistema.

La fase de *Análisis del Sistema* incorpora importantes aspectos y especificaciones a considerar, los cuales son ampliados con más detalle en la siguiente fase *Diseño del Modelo*.

3.3.2.1 PLANTEAMIENTO DEL PROBLEMA

El problema identificado en la fase anterior reúne aspectos preliminares que serán fundamentales para su planteamiento como criptosistema y su posterior modelado y diseño.

La definición del problema a resolver tiene correspondencia al desarrollo del criptosistema que incorpora un conjunto de modelos de decisión y optimización.

Se consideran los siguientes puntos:

- *DEFINIR LA ESTRUCTURA DEL CRIPTOSISTEMA*

El criptosistema está conformado básicamente por los siguientes módulos o componentes.

- ✓ Determinación de la Población
- ✓ Obtención de la muestra.
- ✓ Generación de clave simétrica.
- ✓ Módulo de autómata celular
- ✓ Validación y consistencia

- *IDENTIFICAR MODELOS DE INVESTIGACIÓN DE OPERACIONES.*

El problema a resolver está conformado por un conjunto de subproblemas que están asociados a los módulos establecidos en la sección anterior.

Desde una perspectiva de la investigación de operaciones, los subproblemas están relacionados a determinados tipos de modelos los cuales son categorizados como de decisión y de optimización, y serán descritos con detalle en las secciones posteriores.

La siguiente tabla visualiza los subproblemas más representativos de la investigación, y su correspondiente asociación a una determinada categoría de modelo.

Tabla 3.1 Tipos de modelo asociado al subproblema. Fuente: Elaboración Propia.

Subproblema	Tipo de Modelo
Determinación de la Población	Decisión
Obtención de la muestra	Optimización
Generación de clave simétrica	Optimización
Módulo de autómata celular	Decisión
Validación y consistencia	Decisión

En consecuencia, el planteamiento del problema tal como fue descrito en la sección 1.2 de este documento, consiste fundamentalmente en realizar la simulación de un modelo de autómata celular para el tratamiento del problema del cifrado simétrico de la información.

3.3.2.2 EL PROBLEMA COMO SISTEMA CRIPTOGRÁFICO

Existe una correspondencia biunívoca entre el problema planteado y el criptosistema asociado, que puede ser establecido como:

> *El problema planteado se soluciona mediante el desarrollo de un criptosistema*,

y recíprocamente

> *El criptosistema desarrollado proporciona la solución al problema planteado*

La siguiente figura ilustra el criptosistema conformada por datos y procesos en continua interacción externa con otros sistemas.

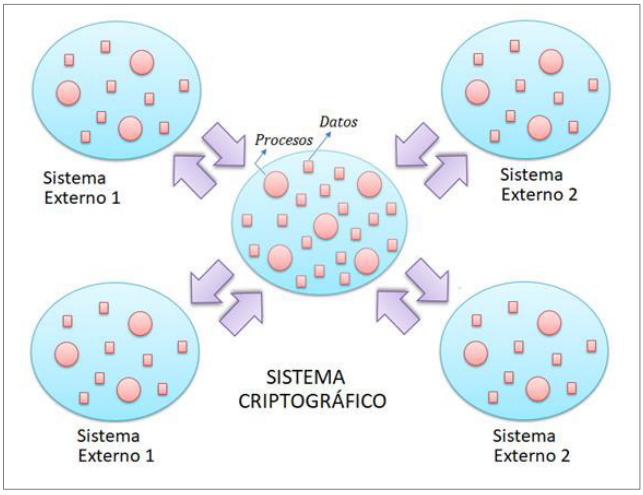

Figura 3.2 Sistema Criptográfico en interacción externa. Fuente: Elaboración Propia.

Los datos representan la información a encriptar y está representado por el conjunto de ficheros localizados en dispositivos de almacenamiento externo en las computadoras de una organización.

Los procesos representan las reglas de transformación expresada en código de programa que actúa sobre los datos de acuerdo a la configuración de los parámetros o variables de estado.

En consecuencia, el criptosistema adopta naturaleza compleja y comportamiento dinámico con parámetros que describen su comportamiento.

- *CONTEXTO E INSTANCIAMIENTO*

El problema del cifrado simétrico de la información es enfocado como un criptosistema que representa su contexto y está conformado por el conjunto de ficheros almacenados en unidades o dispositivos de almacenamiento externo en las computadoras de una organización.

El continuo flujo de entradas y salidas de ficheros, así como las continuas alteraciones en sus contenidos, es un típico ejemplo de un sistema dinámico complejo que evoluciona en el tiempo.

Sea el conjunto enumerable de pasos discretos en el tiempo

$$\{t_1,\ \ t_2,\ \ t_3,\ \dots\ , t_i,\ \dots\}$$

que representa una secuencia ordenada constituyendo una sucesión monótona creciente, dado que $\forall i,\ t_i < t_{i+1}$.

Puesto que el criptosistema adopta naturaleza dinámica, dada su variación en el tiempo, va a generar una sucesión infinita de configuraciones asociadas al problema

$$\{\Gamma_1,\ \ \Gamma_2,\ \ \Gamma_3,\ \dots\ , \Gamma_i,\ \dots\}$$

donde Γ_i representa la configuración o instancia del criptosistema en el instante o paso discreto del tiempo t_i.

De este modo, en un determinado instante t_i, el criptosistema va a adoptar una determinada configuración Γ_i, cuyos parámetros o variables van a describir su comportamiento, tales como cantidad de ficheros, frecuencia de alteración, tipología, tamaño en memoria, etc.

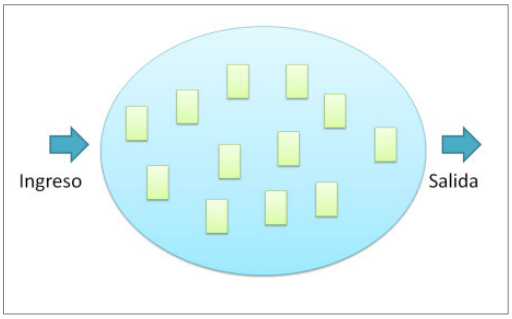

Figura 3.3 Instancia de un criptosistema. Fuente: Elaboración Propia.

Luego, la solución del problema del cifrado simétrico de la información se aplicará a una determina instancia del criptosistema.

- *VARIABLES O PARÁMETROS*

  El criptosistema tiene variables o parámetros que describen su comportamiento global y local, esto es, el comportamiento de todo el sistema de ficheros, de los subsistemas asociados y de cada ítem o elemento constituyente que representa el objeto de estudio.

  Entre los parámetros cabe señalar:

  - ✓ Espacio disponible en el almacenamiento.
  - ✓ Cantidad de ficheros para el procesamiento.
  - ✓ Distribución y frecuencia de tipos de información, tales como texto, imagen, audio, video u otros formatos específicos.

- *ÁMBITO DE APLICACIÓN*

  El contexto de aplicación lo constituye el conjunto de ficheros electrónicos de una organización, tal como una empresa o universidad. Este conjunto representa la población de estudio, a partir del cual se va a seleccionar una muestra consistente y representativa.

  El criptosistema puede manifestarse en varios escenarios, tales como información contenida en un computador personal, información en servidores organizacionales, información en la nube, etc.

  Luego, se tienen los siguientes aspectos a considerar:

  - ✓ Establecer el escenario o ámbito de aplicación.
  - ✓ Determinar la técnica de selección de ficheros.

- ✓ Especificar la cantidad de ficheros a procesar.
- ✓ Definir modelos de optimización en cada fase.
- ✓ Plantear estructuras para organizar y representar la información.
- ✓ Establecer de manera óptima la clave simétrica a utilizar.
- ✓ Establecer el modelo de autómata celular a usar.

3.3.2.3 FUENTES Y TÉCNICAS DE RECOLECCIÓN

En esta sección se plantean las especificaciones respecto a la fuente y técnica de recolección de datos conducentes a su posterior modelado y diseño.

Para tal efecto, se considera los siguientes criterios.

- *OBJETO DE ANÁLISIS*

Los ficheros de almacenamiento externo son componentes de una determinada instancia del criptosistema y como unidades básicas de información constituyen el objeto de análisis de la investigación.

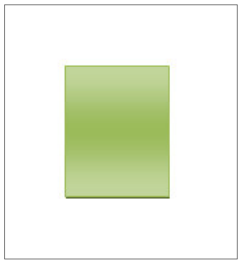

Figura 3.4 Fichero como unidad básica de información. Fuente: Elaboración Propia.

Las unidades básicas de información contienen un conjunto de atributos que permiten su descripción tales como tamaño en bytes, formato o tipo de contenido, etc.

- *POBLACIÓN DE ESTUDIO*

La población de estudio P, asociada a una determinada instancia Γ_i del criptosistema en el tiempo t_i del conjunto de pasos discretos $\{t_1,\ t_2,\ t_3,\ \ldots\ , t_i,\ \ldots\}$, será representada a través una muestra S.

El procesamiento de la muestra S y el consecuente análisis de los resultados van a permitir la generalización de la investigación, es decir, la aplicación del criptosistema a cualesquier de sus instancias, es decir, a cualquier población de ficheros P.

La población P va a ser determinada mediante un instrumento automatizado [12], el cual hace uso de los recursos del sistema operacional subyacente.

- *MUESTRA REPRESENTATIVA*

Las unidades básicas de información representados por los ficheros de almacenamiento externo constituyen la fuente de entrada del criptosistema.

El tamaño de la muestra S que ha sido adoptada en la investigación es $n = 1000$ ficheros.

Al igual que el punto anterior, la generación de la muestra será realizado por simulación a través de un instrumento automatizado a partir de la población P considerada.

12 El instrumento que genera la población, es un procedimiento que invoca al sistema operativo subyacente del equipo, obteniendo la totalidad de archivos contenidos.

3.3.2.4 CONSISTENCIA Y ROBUSTEZ DEL MODELO

La validación del modelo es establecida por su consistencia y robustez.

- CONSISTENCIA DEL MODELO

La consistencia del modelo va a ser sustentada mediante la validación de los resultados.

Para tal efecto, se requiere el diseño e implementación de un objeto o instrumento de validación, el cual determinará la equivalencia de los ficheros inicial y final.

El diseño e implementación del instrumento de validación debe adoptar las siguientes especificaciones:

- Debe estar asociada a la variable *Indicador de Equivalencia*, la cual representará un importante parámetro de decisión.
- Este parámetro decisión debe corresponder a un modelo determinístico adoptando un valor binario del tipo Si/No.

- ROBUSTEZ DEL MODELO

La robustez del modelo debe garantizar su funcionamiento y adaptabilidad a diferentes formatos o tipologías de información.

Así mismo, la adopción de una clave de un tamaño de 1024 bytes con una alta varianza permite maximizar la seguridad del proceso de cifrado/descifrado.

Esto será descrito más adelante, en la sección *Eficiencia de la Clave Simétrica*.

El criptosistema incorpora una aceptable eficiencia computacional[13] en cada una de las fases.

13 La eficiencia computacional se establece por una función de complejidad que determina el desempeño del algoritmo correspondiente.

3.3.2.5 CLAVE SIMÉTRICA DE CIFRADO/DESCIFRADO

La criptografía simétrica requiere la adopción de un mecanismo que garantice los procesos de codificación y recuperación de la información.

Este mecanismo consiste en la utilización de una misma clave, denominada *clave simétrica* a los procesos de cifrado y descifrado de la información.

- *DETERMINACIÓN DE LA CLAVE SIMETRICA*

La clave simétrica a utilizar es un vector de n componentes

$$\Psi = (\Psi_1, \Psi_2, \Psi_3, \dots, \Psi_N)$$

pertenecientes a un determinado alfabeto λ

$$\Psi_i \in \lambda, \qquad 1 \leq i \leq N$$

La varianza v de la clave Ψ se calcula como

$$v = \frac{1}{N}\sum_{i=1}^{N}(\Psi_i - \overline{\Psi})^2$$

donde $\overline{\Psi}$ es la media del vector Ψ

$$\overline{\Psi} = \frac{1}{N}\sum_{i=1}^{N}\Psi_i$$

La varianza de la clave Ψ es directamente proporcional a la eficiencia de su utilización, y debe ser diferente de cero.

Si una clave tiene varianza cero entonces sus N componentes son iguales

$$\Psi = (C, C, C, \dots, C)$$

lo cual no garantiza la eficiencia de la seguridad del proceso de cifrado de la información, propiciando que el descifrado sea muy fácil.

No obstante, si una clave tiene una máxima varianza entonces se maximiza la seguridad del proceso de cifrado de la información.

- *EFICIENCIA DE LA CLAVE SIMETRICA*

Sea la clave simétrica

$$\Psi = (\Psi_1, \Psi_2, \Psi_3, \dots, \Psi_N)$$

$$|\Psi| = N$$

Es conveniente definir el espacio o dominio de valores que van a representar la clave del criptosistema.

Luego, si el esquema de la clave es de 1 byte equivalente a 8 bits, se tienen $2^8 = 256$ posibilidades de representación

Generalizando el esquema a N bytes, equivalente a $8N$ bits, se tiene $2^{8N} = (2^8)^N = 256^N$ posibilidades de representación.

En la presente investigación se trabajará con un esquema de 1024 bytes equivalente a 8192 bits cuya cantidad de posibilidades de representación está dado por 2^{8192}.

$$2^{8192} = (2^4)^{2048} >> 10^{2048}$$

una cantidad absolutamente muy grande desde el punto de vista de procesamiento de datos que dificulta los procedimientos de criptoanálisis, garantizando solidez y robustez al criptosistema.

3.3.2.6 ESPECIFICACIONES DEL AUTÓMATA CELULAR

Para garantizar la simetría del criptosistema, esto es, codificación y decodificación de la información, se va a utilizar un modelo de autómata celular θ basado en los siguientes aspectos:

- *TIPOLOGÍA DEL AUTÓMATA*

  El autómata celular θ debe adoptar las siguientes características:

  - ✓ *Lineal*, el autómata debe ser elemental con vecinos a la derecha e izquierda.
  - ✓ *Circular,* el autómata debe ser circular y debe adoptar reglas que involucren a sus elementos extremos.
  - ✓ *Reversible*, el autómata debe ser bidireccional de modo que sus reglas de transición hagan posible la codificación/decodificación de la información.

- *ESTRUCTURA DEL AUTÓMATA*

  El autómata celular θ estará conformado por T células. Su representación se define en función del paso discreto del tiempo, adoptando múltiples instancias o configuraciones.

  El autómata celular θ tiene asociado una determinada configuración $\theta^{(t)}$ que contiene las instancias de cada célula en el instante t

$$\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_T^{(t)})$$

- *ESTADOS DEL AUTÓMATA*

  Se va a considerar un total de 10 estados que adoptará cada célula del autómata, los cuales pertenecen al conjunto de estados

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

  Expresado de otra manera, el autómata celular realizará la transición de estados sobre los elementos del conjunto Ω.

- *FUNCIÓN DE TRANSFORMACIÓN O MAPEAMIENTO*

  Los estados del autómata θ representan los valores que adoptará durante su evolución y se obtienen aplicando reglas de transición.

  Se va a considerar 10 estados que adoptará cada célula del autómata

  $$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

  Dado el conjunto

  $$\eta = \{0,1,2,3, \dots ,254,255\}$$

  En el autómata celular θ, la función

  $$\zeta: \eta \rightarrow \Omega$$

  genera un mapeamiento[14] de η sobre Ω asignando un determinado estado a cada valor de ocurrencia. Esta función será descrita con detalle en la fase de *Diseño del Sistema*.

- *VECINDAD DEL AUTÓMATA*

  El autómata celular θ debe estar constituido por T células, cada una de las cuales tendrá asociada una vecindad de longitud 3.

  Para un autómata celular de tipo circular, se tiene la definición de la vecindad para la k-ésima célula del autómata, donde $1 \leq k \leq T$:

  - $Vecino\ Izquierdo(k) = \begin{cases} Célula\ (T), & k = 1 \\ Célula\ (k-1), & k > 1 \end{cases}$
  - $Vecino\ Derecho\ (k) = \begin{cases} Célula\ (1), & k = T \\ Célula\ (k+1), & k < T \end{cases}$

14 La función de transformación o mapeamiento se establece mediante un procedimiento algorítmico.

3.3.3 DISEÑO DEL MODELO

Las fases anteriores 3.3.1 y 3.3.2, permitieron la identificación del problema, la descripción de sus especificaciones, el planteamiento del mismo, y el establecimiento de los orígenes y mecanismos de recolección de datos.

Esta fase contiene los elementos orientados al diseño del modelo asociado al criptosistema siendo de vital importancia para su posterior implementación computacional.

3.3.3.1 DETERMINACIÓN DE LA POBLACIÓN

La población *P* a utilizar está asociada a una determinada instancia del criptosistema. Para tal efecto, se consideran los siguientes puntos:

- *SECUENCIA DE FORMATOS O TIPOS DE INFORMACIÓN*

Los ficheros son categorizados de acuerdo a su tipología, la cual determina el formato de información que contienen.

Si T representa el total de tipos de información existentes durante un determinado instanciamiento del criptosistema, entonces la secuencia

$$E_1, E_2, E_3, \dots E_{T-1}, E_T$$

representa todos los formatos existentes.

- *SUBPOBLACIÓN POR FORMATOS O TIPOS DE INFORMACIÓN*

Dada una población P, se requiere determinar la cantidad de elementos que corresponden a un determinado formato de información. Se define la secuencia de frecuencias subpoblacionales

$$N_1, N_2, N_3, \dots N_{T-1}, N_T$$

donde N_i es el total de ficheros correspondiente a cada formato de información E_i.

- *TAMAÑO DE LA POBLACIÓN*

  El tamaño de la población en una determinada instancia del criptosistema, consiste en totalizar las frecuencias subpoblacionales

  $$N_1, N_2, N_3, \dots N_{T-1}, N_T$$

  asociadas a cada tipo de información. Luego, el tamaño de la población P se determina mediante

  $$N = \sum_{i=1}^{T} N_i$$

- *PROCEDIMIENTO DE DETERMINACIÓN DE LA POBLACIÓN*

  Se tiene:

  - Establecer la ruta/unidad de almacenamiento externo del computador como escenario de aplicación.
  - Obtener la información (*nombre*, *tipo*, *tamaño*, etc.) de todos los ficheros contenidos en el equipo y guardarlos en un fichero BUSQUEDA.DAT.
  - A partir del fichero obtenido, generar el fichero POBLACION.DAT el cual tiene la siguiente estructura:
    - Nombre
    - Extensión
    - Tamaño
    - Categoría

  El fichero POBLACION.DAT obtenido constituye la representación física de la población P.

  Para determinar la población P se va a utilizar un instrumento automatizado[15] implementado por el autor.

15 El directorio de archivos contenidos en el equipo, es obtenida mediante comandos de búsqueda con profundidad contenidos en un proceso por lotes.

3.3.3.2 FORMULACIÓN Y DISEÑO DE LA MUESTRA

Las especificaciones de la muestra S de una población de ficheros P y su asociación a una instancia del criptosistema fueron planteadas anteriormente.

Se va a considerar los siguientes puntos:

- *CLUSTERIZACIÓN DE LA MUESTRA*

Los formatos o tipos de contenido, corresponden a determinadas categorías de información y son determinadas por la extensión del nombre del fichero, tales como texto, imagen, audio, video, documento, hoja de cálculo, presentación de diapositivas, plantilla, programa, etc.

Por tanto, se requiere realizar un agrupamiento o clustering al procedimiento de selección de datos de la muestra.

En la presente investigación, se ha propuesto una muestra de tamaño $n = 1000$ elementos distribuidos en $K = 22$ clusters cuya esquema de distribución [16] y tamaño correspondiente se muestra siguiente figura.

<table>
<tr><td rowspan="5">200</td><td rowspan="3">100</td><td rowspan="3">100</td><td rowspan="3">100</td><td>20</td><td>10</td><td>10</td></tr>
<tr><td>20</td><td>10</td><td>10</td></tr>
<tr><td>20</td><td>10</td><td>10</td></tr>
<tr><td rowspan="2">100</td><td rowspan="2">100</td><td rowspan="2">100</td><td>20</td><td>10</td><td>10</td></tr>
<tr><td>20</td><td>10</td><td>10</td></tr>
</table>

Figura 3.5 Organización de clusters de ficheros según tipo de contenido.
Fuente: Elaboración Propia.

La similaridad de los elementos de cada cluster se determina por el formato o tipo de información que contienen.

16 Dada la naturaleza de la investigación, el esquema de representación de cluster ha sido propuesto por el autor.

Luego, cada cluster debe contener una cantidad de ficheros con la misma similaridad, de acuerdo al esquema de organización proporcionado en la figura anterior.

La siguiente tabla visualiza la distribución de ficheros adoptada en la investigación [17] , por cada cluster, totalizando $n = 1000$ representando el tamaño de la muestra.

Tabla 3.2 Distribución de ficheros por clusters. Fuente: Elaboración Propia.

Nro CLUSTER	TIPO DE INFORMACIÓN	CANTIDAD
1	Texto	200
2	Imagen	100
3	Audio	100
4	Video	100
5	Documentos	100
6	Hoja de Calculo	100
7	Presentador de Diapositivas	100
8	Scripts y Programas	20
9		20
10		20
11		20
12		20
13	Otros Formatos	10
14		10
15		10
16		10
17		10
18		10
19		10
20		10
21		10
22		10
	TOTAL	1000

17 El muestreo es obtenido por un proceso de simulación automatizado correspondiente a un problema de optimización con restricciones y descrito en la siguiente página. Los datos presentados (texto, imagen, audio, video, etc. han sido elegidos para fines de comprensión.

- *FORMULACIÓN DEL MUESTREO*

  Sea S una muestra representativa de una población P de ficheros correspondientes a una instancia del criptosistema.

  - La muestra S es obtenida de la población P y debe ser distribuida en un conjunto de K clusters, de acuerdo al criterio de similaridad basada en el formato o tipo de contenido.
  - La secuencia decreciente

    $$n_1, n_2, n_3, \dots n_{K-1}, n_K, \quad n_j \geq n_{j+1}, 1 \leq j \leq K$$

    determina el tamaño de cada cluster. Luego, se cumple

    $$\sum_{j=1}^{K} n_j = n$$

  - Las K frecuencias máximas que corresponden a subpoblaciones de P por formato o tipo de información,

    $$f_1, f_2, f_3, \dots f_{K-1}, f_K, \quad f_j \geq f_{j+1}, 1 \leq j \leq K$$

    se determinan a partir de las frecuencia subpoblacionales $\{N_i\}$ y haciendo uso del modelo

    $$Max\ Z = \sum_{j=1}^{T} N_i\, x_i$$

    sujeto a las restricciones

    $$\sum_{j=1}^{T} x_i = K$$

    $$x_i \in \{0,1\}$$

  - Análogamente, los K formatos, que son seleccionados

    $$e_1, e_2, e_3, \dots e_{K-1}, e_K$$

    son obtenidos del conjunto $\{E_i\}$ de la población.

  - La selección de los elementos que van a conformar cada cluster, se determina aleatoriamente a partir del conjunto de frecuencias máximas $\{f_j\}$ y los formatos de contenido $\{e_j\}$ obtenidos.

- *EXTRACCIÓN ALEATORIA*

  La selección de los elementos poblacionales que van a conformar la muestra representativa, deben considerar:

  - Se requiere K pasos que corresponden a la cantidad de clusters a utilizar.
  - Para un determinado cluster, en cada paso, la selección de elementos es sin sustitución de elementos adoptando probabilidad condicional dependiente de las selecciones de los pasos anteriores.
  - En otras palabras, el tamaño $t = f_i$ correspondiente a una determinada subpoblación con frecuencia máxima, varía dentro de un proceso iterativo, adoptando en cada subpaso de referencia una distribución de probabilidad discreta uniforme. Si el cluster en referencia está formado por m elementos, las probabilidades de las m extracciones se visualizan en la siguiente tabla

Tabla 3.3 Procedimiento de Selección en un Cluster. Fuente: Elaboración Propia.

Extracción	Cantidad	Probabilidad
1	t	$\frac{1}{t}$
2	$t-1$	$\frac{1}{t-1}$
3	$t-2$	$\frac{1}{t-2}$
...	...	...
j	$t-(j-1)$	$\frac{1}{t-(j-1)}$
...	...	...
$m-1$	$t-(m-2)$	$\frac{1}{t-(m-2)}$
m	$t-(m-1)$	$\frac{1}{t-(m-1)}$

  - Se debe garantiza que la frecuencia máxima obtenida f_i asociada a un tipo de información e_i, debe satisfacer

$$f_i \leq n_i, \qquad 1 \leq i \leq K$$

- *PROCEDIMIENTO DE OBTENCIÓN DE LA MUESTRA*

Se tiene:

- Inicializar los valores adoptados en la investigación:
  - Tamaño de la muestra $n = 1000$
  - Número de clusters $K = 22$
  - Secuencia decreciente

$$n_1, n_2, n_3, \dots n_{K-1}, n_K$$

con los valores

$$\{200,100,100,100,100,100,100,20,20,20,20,20,10,10,10,10,10,10,10,10,10,10\}$$

que representan los tamaños de los clusters

- Establecer el fichero generado POBLACION.DAT como lectura
- Obtener el total T de formatos o tipos de información existente en el sistema, verificando que $K \leq T$
- Establecer la secuencia de tipos de información poblacionales

$$E_1, E_2, E_3, \dots E_{T-1}, E_T$$

- Determinar las frecuencias subpoblacionales de acuerdo a $\{E_i\}$

$$N_1, N_2, N_3, \dots N_{T-1}, N_T$$

- Calcular las K frecuencias máximas subpoblaciones y sus correspondientes tipos de información a partir de los conjuntos $\{N_i\}$ y $\{E_i\}$ respectivamente.

$$f_1, f_2, f_3, \dots f_{K-1}, f_K$$

$$e_1, e_2, e_3, \dots e_{K-1}, e_K$$

- Realizar la selección aleatoria de elementos, para un cluster determinado, usando las especificaciones en del punto anterior

El procedimiento descrito, describe el proceso de obtención de la muestra.

La implementación de este procedimiento es un instrumento, consistente en un proceso automatizado realizado por el autor.

3.3.3.3 CRIPTOGRAFÍA Y CRIPTOANÁLISIS

El criptosistema debe garantizar la criptografía y el criptoanálisis, esto es, debe realizar el cifrado/descifrado mediante una clave simétrica que garantiza correspondientemente la codificación/decodificación de la información.

Los ficheros deben ser cifrados/descifrados de modo que garanticen la confidencialidad y reserva de la información, adoptando una razonable complejidad de procesamiento.

Los siguientes criterios son descritos y son importantes para su posterior implementación.

- *ESQUEMA DEL CIFRADO/DESCIFRADO SIMETRICO*

Cada elemento del criptosistema será cifrado usando una determinada clave para garantizar la recuperación de la información, según se ilustra en el siguiente esquema

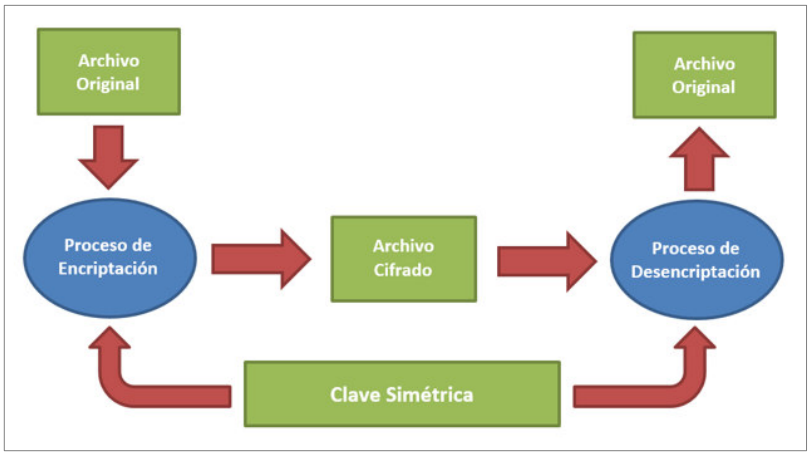

Figura 3.6 Cifrado/Descifrado Simétrico de la Información. Fuente: Elaboración Propia.

3.3.3.4 FLUJO EXTERNO DE INFORMACIÓN

- *ESPECIFICACIÓN DEL ALFABETO DE SIMBOLOS*

  El criptosistema requiere la utilización de un alfabeto de símbolos λ.

  Sea η un subconjunto de los enteros no negativos definido como

  $$\eta = \{0,1,2,3, \dots ,254,255\}$$

  asociado biunívocamente a un conjunto λ finito numerable

  $$\lambda = \{\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_{254}, \lambda_{255}\} = \{\lambda_i / i \in \eta\}$$

  Existe una correspondencia biunívoca entre los elementos de η y λ.

  Cada elemento λ_i está asociado a cada secuencia binaria de 8 bits, según se muestra en la figura

	b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	1	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
253	1	1	1	1	1	1	0	1
254	1	1	1	1	1	1	1	0
255	1	1	1	1	1	1	1	1

Figura 3.7 Secuencia binaria de 8 bits. Fuente: Elaboración Propia.

- *FICHEROS DE FLUJO DE BYTES*

  Sea F un fichero de información perteneciente al almacenamiento externo en una determinada instancia del sistema de ficheros.

  Luego, F puede ser definido como un vector F perteneciente a un espacio L-dimensional de L bytes

  $$\mathrm{F} = (\mathrm{F}_0, \mathrm{F}_1, \mathrm{F}_2, \dots, \mathrm{F}_{\mathrm{L}-1})$$

  Cada componente $\mathrm{F_j}$ pertenece al mismo alfabeto de símbolos λ

  $$\mathrm{F_j} \in \lambda, 0 \leq j \leq L-1$$

Desde el punto de vista físico, cada fichero es un flujo o secuencia finita de L bytes, los cuales se encuentran localizados en una determinada posición que constituye su dirección física.

El direccionamiento es a partir de 0, tal como se indica en la figura

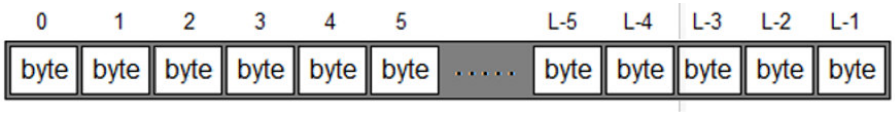

Figura 3.8 Direccionamiento en fichero de L bytes. Fuente: Elaboración Propia.

- *ORGANIZACIÓN Y ESTRUCTURA DE DATOS*

La lectura del fichero de L bytes, se realizará de modo iterativo en un determinado número de pasos secuenciales dependiendo directamente de su tamaño.

Por tanto, se requiere disponer de una estructura de datos que permita contener la lectura de cada grupo de N bytes del fichero correspondiente a cada paso.

Figura 3.9 Vector de Lectura de Bloque de N bytes. Fuente: Elaboración Propia.

Dicho de otro modo, durante la operación de lectura del fichero, en cada paso del bucle, se cargará el vector V con N elementos leidos directamente del fichero.

Esto significa que en el j-esimo paso de lectura, el vector V contendrá los elementos $(j-1)N$ al $jN-1$ del fichero.

	Posición	
Paso	Inicio	Fin
1	0	N-1
2	N	2N-1
3	2N	3N-1
...	....	...
j	(j-1)N	jN-1
...	....	...
...	....	...
...	....	...

Figura 3.10 Direccionamiento de Lectura de Bloque de N bytes. Fuente: Elaboración Propia.

- *ESTRUCTURA INTERNA DE INFORMACIÓN*

El vector F correspondiente a un determinado fichero de información $\mathbb{F}$ puede ser asociado biunívocamente a una estructura vectorial interna β conformado por T bloques de N bytes.

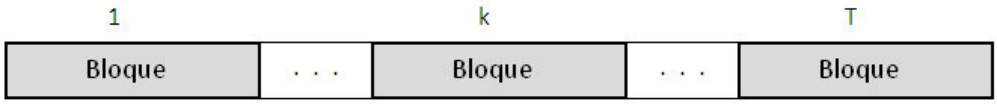

Figura 3.11 Estructura interna de información. Fuente: Elaboración Propia.

En símbolos:

$$\beta = (\beta_1, \beta_2, \beta_3, \ldots, \beta_T), \qquad 1 \leq k \leq T$$

Cada bloque β_k del vector constituye la unidad fundamental del proceso y se denomina bloque interno de información. Luego

- Cuando los T bloques son de igual tamaño N

$$|\beta_k| = N, \qquad 1 \leq k \leq T$$

$$L = N * T$$

- Cuando el último bloque T del autómata tiene tamaño E<N.

$$|\beta_k| = \begin{cases} N, & 1 \leq k < T \\ E, & k = T \end{cases}$$

$$L = N * (T - 1) + E$$

Para fines de comprensión y facilidad en el diseño e implementación, se va a considerar todos los bloques de igual tamaño.

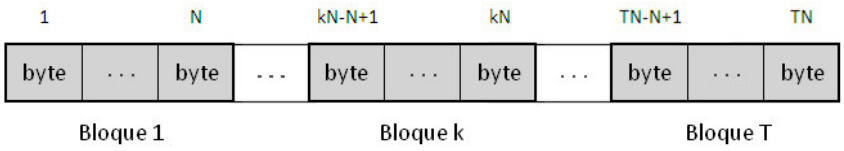

Figura 3.12 Organización de bytes por bloque. Fuente: Elaboración Propia.

Luego, el k-ésimo bloque β_k, $1 \leq k \leq T$, se indexa del siguiente modo

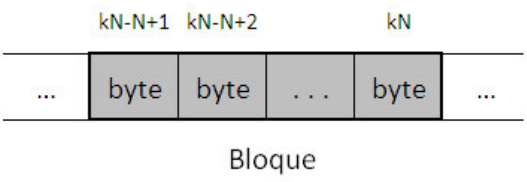

Figura 3.13 Indexación del k-ésimo bloque

3.3.3.5 CLAVE SIMÉTRICA DE CIFRADO/DESCIFRADO

El proceso correspondiente a la generación de la clave simétrica a utilizar es descrito en las siguientes secciones

- *DETERMINACIÓN DE LA CLAVE*

La generación de la clave simétrica requiere la adopción de un procedimiento que maximiza la varianza sobre un conjunto de claves aleatoriamente generadas

$$\{\Psi^{(1)}, \Psi^{(2)}, \Psi^{(3)}, \quad \dots, \Psi^{(n-1)}, \Psi^{(n)}\}$$

donde cada clave $\Psi^{(k)}$ es definida como un vector N-dimensional

$$\Psi^{(k)} = (\Psi_1^{(k)}, \Psi_2^{(k)}, \Psi_3^{(k)}, \dots, \Psi_N^{(k)})$$

Así mismo, cada clave $\Psi^{(k)}$ tiene su correspondiente varianza v_k que es calculada sobre sus elementos

$$v_k = \frac{1}{N}\sum_{i=1}^{N}\left(\Psi_i^{(k)} - \bar{\Psi}^{(k)}\right)^2$$

donde $\bar{\Psi}^{(k)}$ es la media del vector $\Psi^{(k)}$

$$\bar{\Psi}^{(k)} = \frac{1}{N}\sum_{i=1}^{N}\Psi_i^{(k)}$$

Por lo tanto, la varianza máxima es definida como

$$v_M = max\{v_1, v_2, v_3, \dots, v_n\}$$

y está asociada a la clave $\Psi^{(M)}$, cuya obtención se va a realizar por simulación mediante un proceso iterativo de $n = 100$ pasos.

Si v_M representa la varianza máxima del conjunto de claves

$$\{\Psi^{(1)}, \Psi^{(2)}, \Psi^{(3)}, \dots, \Psi^{(n)}\},$$

entonces $\Psi^{(M)}$ es la clave deseada.

- *COMPLEJIDAD DE LA CLAVE SIMETRICA*

  La longitud de la clave simétrica adoptada debe ser equivalente al tamaño del bloque interno de información. De este modo, si Ψ representa la clave del criptosistema, se tiene

  $$\Psi = (\psi_1, \Psi_2, \Psi_3, \dots, \Psi_N),\ 1 \leq i \leq N$$

  $$|\Psi| = n$$

  Si la clave simétrica tiene un esquema de 1 byte igual a 8 bits, se tienen $2^8 = 256$ posibilidades de representación

  En la presente investigación, la clave simétrica tiene un esquema de 1024 bytes equivalente a 8192 bits. El número de posibilidades

  $$2^{8192} = (2^4)^{2048} >> 10^{2048}$$

  constituye una cantidad absolutamente muy grande desde el punto de vista computacional proporcionando robustez al criptosistema.

- *GENERACIÓN DE LA CLAVE*

  El siguiente cuadro ilustra las n iteraciones a partir del cual se obtiene la máxima varianza que representara la clave a utilizar

  Tabla 3.4 Cuadro de iteraciones para generación de claves

Iteración	Clave Generada	Varianza
1	$\Psi^{(1)} = (\Psi_1^{(1)}, \Psi_2^{(1)}, \Psi_3^{(1)}, \dots, \Psi_N^{(1)})$	v_1
2	$\Psi^{(2)} = (\Psi_1^{(2)}, \Psi_2^{(2)}, \Psi_3^{(2)}, \dots, \Psi_N^{(2)})$	v_2
3	$\Psi^{(3)} = (\Psi_1^{(3)}, \Psi_2^{(3)}, \Psi_3^{(3)}, \dots, \Psi_N^{(3)})$	v_3
$\vdots$	$\vdots$	$\vdots$
n	$\Psi^{(n)} = (\Psi_1^{(n)}, \Psi_2^{(n)}, \Psi_3^{(n)}, \dots, \Psi_N^{(n)})$	v_n

  donde

  $$v_k = \frac{1}{N}\sum_{i=1}^{N}\left(\Psi_i^{(k)} - \bar{\Psi}^{(k)}\right)^2,\ \ k = 1,2,3,\dots,n$$

  $$\bar{\Psi}^{(k)} = \frac{1}{N}\sum_{i=1}^{N}\Psi_i^{(k)},\quad k = 1,2,3,\dots,n$$

3.3.3.6 AUTÓMATA CELULAR

De acuerdo a las especificaciones indicadas en la fase anterior, en esta fase se describe el diseño del autómata celular.

- *TIPOLOGÍA DEL AUTÓMATA CELULAR*

El autómata celular debe ser lineal, circular y reversible.

- *Lineal*, porque puede ser representado como un vector unidimensional con vecinos a la izquierda y derecha.
- *Circular*, es una extensión de la característica lineal del autómata. El vecino izquierdo del primer elemento es el último. Análogamente, el vecino derecho del último elemento es el primero.
- *Reversible*, el cifrado/descifrado de la información, se maneja mediante de reglas de transición elementales, tales como la complementación y reversión de una secuencia de bits.

La siguiente figura ilustra el modelo de autómata conformado por T elementos o células

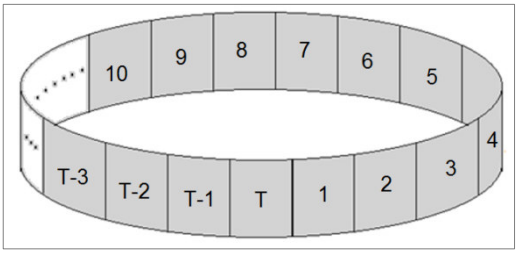

Figura 3.14 Modelo de Autómata Celular de T células. Fuente. Elaboración Propia.

- *ESTADOS DEL AUTÓMATA*

Los estados que adoptara el autómata celular, establecidos en el conjunto

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

han sido obtenido mediante una función ζ de transformación o mapeamiento de los valores del conjunto

$$\eta = \{0,1,2,3, \dots ,254,255\}$$

- *FUNCIÓN DE TRANSFORMACIÓN O MAPEAMIENTO*

La función de transición de estados en el autómata celular

$$\zeta: \eta \rightarrow \Omega$$

realiza una transformación sobre el conjunto η generando una imagen en Ω.

Existen muchas modalidades para obtener esta transformación. El mecanismo general es particionar el conjunto η en 10 subconjuntos

$$\eta_0, \eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7, \eta_8, \eta_9$$

asignando a cada elemento del subconjunto η_i el estado i.

No obstante, la función ζ utilizada es una función algorítmica denominado RDS (*Reduced Digit Sum*) el cual proporciona la siguiente distribución de frecuencias o probabilidades de ocurrencia

Tabla 3.5 Estados del Autómata y Probabilidad de Ocurrencia.
Fuente: Elaboración Propia.

Estado	Frecuencia	Probabilidad
0	1	0.00390625
1	29	0.11328125
2	29	0.11328125
3	29	0.11328125
4	28	0.10937500
5	28	0.10937500
6	28	0.10937500
7	28	0.10937500
8	28	0.10937500
9	28	0.10937500
	256	**1.00000000**

El algoritmo de transformación y mapeamiento de estados RDS asociada a la función ζ se muestra a continuación

```
Function RDS(X)
   While 10<X
     S =0
     While X!=0
       D = mod(X/10)
       X = int(X/10)
       S = S + D
     End
     X= S
   End
   Return X
End
```

- *VECINDAD DEL AUTÓMATA*

El autómata está conformado por T células. Cada célula corresponde a una determinada vecindad.

Para la k-esima célula del autómata celular, donde $1 \leq k \leq T$, se tiene

- Si la célula actual se encuentra en los extremos del autómata. Hay dos posibilidades:
  - Si $k = 1$ entonces:

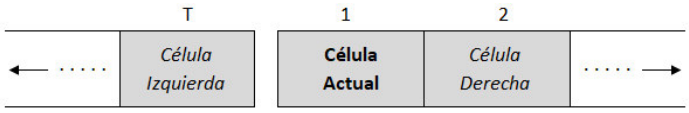

Figura 3.15 Vecindad de la primera célula como elemento extremo.

  - Si $k = T$ entonces:

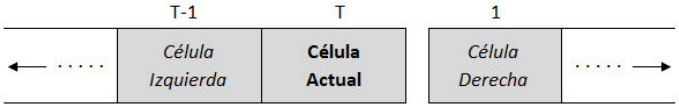

Figura 3.16 Vecindad de la última célula como elemento extremo.

- Si la célula actual no se encuentra en los extremos del autómata:

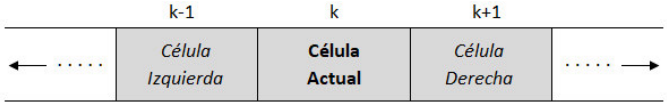

Figura 3.17 Vecindad de un elemento no extremo.

Lo anterior puede ser resumido

- $Vecino\ Izquierdo(k) = \begin{cases} Célula\ (T), & k = 1 \\ Célula\ (k-1), & k > 1 \end{cases}$
- $Vecino\ Derecho\ (k) = \begin{cases} Célula\ (1), & k = T \\ Célula\ (k+1), & k < T \end{cases}$

El total de vecindades en el autómata celular es equivalente al número de células que contiene.

Si existen T células, entonces existen T vecindadesde con longitud 3.

- *EVOLUCIÓN DEL AUTÓMATA CELULAR*

Las reglas de transición de estados están asociados a la evolución del autómata celular a pasos discretos del tiempo.

Tal como se indicó, se van a considerar 10 estados que adoptarán cada célula componente del autómata, los cuales están establecidos en el conjunto Ω

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

La evolución del autómata celular θ en el tiempo genera diferentes configuraciones que representan la sucesión de instancias $\{\theta^{(t)}\}$ del autómata durante la variación discreta del tiempo:

$$\{\theta^{(t)}/t = 0,1,2,3, \dots\} = \{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \dots\}$$

donde:

- $\theta^{(0)}$ Configuración inicial del autómata celular
- $\theta^{(1)}$ Configuración del autómata celular en el paso del tiempo 1
- $\theta^{(2)}$ Configuración del autómata celular en el paso del tiempo 2
- $\theta^{(3)}$ Configuración del autómata celular en el paso del tiempo 3

. . .

Una configuración o instancia $\theta^{(t)}$ del autómata celular θ en el tiempo t es un vector que la configuración de estados de todas las células del autómata en el instante o paso discreto del tiempo t.

$$\theta^{(t)} = \left(\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_T^{(t)}\right), \qquad \theta_k^{(t)} \epsilon\, \Omega, \qquad k = 0,1,2, \dots, T$$

De este modo, $\theta_k^{(t)}$ contiene la instancia o estado de la k-esima célula del autómata en el instante o paso discreto del tiempo t.

Luego, para cada bloque β_k de información, se tiene asociado una sucesión de estados $\theta_k^{(t)}, t = 0,1,2, \dots, P$, durante la variación discreta del tiempo, que representa su historial de evolución.

$$H_k = (\theta_k^{(0)}, \theta_k^{(1)}, \dots, \theta_k^{(t)}, \dots \theta_k^{(P)})$$

donde P es el periodo que establece el ciclo de repetición del sistema.

Por tanto, al extender el historial a todo el autómata celular, se tiene

$$H = \begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ \vdots \\ H_T \end{bmatrix} = \begin{bmatrix} \theta_1^{(0)} & \theta_1^{(1)} & \theta_1^{(2)} & \dots & \theta_1^{(P)} \\ \theta_2^{(0)} & \theta_2^{(1)} & \theta_2^{(2)} & \dots & \theta_2^{(P)} \\ \theta_3^{(0)} & \theta_3^{(1)} & \theta_3^{(2)} & \dots & \theta_3^{(P)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_T^{(0)} & \theta_T^{(1)} & \theta_T^{(2)} & \dots & \theta_T^{(P)} \end{bmatrix}$$

- REGLAS DE CONFIGURACIÓN Y TRANSICIÓN DE ESTADOS

Las reglas de transición determinan la evolución del autómata durante la variación discreta del tiempo.

La obtención de la nueva configuración $\theta^{(t+1)}$ del autómata celular depende directamente del estado anterior $\theta^{(t)}$.

La configuración $\theta_k^{(t+1)}$ de la k-esima célula, es función[18] de su correspondiente vecindad en la configuración anterior $\theta_k^{(t)}$

$$\theta^{(t+1)}_k = f(\theta_{k-1}^{(t)}, \theta_k^{(t)}, \theta_{k+1}^{(t)})$$

y requiere el producto cartesiano del conjunto de estados Ω

$$\Omega x \Omega x \Omega = \{(0,0,0), (0,0,1), \dots, (9,9,9)\}$$

cuya partición determina la funcionalidad o mecanismo de acción según se ilustra en la tabla

Tabla 3.6 Funciones de Transformación de Bloque. Fuente: Elaboración Propia.

Rango de Estados Tiempo t			Estado Tiempo t+1
Vecina Izquierda	Célula	Vecina Derecha	Acción
Del (0,0,0) Al (0,9,9)			Rotación Derecha
Del (1,0,0) Al (1,9,9)			Rotación Izquierda
Del (2,0,0) Al (2,9,9)			Unión Impar/Par
Del (3,0,0) Al (3,9,9)			Unión Par/Impar
Del (4,0,0) Al (4,9,9)			Transposición Par/Impar Ascendente
Del (5,0,0) Al (5,9,9)			Transposición Par/Impar Descendente
Del (6,0,0) Al (6,9,9)			EquiDistancia Ascendente Impar
Del (7,0,0) Al (7,9,9)			EquiDistancia Descendente Impar
Del (8,0,0) Al (8,9,9)			EquiDistancia Ascendente Par
Del (9,0,0) Al (9,9,9)			EquiDistancia Descendente Par

18 Para denotar la función de la vecindad de la k-ésima célula en el estado anterior, se utiliza f como representación de la dependencia existente.

3.3.3.7 ALGORITMOS DE TRANSFORMACIÓN DE BLOQUE

La funcionalidad especificada en la tabla anterior, requiere ser diseñada mediante procedimientos algorítmicos.

Los siguientes algoritmos son fundamentales para realizar las operaciones de transformación de bloque en cada paso de transición de estados.

Los algoritmos más elementales, lo constituyen la complementación y reversión de una secuencia de bytes.

- *COMPLEMENTACIÓN*

Cada byte que conforma el k-ésimo bloque interno de información debe ser complementado respecto a 255.

```
// Complemento de Vector de bits
Procedure Complement(V[], n)
    For i=0 To n-1
        V[i]=(255-V[i])
    EndFor
EndProcedure
```

- *REVERSIÓN*

Cada byte que conforma el bloque debe ser intercambiado con su reverso que es el elemento que tienen igual distancia desde el centro.

```
// Reverso de Vector de bits
Procedure Reverse(V[], n)
    If Mod(n/2)==0 Then
       P=int((n-1)/2)
    Else
       P=int(n/2)
    EndIf
    For i=0 To P
        TMP  = V[i]
        V[i] = V[n-i-1]
        V[n-i-1] = TMP
    EndFor
EndProcedure
```

Los siguientes algoritmos realizan la rotación de los elementos de un bloque de información hacia la derecha e izquierda respectivamente.

- *ROTACIÓN A LA IZQUIERDA*

```
// Rotacion del k-ésimo bloque 't' veces a la Izquierda
Procedure RotaryLeft(V[], n, t)
    For k=1 To t
        Tmp=V[0]
        For i=1 To n-1
            V[i-1]=V[i]
        EndFor
        V[n-1]=Tmp
    EndFor
EndProcedure
```

- *ROTACIÓN A LA DERECHA*

```
// Rotacion del k-ésimo bloque 't' veces a la Derecha
Procedure RotaryRight(V[], n, t)
    For k=1 To t
        Tmp=V[n-1]
        For i=n-2 DownTo 0
            V[i+1]=V[i]
        EndFor
        V[0]=Tmp;
    EndFor
EndProcedure
```

Se observa la equivalencia de las operaciones `RotaryRight(k)` y `RotaryLeft(N-k)`.

Los algoritmos siguientes realizan el incremento y decremento en los elementos de un bloque, de un valor escalar constante $n, 0 \leq n \leq 255$.

- *INCREMENTO CONSTANTE*

```
// Incremento Constante del Bloque
Procedure Increase_Constant(V[], n, t)
    For i=0 To n-1
        V[i] = Mod(V[i]+t)/256)
    EndFor
EndProcedure
```

- *DECREMENTO CONSTANTE*

```
// Decremento Constante del Bloque
Procedure Decrease_Constant(V[], n, t)
    For i=0 To n-1
        V[i] = Mod((V[i]-t)/256)
    EndFor
EndProcedure
```

Análogamente, los siguientes algoritmos realizan el incremento y decremento en los elementos de un bloque, de un vector de elementos $W, 0 \leq w_i \leq 255$.

- *INCREMENTO VARIABLE*

```
// Incremento Variable del Bloque
Procedure Increase_Variable(V[], n, KEY[])
    For i=0 To n-1
        V[i] = Mod(V[i]+K[i])/256)
    EndFor
EndProcedure
```

- *DECREMENTO VARIABLE*

```
// Decremento Variable del Bloque
Procedure Decrease_Variable(V[], n, KEY[])
    For i=0 To n-1
        V[i] = Mod((V[i]-K[i])/256)
    EndFor
EndProcedure
```

A continuación, se exponen algoritmos que realizan transformaciones por posicionamiento en una secuencia de bytes.

- *UNIÓN IMPAR/PAR Y UNIÓN PAR/IMPAR*

```
// Union de Elementos Impar/Par
Procedure JoinImparPar(BLK, n, Mode)
    If Mode==1 Then
       P=0
       For i=0 To n-1
            If Mod(i/2)==0 Then
               W[P]=BLK[i]
               P=P+1
            EndIf
       EndFor
       For i=0 To n-1
            If Mod(i/2)!=0 Then
               W[P]=BLK[i]
               P=P+1
            EndIf
       EndFor
       For i=0 To n-1
            BLK[i]=W[i]
       EndFor
    Else
       t=Choose(Mod(n/2)==0, n/2, (n+1)/2)
       For i=0 To t-1
            W[2*i]=BLK[i]
       EndFor
       For i=t;i<=n-1
            W[2*(i-t)+1]=BLK[i]
       EndFor
       For i=0 To n-1
            BLK[i]=W[i]
       EndFor
    EndIf
EndProcedure
```

```
// Union de Elementos Par/Impar
Procedure JoinParImpar(BLK, n, Mode)
    If Mode==1 Then
       P=0
       For i=0 To n-1
            If Mod(i/2)!=0 Then
               W[P]=BLK[i]
               P=P+1
            EndIf
       EndFor
       For i=0 To n-1
            If Mod(i/2)==0 Then
               W[P]=BLK[i]
               P=P+1
            EndIf
       EndFor
       For i=0 To n-1
            BLK[i]=W[i]
       EndFor
    Else
       t=Choose(Mod(n/2)==0, n/2, (n-1)/2)
       For i=0 To t-1
            W[2*i+1]=BLK[i]
       EndFor
       For i=t To n-1
            W[2*(i-t)]=BLK[i]
       EndFor
       For i=0 To n-1
          BLK[i]=W[i]
       EndFor
    EndIf
EndProcedure
```

- *EQUIDISTANCE_INSIDE_IMPAR*

```
//Intercambio de Elementos con Posiciones Simetricas
Procedure EquiDistanceInsideAscending(BLK, n, Mode)
    If Mode==1 Then
        P = Choose(Mod(n/2)==0, n/2, (n+1)/2)-1
        k = 0
        For i=0 To P
            W[k]  = BLK[i]
            If k<n-1 Then
               W[k+1]= BLK[n-i-1]
            EndIf
            k = k + 2
        EndFor
    Else
        k = -1
        For i=0 To n-1
            If Mod(i/2)==0 Then
               k = k + 1
               W[k]= BLK[i]
            Else
               W[n-k-1]= BLK[i]
            EndIf
        EndFor
    EndIf
    For i=0 To n-1
        BLK[i]=W[i]
    EndFor
EndProcedure
```

3.3.3.8 CONSISTENCIA Y ROBUSTEZ DEL MODELO

La validación del modelo debe sustentarse en consistencia y robustez, características que serán descritas en esta sección.

- *CONSISTENCIA DEL MODELO*

La consistencia del modelo será establecida mediante un instrumento de validación, de carácter determinístico, cuyo diagrama se ilustra a continuación.

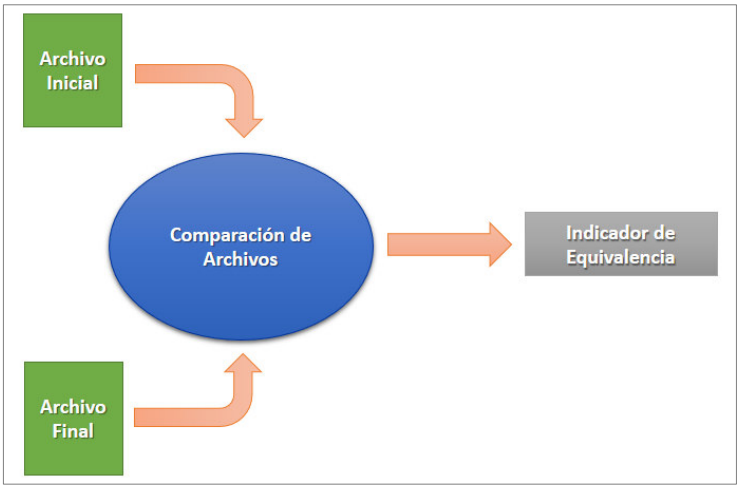

Figura 3.18 Instrumento de validación para la consistencia del modelo.
Fuente. Elaboración Propia.

Este instrumento contiene un parámetro de salida, el cual permite decidir la equivalencia de los ficheros inicial y final.

El diseño correspondiente consiste en una función algorítmica con dos parámetros de entrada representados por los ficheros inicial y final, y un parámetro de salida cuyo valor lógico determina la equivalencia de los ficheros referidos.

```
Function ConsistentModel(FileIni, FileEnd)
   Assign(FileIni, F1)
   Assign(FileEnd, F2)
   Open(F1, READ)
   Open(F2, READ)
   Sw=BinaryCompare(F1, F2)
   Close(F1)
   Close(F2)
   Return Sw
End
```

- *ROBUSTEZ DEL MODELO*

Para que el modelo sea robusto, debe adoptar las siguientes consideraciones:

- Debe soportar cualquier tipología de información, garantizando su funcionamiento y adaptabilidad a diferentes formatos de almacenamiento.
- Debe adoptar una aceptable eficiencia de procesamiento.

3.3.4 CONSTRUCCIÓN DEL MODELO

En esta sección se exponen de manera restringida, únicamente algunos elementos que corresponden a la implementación computacional del criptosistema, cuyo análisis y diseño fueros descritas en las fases anteriores.

- *ESPECIFICACIÓN DE ALFABETO*

El alfabeto se implementa computacionalmente mediante una estructura vectorial estática, esto es, de tamaño fijo igual a 256, cuya definición se muestra a continuación.

```
public static byte ALPHABET[] = new byte[256];
```

El contenido del alfabeto se establece en el siguiente código

```
public static void Create_Alphabet() {
    for(int i=0;i<=255;i++) {
        ALPHABET[i]=(byte)(i);
    }
}
```

- *GENERACIÓN DE LA CLAVE*

Se generan múltiples ocurrencias de clave simétrica en el fichero “KEY.TXT“ por un procedimiento que maximiza la varianza.

El fichero KEY.TXT se muestra a continuación:

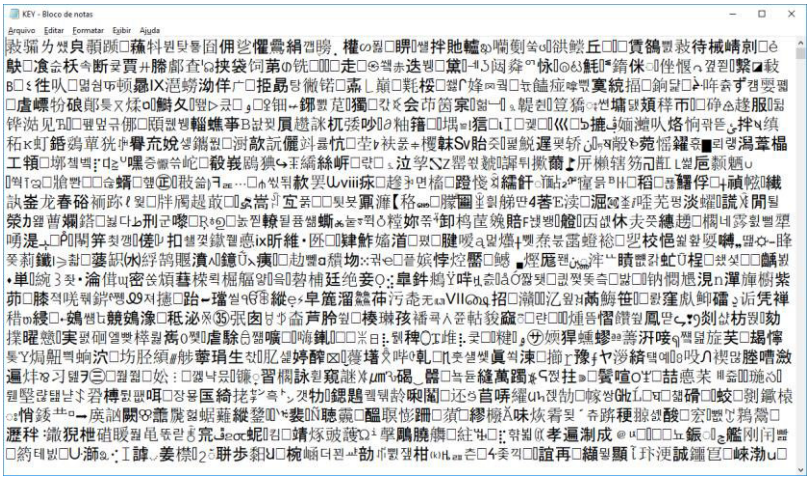

Figura 3.16 Fichero de Claves Generado. Fuente: Elaboración Propia.

En el siguiente código se efectúan 1000 simulaciones para determinar una sola ocurrencia de la clave que contiene la varianza máxima.

```
private void Generate_Key(String FileKey) {
byte TMP[] = new byte[NN];
byte KEY[] = new byte[NN];
double S,E,V,VMax;
FileOutputStream FOS = null;
BufferedOutputStream BOS = null;
    try {
        FOS = new FileOutputStream (FileKey);
        BOS = new BufferedOutputStream(FOS);
        for(int k=1;k<=MM;k++) {
            VMax = 0;
            for(int j=1;j<=1000;j++) {
                for(int i=0;i<=NN-1;i++) {
                    TMP[i]=(byte)(Math.random()*256);
                }
                S=0;
                for(int i=0;i<=NN-1;i++) {
                    S = S + (int)(TMP[i]);
                }
                E=S/NN;
                S=0;
                for(int i=0;i<=NN-1;i++) {
                    S = S + Math.pow(E - (int)(TMP[i]), 2);
                }
                V=S/NN;
                if(VMax < V) {
                   VMax = V;
                   for(int i=0;i<=NN-1;i++) {
                       KEY[i] = TMP[i];
                   }
                }
            }
            BOS.write(KEY,0,NN);
        }
    }
    catch (Exception EE) {
    }
    finally {
        try {
          BOS.close();
        }
        catch (Exception EE) {
        }
     }
}
```

- *OBTENCIÓN DE LA VECINDAD*

Sea T la cantidad de bloques de información existentes.

El siguiente código

```
private int ObtainNeighborhood(int XX[], int k, int TT) {
int AA, BB, WW;
    if(TT==1) {
        AA = XX[0];   BB = XX[0];
        WW = XX[0];
    }
    else {
        if(TT==2) {
            if(k==0) {
                AA = XX[1];   BB = XX[1];
            }
            else {
                AA = XX[0];   BB = XX[0];
            }
            WW = XX[k];
        }
        else {
            if(k==0) {
                AA = XX[TT-1];   BB = XX[1];
            }
            else {
                if(k==(TT-1)) {
                   AA = XX[TT-2];   BB = XX[0];
                }
                else {
                   AA = XX[k-1];    BB = XX[k+1];
                }
            }
            WW = XX[k];
        }
    }
    return 100*AA + 10*WW + BB;
}
```

genera un número entero de 3 dígitos decimales correspondiente a la vecindad actual. Este número será procesado en la rutina *ActionProcess*.

- *PROCESAR VECINDAD*

```
private void ActionProcess(byte VV[], int SW, int n, int EE) {
Funciones FNC = new Funciones();
    if(EE<100) {
        switch(SW) {
                case  1: FNC.RotaryRight(VV,n,7); break;
                case -1: FNC.RotaryLeft(VV,n,7); break;
                default:
        }
    }
    else {
        if(EE<200) {
            switch(SW) {
                    case  1: FNC.RotaryLeft(VV,n,7); break;
                    case -1: FNC.RotaryRight(VV,n,7); break;
                    default:
            }
        }
        else {
            if(EE<300) {
                FNC.JoinImparPar(VV,n,SW);
            }
            else {
                if(EE<400) {
                    FNC.JoinParImpar(VV,n,SW);
                }
                else {
                    if(EE<500) {
                        FNC.TransParImpar_Ascending(VV,n,SW);
                    }
                    else {
                        if(EE<600) {
                            FNC.TransParImpar_Descending(VV,n,SW);
                        }
                        else {
                            if(EE<700) {
                                FNC.EquiDistance_Inside_Impar(VV,n,SW);
                            }
                            else {
                                if(EE<800) {
                                    FNC.EquiDistance_Inside_Par(VV,n,SW);
                                }
                                else {
                                    if(EE<900) {
                                        FNC.EquiDistance_Outside_Impar(VV,n,SW);
                                    }
                                    else {
                                        FNC.EquiDistance_Outside_Par(VV,n,SW);
                                    }
                                }
                            }
                        }
                    }
                }
            }
        }
    }
}
```

3.3.5 SIMULACIÓN Y TESTING

Las pruebas realizadas se aplicaron sobre un total de 1000 ficheros.

- **PROCESO DE RECOLECCIÓN DE DATOS**

La simulación de los procesos de muestreo tomando en consideración las especificaciones señaladas, genera los siguientes elementos constituyentes de la muestra

- TEXTO

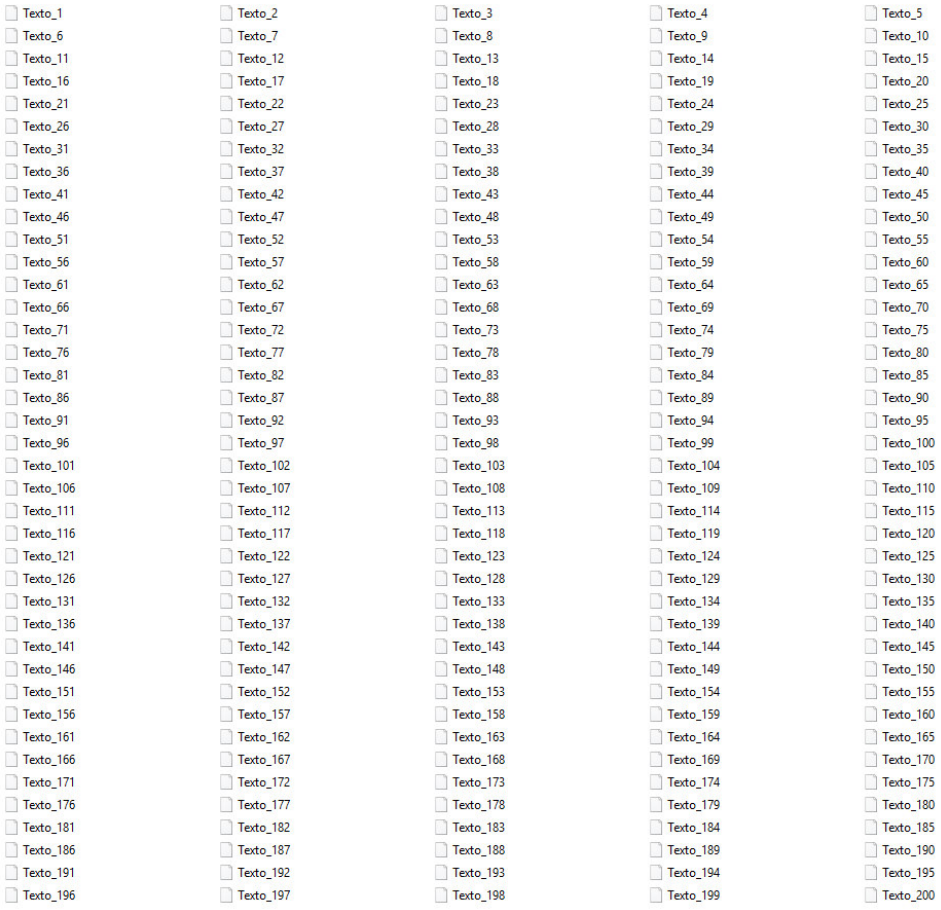

Figura 3.20 Muestra de Ficheros Texto. Fuente: Elaboración Propia.

- IMÁGEN

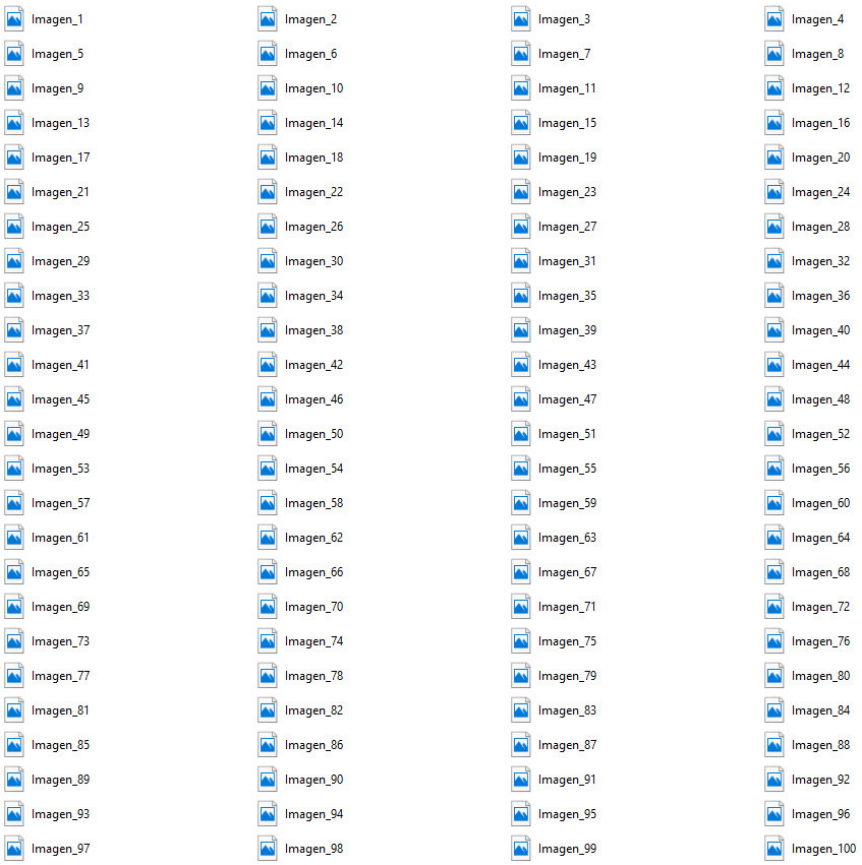

Figura 3.21 Muestra de Ficheros Imagen. Fuente: Elaboración Propia.

- AUDIO

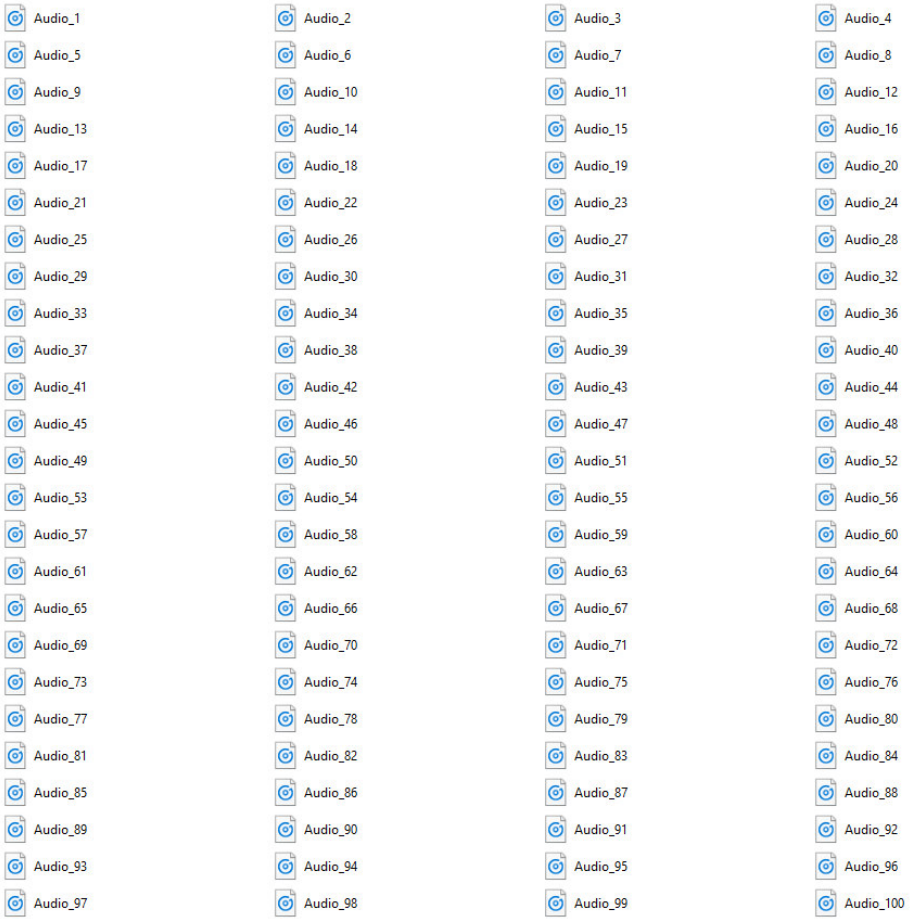

Figura 3.22 Muestra de Ficheros Audio. Fuente: Elaboración Propia.

- VIDEO

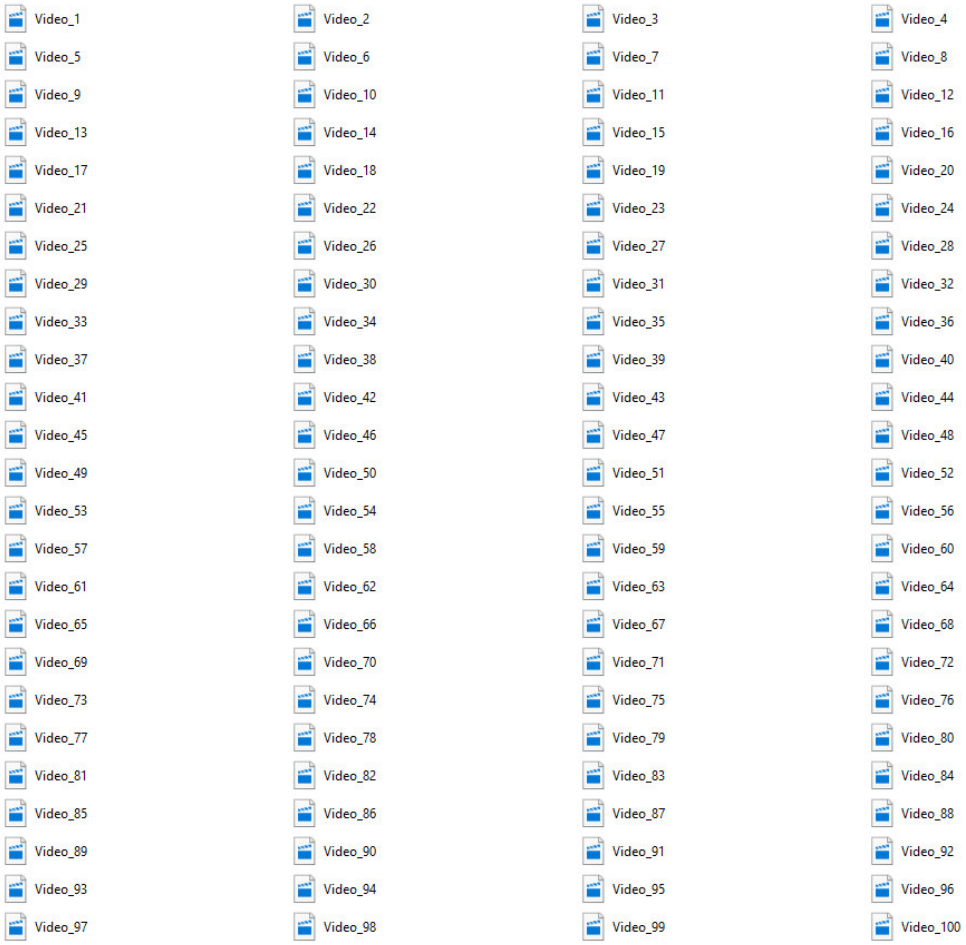

Figura 3.23 Muestra de Ficheros Video. Fuente: Elaboración Propia.

- DOCUMENTOS

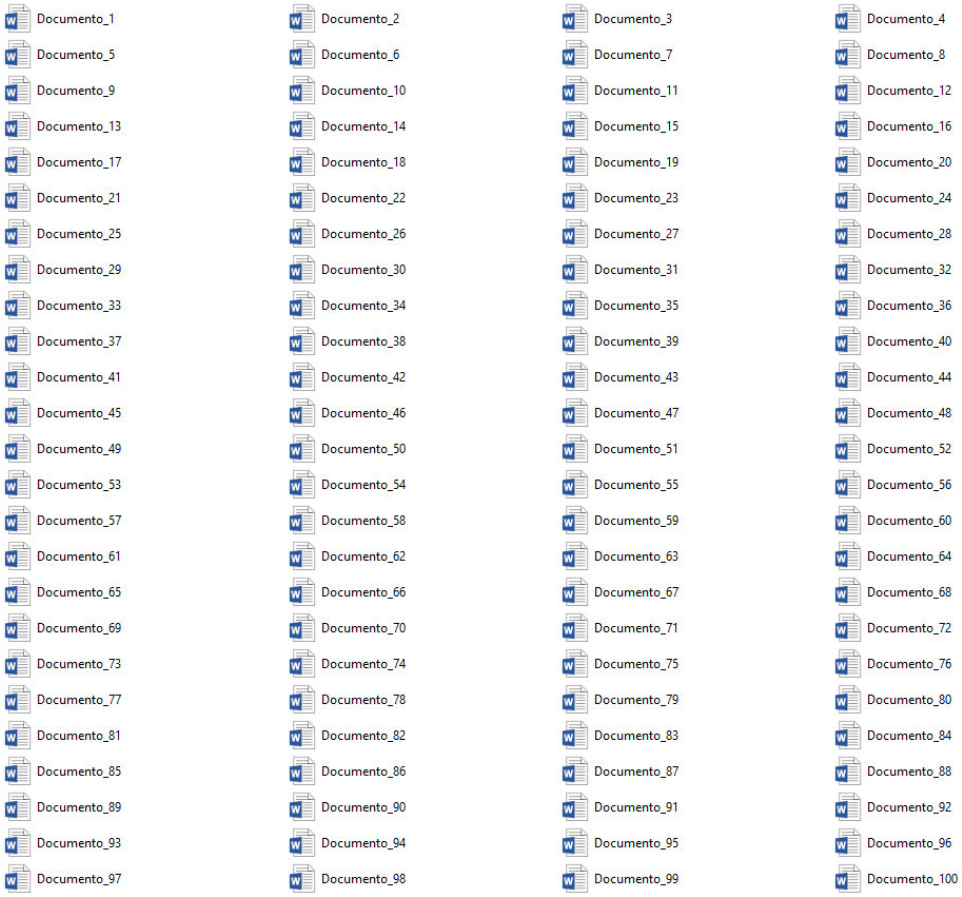

Figura 3.24 Muestra de Ficheros Documento. Fuente: Elaboración Propia.

- HOJA DE CÁLCULO

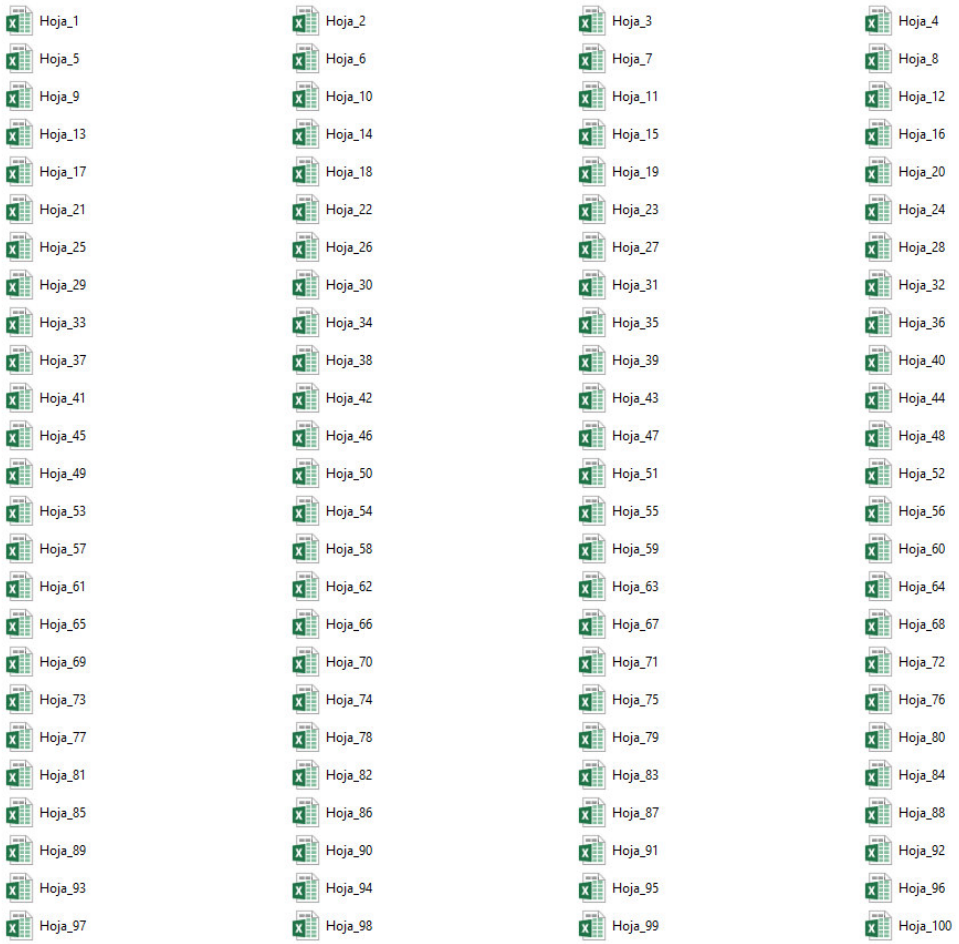

Figura 3.25 Muestra de Ficheros Hoja de Cálculo. Fuente: Elaboración Propia.

- DIAPOSITIVAS

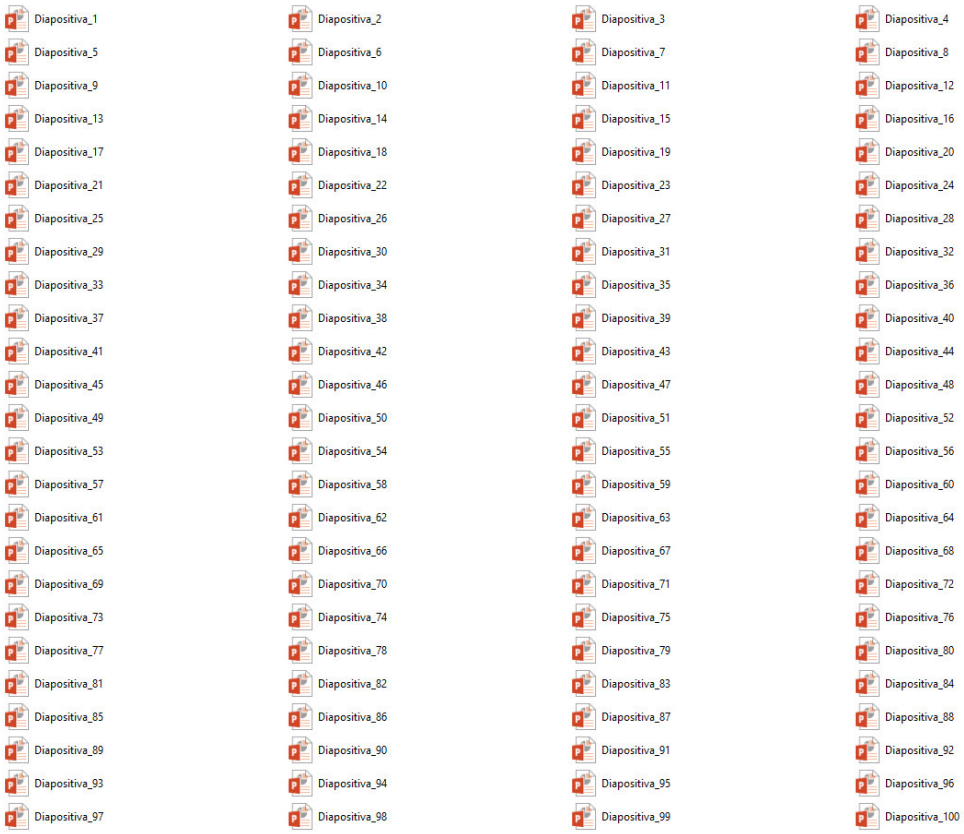

Figura 3.26 Muestra de Ficheros Diapositiva. Fuente: Elaboración Propia.

- SCRIPTS Y PROGRAMAS

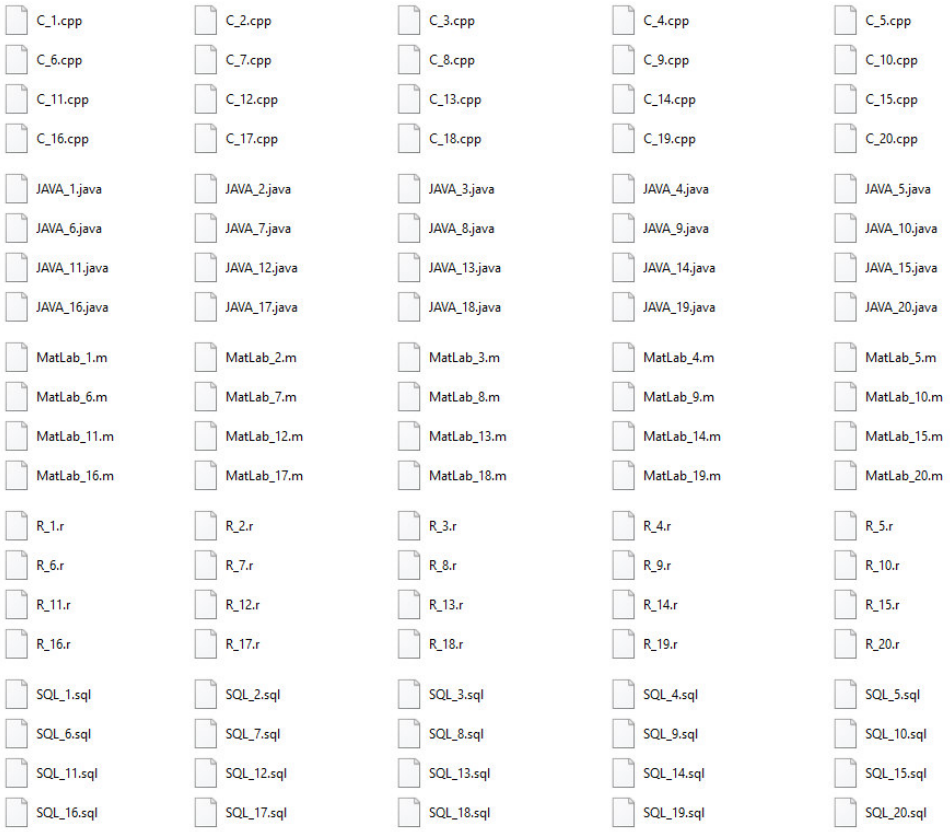

Figura 3.27 Muestra de Ficheros Script y Programas. Fuente: Elaboración Propia.

- OTROS FORMATOS

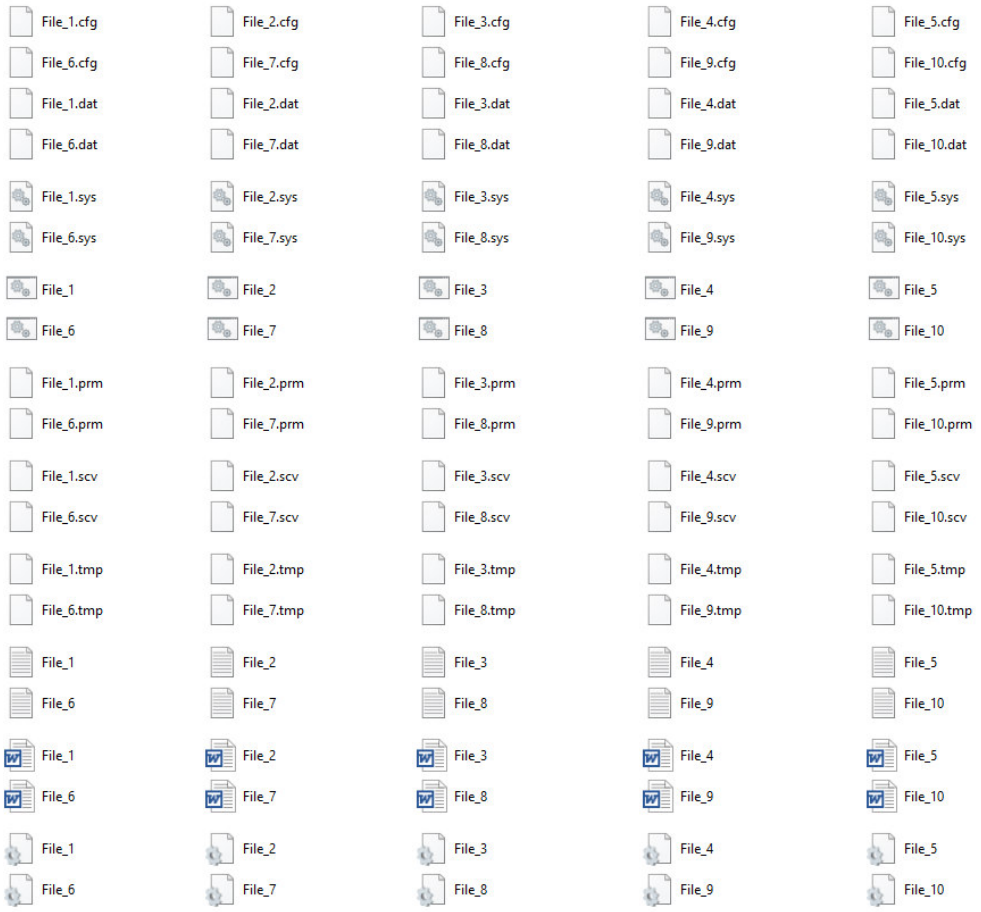

Figura 3.28 Muestra de Ficheros de otros formatos. Fuente: Elaboración Propia.

- **PROCESO DE CIFRADO Y DESCIFRADO**

Las siguientes pantallas ilustran el proceso correspondiente a ficheros Texto y de Imagen respectivamente, en la cual se evidencia la equivalencia de los ficheros Inicial y Final.

Nro	LONGITUD	TIPO	COMPRESION	EQUIVALENCIA
1	3568	Texto	0	Si
2	631	Texto	0	Si
3	2699	Texto	0	Si
4	3411	Texto	0	Si
5	1579	Texto	0	Si
6	2635	Texto	0	Si
7	1031	Texto	0	Si
8	3865	Texto	0	Si
9	3181	Texto	0	Si
10	3914	Texto	0	Si
11	1694	Texto	0	Si
12	914	Texto	0	Si
13	3854	Texto	0	Si
14	2701	Texto	0	Si
15	2328	Texto	0	Si
16	2874	Texto	0	Si
17	429	Texto	0	Si
18	2564	Texto	0	Si
19	389	Texto	0	Si
20	2570	Texto	0	Si

Figura 3.29 Procesamiento de ficheros de texto. Fuente: Elaboración Propia.

Nro	LONGITUD	TIPO	COMPRESION	EQUIVALENCIA
1	5103	Imagen	0	Si
2	1304	Imagen	0	Si
3	7028	Imagen	0	Si
4	1295	Imagen	0	Si
5	9628	Imagen	0	Si
6	1466	Imagen	0	Si
7	8702	Imagen	0	Si
8	10828	Imagen	0	Si
9	7847	Imagen	0	Si
10	7947	Imagen	0	Si
11	10608	Imagen	0	Si
12	3543	Imagen	0	Si
13	7795	Imagen	0	Si
14	2901	Imagen	0	Si
15	2240	Imagen	0	Si
16	2182	Imagen	0	Si
17	3516	Imagen	0	Si
18	4648	Imagen	0	Si
19	5208	Imagen	0	Si
20	9256	Imagen	0	Si

Figura 3.30 Procesamiento de ficheros de imagen. Fuente: Elaboración Propia.

- **TRANSICIÓN DE ESTADOS DEL AUTÓMATA**

Los estados representan el dominio de valores que adopta cada célula componente del autómata y representan un sistema dinámico complejo que evoluciona a pasos discretos.

En este caso, se ha considerado el tamaño de la clave a 1024 bytes.

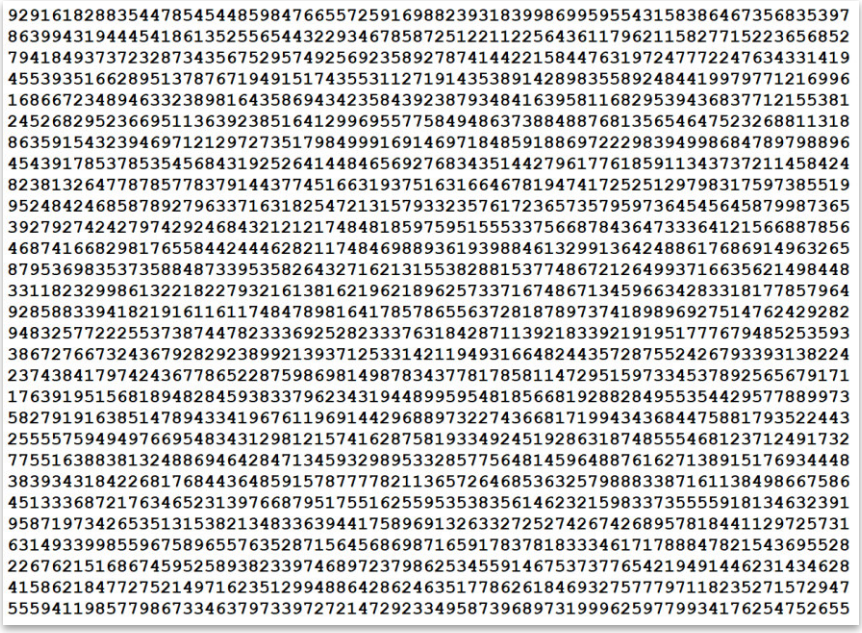

Figura 3.31 Evolución del autómata celular. Fuente: Elaboración Propia.

3.3.6 VALIDACIÓN DE RESULTADOS

Los resultados son validados mediante el instrumento diseñado descrito en las fases anteriores.

- **Presentación de Resultados**

Tal como fue indicado, la validación de los resultados es determinado por el instrumento de validación, que constituye una herramienta de decisión, que validará la equivalencia de los ficheros inicial y final.

El siguiente cuadro presenta los resultados obtenidos de un total de 1000 ficheros que representan las unidades de análisis de la presente investigación.

Tabla 3.7 Resultados del proceso de una muestra de ficheros. Fuente: Elaboración Propia.

Tipo de Archivo de Información	Total de Unidades Procesadas	Tasa de Compresión Información Cifrada	Frecuencia Relativa de Error de Información Descifrada	Tasa de Equivalencia de Información Descifrada
Texto	200	0%	0	100%
Imagen (*Todos los Formatos*)	100	0%	0	100%
Audio (*Todos los Formatos*)	100	0%	0	100%
Video (*Todos los Formatos*)	100	0%	0	100%
Documento Word	100	0%	0	100%
Hoja de Cálculo	100	0%	0	100%
Diapositivas	100	0%	0	100%
Scripts y Programas de Cálculo Científico	100	0%	0	100%
Otros Formatos	100	0%	0	100%

Dicho cuadro proporciona el resultado del procesamiento del criptosistema, aplicado sobre una muestra de 1000 ficheros entre diferentes tipologías y formatos.

Los ficheros de información fueron obtenidos de unidades de disco externo de diferentes computadores.

Los indicadores que son considerados son los siguientes:

- *Tipo de Fichero*, Que representa la categoría de la unidad de información
- *Total de Unidades Procesadas*, Cantidad de unidades por tipo de fichero
- *Tasa de Compresión*, se refiere a la posible variación resultante durante el proceso de cifrado.
- *Frecuencia Relativa de Error*, Es una variable que contiene la proporción de diferencias existentes entre los bytes del fichero original y los del fichero descifrado, respecto al total de bytes del fichero.
- *Tasa de Equivalencia de Información*, es el caso contrario del punto anterior. Consiste en determinar la proporción de igualdades existentes entre los bytes del fichero original y los del fichero descifrado, respecto al total de bytes del fichero.

3.3.7 IMPLANTACIÓN Y PUESTA EN MARCHA

La implantación del modelo y su puesta en marcha, constituye en otras palabras, su conformidad y adaptación a un determinado escenario real.

En el Anexo 1, se adjunta una Propuesta de Solución Técnica para la Facultad de Ciencias Matemáticas [19] de la Universidad Nacional Mayor de San Marcos consistente en el *Desarrollo e Implementación de un Sistema de Protección y Seguridad de la Información* orientada a garantizar la confidencialidad y reserva de la información.

Esta solución técnica representa la aplicación de la investigación a un escenario real.

19 Esta solución técnica puede ser extendida a todas las facultades y dependencias de la Universidad.

Capítulo 4

RESULTADOS Y DISCUSIÓN

4.1 ANÁLISIS, INTERPRETACIÓN Y DISCUSIÓN DE RESULTADOS

El desarrollo de la investigación tiene el propósito de diseñar e implementar un criptosistema, donde el resultado final es codificar y decodificar los ficheros de datos utilizando una determinada clave simétrica, utilizando posteriormente un instrumento de validación de resultados. No obstante, es importante señalar que existen varios resultados de carácter intermedio en cada fase correspondiente a la metodología utilizada.

Las siguientes secciones representan categorías que permiten describir los resultados de la investigación.

4.1.1 PROCESOS DE SIMULACIÓN COMPUTACIONAL

Los procesos de simulación fueron aplicados en cada subproblema y pueden ser interpretados como una experimentación de los modelos asociados.

Los procesos de simulación son de gran utilidad para resolver problemas y garantizan su puesta en marcha.

4.1.2 DETERMINACIÓN DE LA POBLACIÓN

La población bajo estudio fue determinada por un proceso automatizado que hace uso de los recursos del sistema operacional subyacente.

Desde el punto de vista técnico, es un proceso fácil de implementar.

4.1.3 OBTENCIÓN DE LA MUESTRA REPRESENTATIVA

La muestra representativa fue obtenida por conveniencia involucrando formatos de archivo más representativos en categorías. La selección en cada categoría fue probabilística.

Por las características de la investigación, el muestreo puede ser alterado de acuerdo a criterios o aspectos técnicos de las unidades de almacenamiento en la organización.

Para su implementación, necesariamente se requiere el uso de paquetes estadísticos o herramientas de desarrollo de programas.

4.1.4 GENERACIÓN DE CLAVE SIMETRICA

La clave simétrica se obtiene mediante un procedimiento que maximiza la varianza a partir de un conjunto de iteraciones. Tiene su máximo punto de eficiencia cuando su varianza es alta.

Para obtener una clave simétrica razonable se deben realiza simulaciones haciendo uso de paquetes o herramientas de desarrollo de programas.

4.1.5 RIESGOS ASOCIADOS AL CRIPTOANÁLISIS

La adopción de una clave simétrica de 1024 bytes equivalente a 8192 bits, garantiza la robustez y potencia del criptosistema minimizando el riesgo de cualquier procedimiento de criptoanálisis que pretenda efectuar el descifrado de la información.

4.1.6 FORMATOS Y TIPOLOGIA DE INFORMACIÓN

Los resultados obtenidos corresponden a los formatos y tipos de ficheros establecidos en la muestra representativa. Sin embargo, cualquier formato de información debe ser soportado por el criptosistema.

El criptosistema diseñado permite cifrar la información correspondiente a cualquier tipo de información puesto que realiza el procesamiento de cada fichero como unidad funcional realizando el tratamiento a nivel de bytes.

4.2 PRUEBAS DE HIPÓTESIS

En esta sección se consideran los siguientes aspectos las cuales tienen correspondencia.

4.2.1 DESCRIPCIÓN DE LA PRUEBA ESTADÍSTICA

La criptografía simétrica garantiza el cifrado y descifrado de la información usando la misma clave. No obstante, el autómata celular implementado es una representación matemática y computacional que evoluciona en el tiempo mediante reglas de transición especificadas en el diseño e implementadas en el código fuente.

Por tanto, se requiere probar si fue posible efectuar el cifrado simétrico de la información. Esto es posible mediante la comparación de los ficheros originales de la muestra que representan las entradas al criptosistema, respecto a su correspondiente fichero de descifrado que constituyen la salida.

La comparación de los ficheros se va a determinar mediante un instrumento de validación, el cual va a determinar la validez del modelo.

4.2.2 DETERMINACIÓN DE LA PROBABILIDAD

La comparación de un fichero original y su respectivo fichero descifrado, puede ser enfocada como un ensayo de Bernoulli denominado también experimento de *éxito* o *fracaso* que pueden ser definidos como sigue:

- ***Éxito*** : Los ficheros original y descifrado son iguales
- ***Fracaso*** : Los ficheros original y descifrado son diferentes

En este caso interesa determinar la probabilidad del *éxito* o del *fracaso*. Tomando como referencia el cuadro anterior y usando la ley de los grandes números en la cual se establece que "*cuando un experimento se realiza un número infinito de veces, la frecuencia relativa de la variable se aproxima a su probabilidad*".

Luego, se puede definir la probabilidad del siguiente modo:

$$P(x_i) = \begin{cases} 1, & \text{Exito} \\ 0, & \text{Fracaso} \end{cases}$$

Donde x_i es una variable aleatoria de Bernoulli, y $\{x_i\}$ representa la muestra seleccionada, $i = 1,2,3, \ldots . n$

Una muestra o sucesión de n variables aleatorias de Bernoulli forma una muestra aleatoria Binomial con parámetros p y n.

4.2.3 TOTAL DE LA MUESTRA

Los ficheros de información representan las unidades de análisis de la investigación. El total de ficheros de información que van ser procesados es $n = 1000$, y constituye el tamaño de la muestra.

El siguiente grafico de barras ilustra los tipos de información procesados que representan las unidades de análisis.

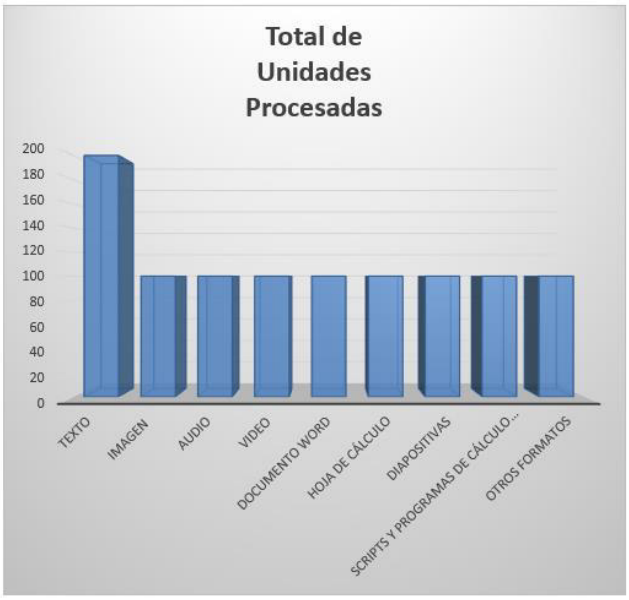

Figura 4.1 Tipología de Ficheros Procesados. Fuente: Elaboración Propia.

Luego, la muestra Binomial con parámetros p=1 y $n = 1000$ se define como una secuencia de $n = 100$ variables aleatorias de Bernoulli

$$x_1, x_2, x_3, \ldots, x_{n-1}, x_n$$

Donde cada elemento $p(x_i) \in \{0,1\},\ i = 1,2,3, \ldots, n$

4.2.4 PRUEBAS DE HIPÓTESIS

La prueba de hipótesis asociada a la presente investigación, se puede formular del siguiente modo:

H_0: HIPÓTESIS NULA

No es posible realizar el modelado y simulación de un autómata celular para el tratamiento del problema del cifrado simétrico de la información.

En este caso, se requiere establecer la no equivalencia de los parámetros muestral y poblacional

$$p \neq P$$

H_1: HIPÓTESIS ALTERNATIVA

Si es posible realizar el modelado y simulación de un autómata celular para el tratamiento del problema del cifrado simétrico de la información.

Esto consiste, en establecer la igualdad de los parámetros muestral y poblacional

$$p = P$$

La siguiente tabla ilustra la validación de los ficheros procesados

Tabla 4.1 Resultados del proceso de una muestra de ficheros. Fuente: Elaboración Propia.

Tipo de Archivo de Información	Total de Unidades Procesadas	Tasa de Compresión Información Cifrada	Frecuencia Relativa de Error de Información Descifrada	Tasa de Equivalencia de Información Descifrada
Texto	200	0%	0	100%
Imagen (*Todos los Formatos*)	100	0%	0	100%
Audio (*Todos los Formatos*)	100	0%	0	100%
Video (*Todos los Formatos*)	100	0%	0	100%
Documento Word	100	0%	0	100%
Hoja de Cálculo	100	0%	0	100%
Diapositivas	100	0%	0	100%
Scripts y Programas de Cálculo Científico	100	0%	0	100%
Otros Formatos	100	0%	0	100%

Luego, del cuadro anterior, se establece que la proporción de equivalencias respecto al total, esto es, la frecuencia relativa de equivalencias es 1. En consecuencia, generalizando de acuerdo a la *Ley de los grandes números*, se rechaza la hipótesis H_0.

Por lo tanto, el contraste de hipótesis acepta H_1.

El problema de la investigación requiere que todos los ficheros sean equivalentes a su correspondiente descifrado, aun cuando el total n sea un número absolutamente grande.

Esto significa que, en el peor caso, si p< 1, entonces no se aceptará la solución planteada, invalidando por consiguiente la solución.

4.3 PRESENTACIÓN DE RESULTADOS

Las pruebas realizadas se aplicaron sobre un total de 1000 ficheros con diferentes formatos o tipologías de información.

- **INFORMACIÓN TEXTO**

TEXTO ORIGINAL

El continuo desarrollo tecnologico ha generado impacto en la sociedad y en las organizaciones, propiciando cambios profundos en sus procesos, operaciones y sistemas.

Uno de los grandes desafíos en las organizaciones es la seguridad y control de acceso a la información y su disponibilidad de manera íntegra, veraz y oportuna como base fundamental para la óptima toma de decisiones.

La información representa un importante activo en las organizaciones constituyéndose como un factor crítico de éxito para afianzar su posicionamiento y ventaja competitiva. Sin embargo, requiere el desarrollo e implementación de mecanismos y procedimientos que garanticen su protección y seguridad.

El problema del cifrado de la información desde una óptica de los sistemas dinámicos complejos utilizando modelos de simulación de tiempo discreto basado en autómatas celulares, representa el objeto de estudio de la presente investigación.

Para tal efecto, el conjunto de archivos que se encuentran localizados en repositorios o unidades de almacenamiento externo en computadoras personales o servidores organizacionales constituye el escenario de aplicación.

En los sistemas criptográficos, la privacidad, confidencialidad y reserva de la información es de gran novedad y absoluto interés adoptando un grado de alta importancia, prioridad y viabilidad, constituyéndose como la base fundamental de la seguridad y protección de la información en los sistemas organizacionales.

Desde épocas remotas, el hombre ha desarrollado técnicas y procedimientos orientados a la protección y seguridad de la información de modo que garanticen su reserva, privacidad y confidencialidad.

La información representa el insumo principal que genera conocimiento, base fundamental para la óptima toma de decisiones. De este modo la información adopta un rol preponderante constituyéndose en un factor crítico de éxito en

Figura 4.1 Texto Original. Fuente: Elaboración Propia.

TEXTO CIFRADO

Figura 4.2 Texto Cifrado. Fuente: Elaboración Propia.

- **INFORMACIÓN IMAGEN**

IMAGEN ORIGINAL

Figura 4.3 Imagen Original. Fuente: Elaboración Propia.

IMAGEN CIFRADA

Figura 4.4 Imagen Cifrada. Fuente: Elaboración Propia.

IMAGEN ORIGINAL

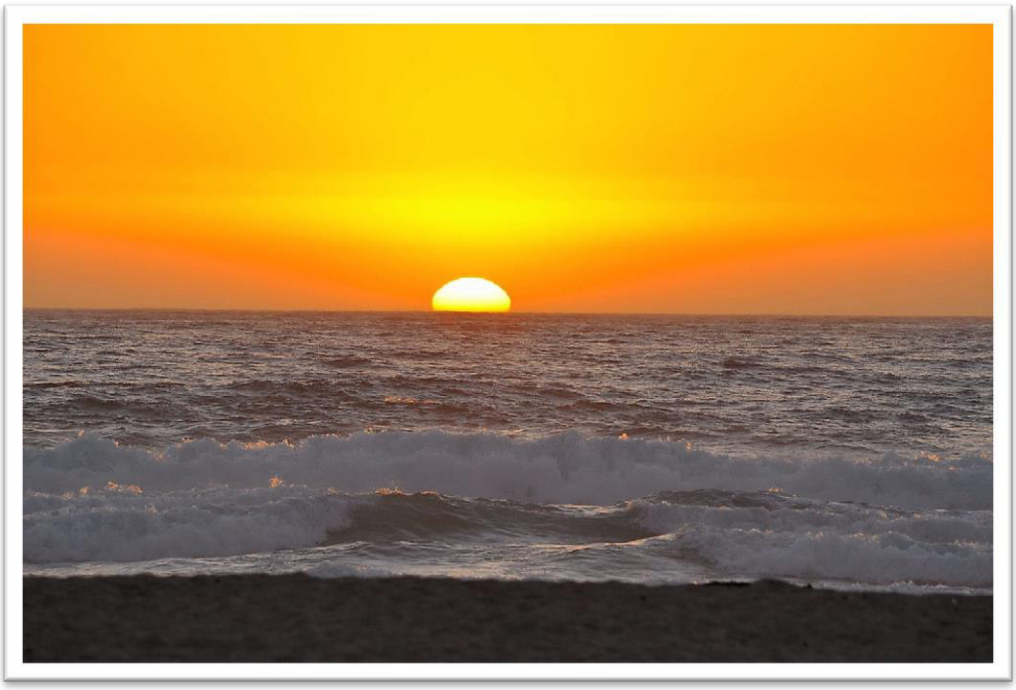

Figura 4.5 Imagen Original. Fuente: Elaboración Propia.

IMAGEN CIFRADA

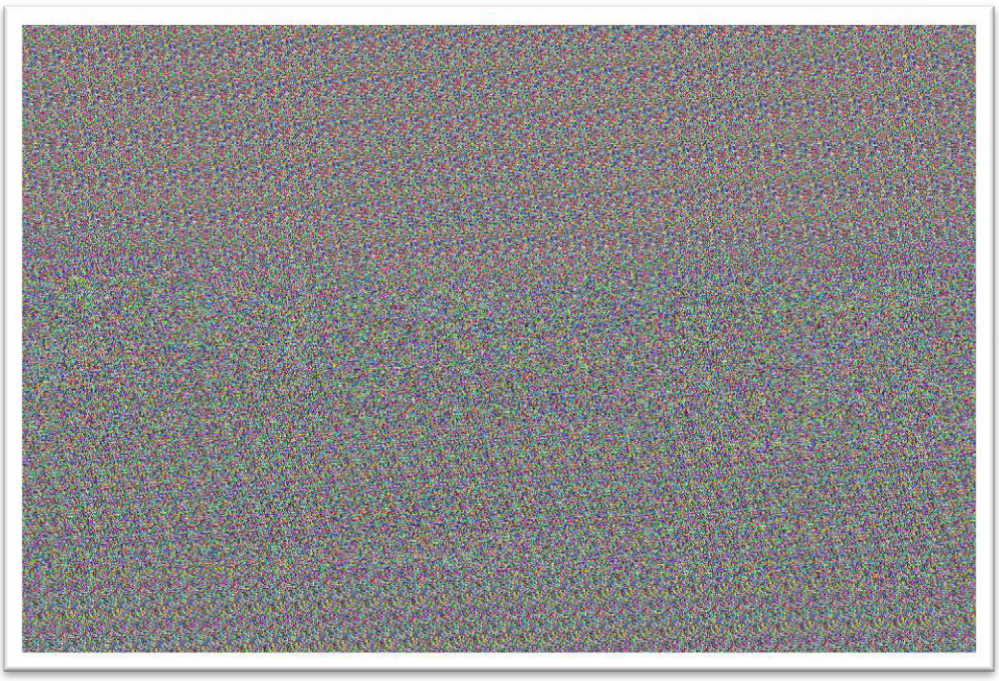

Figura 4.6 Imagen Cifrada. Fuente: Elaboración Propia.

IMAGEN ORIGINAL

Figura 4.7 Imagen Original. Fuente: Elaboración Propia.

IMAGEN CIFRADA

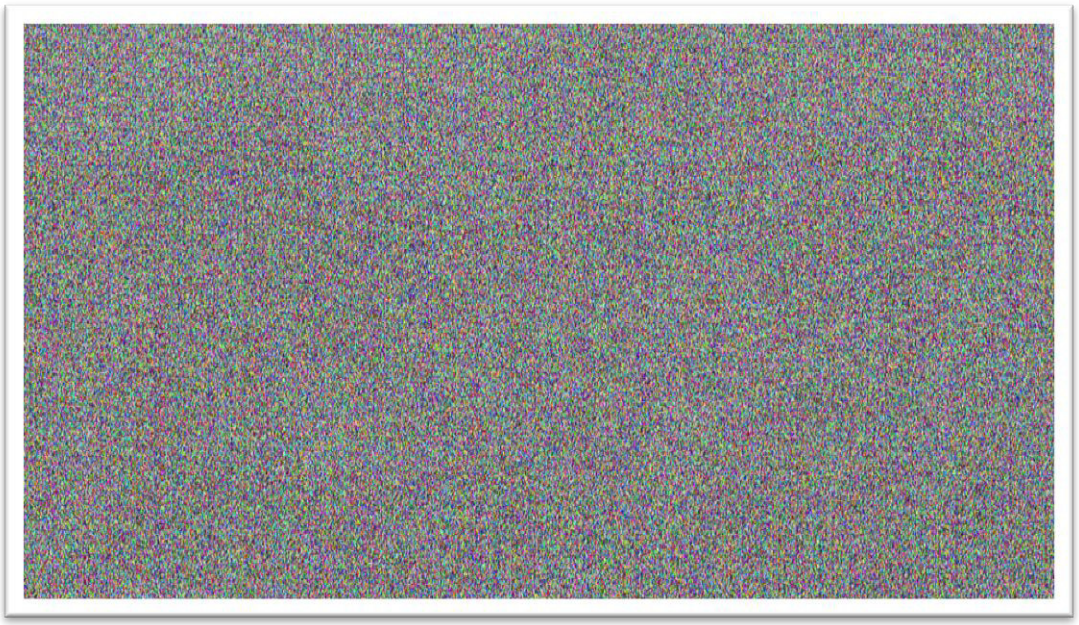

Figura 4.8 Imagen Cifrada. Fuente: Elaboración Propia.

Capítulo 5

IMPACTOS

5.1 PROPUESTA PARA LA SOLUCIÓN DEL PROBLEMA

El presente documento incorpora como un valor agregado al desarrollo de la investigación una Propuesta consistente en un *Sistema de Gestión de Confidencialidad y Reserva de Archivos de Información para la Universidad Nacional Mayor de San Marcos* (Proyecto SGCRAI-UNMSM) [20].

Esta Propuesta permitirá generar la solución del problema general de la investigación, y con aplicación a todas las áreas o unidades orgánicas de la UNMSM tales como facultades o dependencias.

La Propuesta tiene costo CERO y contiene los siguientes aspectos:

- Periodo y lugar de ejecución
- Descripción y alcance del servicio
- Garantía del producto
- Especificaciones Técnicas
- Cronograma de actividades

5.2 COSTOS DE IMPLEMENTACIÓN DE LA PROPUESTA

La implementación de la Propuesta tiene costo CERO. Para tal efecto, la gestión del Proyecto de desarrollo e implementación de la Propuesta requiere ser gestionado por el autor, con el apoyo de los recursos humanos y tecnológicos que establezca la UNMSM[21].

---

20 La Propuesta Técnica-Económica del Proyecto SGCRAI-UNMSM es descrita brevemente en la sección Anexo 1 de este documento.

21 La UNMSM como entidad interesada, deberá efectuar de manera oportuna las coordinaciones con el autor, para llevar a cabo la ejecución del Proyecto.

5.3 BENEFICIOS QUE APORTA LA PROPUESTA

La ejecución del Proyecto SGCRAI-UNMSM va a generar beneficios, entre las cuales vale citar:

- *Aporte a la sociedad de una cultura de confidencialidad y reserva de la información*.

  El fenómeno BigData es una actual tendencia tecnológica caracterizada por la gestión y tratamiento de grandes volúmenes de información provenientes de fuentes masivas de datos tales como blogs o redes sociales, propiciando un real escenario, para establecer y garantizar la confidencialidad y reserva de la información.

- *Incremento del rendimiento y productividad en los procesos, operaciones y sistemas organizacionales*.

  La garantía de confidencialidad y reserva de la información, minimiza los riesgos que vulneran su protección y seguridad, incrementando la confianza de los usuarios y propiciando mayor disponibilidad de tiempo y esfuerzo en las operaciones y procesos organizacionales.

- *Organización Sistemática de la Información.*

  La obtención de los beneficios establecidos en los puntos anteriores permitirá organizar la información de manera ordenada y sistemática.

CONCLUSIONES

Las siguientes conclusiones fueron obtenidas como producto del desarrollo de la investigación:

- **La simulación computacional es una poderosa herramienta experimental**

  El desarrollo de la investigación fue posible mediante el uso de esta poderosa herramienta de la investigación de operaciones. La simulación permite experimentar con datos provenientes de una muestra, anticipándose de este modo a su posterior adaptación a un escenario real y reduciendo de manera significativa los costos.

- **Autómatas celulares como modelos de sistemas dinámicos complejos de tiempo discreto**

  La investigación incorporó modelos basados en autómatas celulares para el desarrollo del criptosistema simétrico. Los modelos basados en autómatas celulares permiten representar a sistemas dinámicos complejos a partir de un conjunto discreto de reglas simples de evolución o transición de estados. Esta alternativa representa una gran ventaja respecto a la utilización de modelos matemáticos que incorporan ecuaciones diferenciales, complicando y haciendo algunas veces imposible su tratamiento y resolución debido al elevado costo de procesamiento que genera su implementación computacional.

- **Utilización de claves simétricas seguras**

  La adopción de una clave que posee varianza máxima entre sus elementos, constituye una excelente técnica para el desarrollo del criptosistema simétrico, incrementando la protección y seguridad de la información y minimizando los riesgos de un potencial criptoanálisis.

RECOMENDACIONES

Para establecer y garantizar la confidencialidad y reserva de la información en las organizaciones, se proponen las siguientes recomendaciones:

- **Directivas de confidencialidad y reserva de la información**

  Las organizaciones deben establecer e implantar directivas o políticas de confidencialidad y reserva de la información, específicamente aquella que adopta un valor crítico dado que su acceso no controlado pone en riesgo los planes y proyectos.

- **Almacenamiento de archivos en modo cifrado**

  Los archivos de información deben ser guardados en las unidades de almacenamiento externo, en modo cifrado mediante el uso de una clave determinada. De este modo, el acceso a la información será realizada a través de determinados controles y su lectura mediante herramientas especializadas.

- **Claves simétricas para cada archivo**

  En un escenario caracterizado por el manejo de información de carácter estrictamente personal, utilizar claves simétricas específicas para cada archivo de información del repositorio de almacenamiento externo, representa una excelente alternativa para la seguridad y protección de la información. Sin embargo, se requiere un continuo y riguroso control de las claves, generando una sobrecarga de tareas, un alto costo de mantenimiento la posible inducción al error y consecuentemente el colapso del sistema.

- **Claves simétricas por categoría**

  En otro contexto de aplicación diferente del punto anterior, donde la información de los usuarios es compartida a todo un equipo de trabajo tal como un proyecto, usar una clave simétrica para todos los archivos del mismo grupo representa una buena práctica, que se sustenta en los criterios de proximidad o similaridad.

TRABAJO FUTURO

La criptografía y el criptoanálisis de la información usando una clave simétrica son dos procesos complementarios que adoptan alta importancia y requieren continuamente incrementar la garantía de seguridad y protección de la información para minimizar el riesgo de posibles ataques externos.

Para tal efecto, se proponen las siguientes actividades futuras que permitan fortalecer este trabajo de investigación:

- Extender el autómata celular elemental circular utilizado a un modelo multidimensional, periódico y estocástico.
- Generar cifrado/descifrado de la información de acuerdo a criterios de similaridad mediante el uso y aplicación de cluster.
- Incorporar un procedimiento de compactación en el proceso de cifrado de la información.

REFERENCIAS BIBLIOGRÁFICAS

[1] Baranger, Michel. (2001). *Chaos, Complexity, and Entropy. A physics talk for non-physicists*. Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics. MIT, Cambridge.

[2] Boccara, Nino. (2010). *Modeling Complex Systems*. New York, Dordrecht, Heidelberg, London: Springer.

[3] Cavada Benech, Cristián. (2007). *Aplicación de Autómatas Celulares en la Predicción del Movimiento de Precios de Bienes Raíces. Universidad de Chile*.

[4] Ferguson, Niels & Schneier, Bruce & Kohno, Tadayoshi. (2010). *Cryptography Engineering: Design Principles and Practical Applications*. Indianapolis, Indiana, USA: Wiley Publishing, Inc.

[5] Gracia, Diego. *Ciencia y Filosofía*. (2005). Madrid, Spain.

[6] Hernández Encinas, L. & Hernández Encinas, A. & Hoya White, S. & Martín del Rey, A. & Rodríguez Sánchez, G. (2004). *Cifrado de imágenes usando autómatas celulares con memoria*.

[7] Hillier, Frederick S. & Lieberman, Gerald J. (2010). *Introducción a la Investigación de Operaciones*. México, D.F.: McGraw-Hill Interamericana Editores, S.A. de C.V.

[8] Johansen Bertoglio, Oscar. (1993). *Introducción a la Teoria General de Sistemas*. México, D.F.: Editorial LIMUSA S.A. de C.V. Grupo Noriega Editores.

[9] Karafyllidis, Ioannis & Thanailakis, Adonios. (1996). *A model for predicting forest fire spreading using cellular automata*. Democritus University of Thrace, Department of Electrical and Computer Engineering, Xanthi, Greece.

[10] Kohutek, Carolina & Zanotello, Marcelo. (2009). *Autômatos Celulares e suas aplicações na Simulação de Fenômenos em Ciência dos Materiais*. Universidade Federal do ABC.

[11] Ladyman, James. (2002). *Understanding Philosophy of Science*. Routledge Taylor & Francis Group. New York, USA and London UK

[12] Ladyman, James & Lambert, James. (2012). *What is a Complex System?*. University of Bristol, U.K.

[13] Meadows, Donella H. (2008). *Thinking in Systems*. London, UK: Earthscan.

[14] Muthuswamy, Bharathwaj & Ellithorpe, Jonathan D. (2008). *A Cellular Automaton For Factoring Integers*.UC BERKELEY, EECS DEPT., SPRING 2008. TECHNICAL REPORT 1.

[15] Ortiz, Miky & Olivares, Paulo. (2009). *Investigación de Operaciones*. Editorial Macro. Lima, PERÚ

[16] Parra, Jorge & Perozo, Niriaska. (2017). *Un método cooperativo para la segmentación de imágenes basado en autómatas celulares*. Revista INGENIERÍA UC. ISSN: 1316-6832. Universidad de Carabobo. Venezuela. Revista Ingenieria UC, Vol. 24, No. 1, Abril 2017 53 – 62.

[17] Prawda Witenberg, Juan. (2004). *Métodos y Modelos de Investigación de Operaciones*. México: Limusa Noriega Editores.

[18] Ptolemaeus, Claudius. (2014). *System Design, Modeling, and Simulation using Ptolemy II*. Ptolemy.org.

[19] Rosenberg, Alex. (2005). Philosophy of Science: A contemporary introduction. Routledge Taylor & Francis Group. New York, USA and London UK.

[20] Senge, Peter M. (2005). *La Quinta Disciplina: Cómo impulsar el aprendizaje en la organización inteligente*. Ediciones Granica S.A. Buenos Aires.

[21] Seredynski, Franciszek & Bouvry, Pascal & Zomaya, Albert Y. (2003). *Secret Key Cryptography with Cellular Automata*. IPDPS International Parallel and Distributed Processing Symposium. Nice, France.

[22] Stallings, William. (2014). *Cryptography and Network Security: Principles and Practice*. Pearson Education, Inc.

[23] Taha, Hamdy A. (2012). *Investigación de Operaciones*. Naucalpan de Juárez, Estado de México: Pearson Educación de México, S.A. de C.V.

[24] Uribe, Carmenza. (2007). *Ciencia, tecnología y sociedad: Evolución y revoluciones*.

[25] Van Tilborg, Henk C.A. (2000). *Fundamentals of Cryptology*: *A Professional Reference and Interactive Tutorial.* Kluwer Academic Publishers. Boston/Dordrecht/London.

[26] Villarreal Zapata, Elena & Ordaz Salazar, Francisco Cruz. (2011). *Autómatas celulares elementales aplicados a la encriptación de datos*.

[27] Von Bertalanffy, Ludwig. (1989). *Teoría General de los Sistemas. México*: Fondo de Cultura Económica.

[28] Von Neumann, John. (1966). *Theory of Self-Reproducing Automata*.

[29] Winston, Wayne L. (2005). *Investigación de Operaciones, Aplicaciones y Algoritmos*. México: International Thomson Editores S.A.

ANEXO 1

PROPUESTA TÉCNICA - ECONOMICA

PROYECTO SGCRAI-UNMSM: Sistema de Gestión de Confidencialidad y Reserva de Archivos de Información para la Universidad Nacional Mayor de San Marcos

1.1 OBJETIVO

Establecer y garantizar la confidencialidad y reserva de los archivos de información localizados en unidades de almacenamiento externo en computadores y servidores de la Universidad Nacional Mayor de San Marcos (UNMSM), mediante el desarrollo e implementación del sistema criptográfico SGCRAI-UNMSM.

1.2 PERIODO DE EJECUCION

El periodo de realización del Proyecto SGCRAI-UNMSM está estimado en 18 semanas.

1.3 LUGAR DE EJECUCION

El Proyecto SGCRAI-UNMSM será realizado en las instalaciones de la Universidad para facilitar los procesos de supervisión y control en cada fase del desarrollo.

1.4 DESCRIPCIÓN DEL SERVICIO

El Proyecto SGCRAI-UNMSM se sustenta en los siguientes ítems:

- Instalación de los módulos ejecutables del Sistema.
- Configuración de los parámetros del Sistema.
- Transferencia y migración de datos.
- Guía de Operación para los usuarios finales.
- Manual Técnico para usuarios administradores.
- Soporte, capacitación y asistencia técnica a los usuarios

1.5 ALCANCE DEL SERVICIO

El Proyecto SGCRAI-UNMSM tiene alcance a todas las oficinas o unidades de la Universidad establecidos en su estructura orgánica, tales como facultades y dependencias.

1.6 ESPECIFICACIONES TÉCNICAS

El Sistema adoptará las siguientes especificaciones:

- Software con licencia abierta para ser instalado y configurado en cualquier área orgánica de la Universidad.
- Aplicación concurrente y multiusuario
- Sistema multiusuario de tipo Desktop, con orientación a Internet y escalable a plataforma móvil
- Documentación en línea con asistencia permanente al usuario.
- Base de datos portable y escalable a plataforma Cloud.
- Mecanismos de control y seguridad de acceso a la información

1.7 GARANTÍA DEL SERVICIO

La garantía del servicio será permanente. La administración del Sistema será realizada por la Universidad.

1.8 COSTOS

Costo CERO para su implantación en la Universidad la cual incluye todas sus facultades y dependencias.

1.9 CRONOGRAMA DE ACTIVIDADES

El calendario de tareas asociado al Proyecto SGCRAI-UNMSM se describe a continuación.