Temas 6.1.3 y 6.1.4
Búsquedas y Seguridad

Integrantes:
Díaz Hernández Braulio
González Escamilla Felipe Yael
Islas Gasca Ian Jair

Introducción

En la era digital actual, la seguridad informática y la capacidad de buscar información de manera eficiente son fundamentales para el funcionamiento de la sociedad moderna. Desde los primeros buscadores web hasta los avanzados protocolos de seguridad, la tecnología ha evolucionado considerablemente para satisfacer las crecientes demandas de usuarios y organizaciones. Este análisis explora la evolución de las herramientas de búsqueda en la web, destacando hitos importantes y arquitecturas fundamentales, así como los complejos mecanismos de seguridad como IPSec y SSL que protegen las comunicaciones y datos en internet. Además, se aborda la necesidad de estrategias de seguridad robustas para mitigar las amenazas crecientes en el entorno digital.un entorno digital seguro y eficiente para todos.

Búsqueda en la Web

Historia de las herramientas de búsqueda en la Web:

I. En 1990, apareció el primer buscador llamado Archie, que se conectaba a servidores a través de FTP.

I. Poco después, en 1991, surgió Gopher, el primer sistema de almacenamiento y recuperación de información a través de Internet.

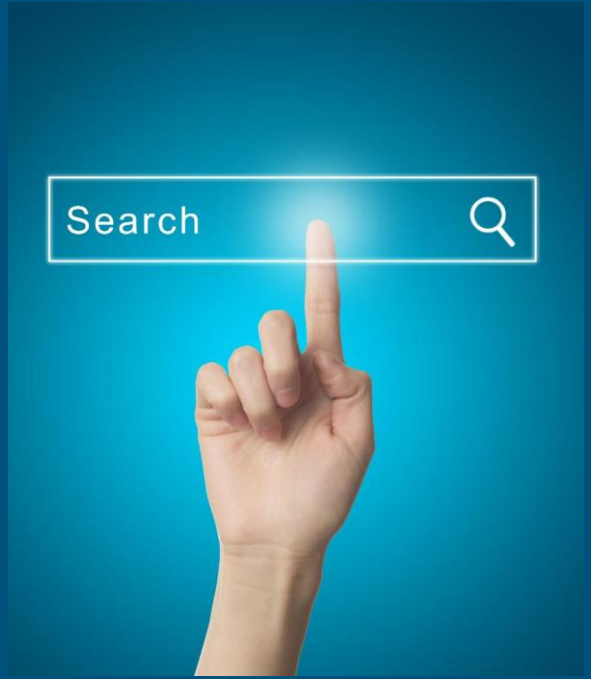
I. En 1993, Tim Berners-Lee creó el primer sitio web, seguido en el mismo año por el primer robot web llamado World Wide Web Wanderer.

I. El primer buscador web fue Aliweb en 1994, seguido por otros como Galaxy, WebCrawler y Lycos en 1994.

I. En 1995, AltaVista ofrecía búsqueda de documentos multimedia, y en 1996, Larry Page y Sergey Brin comenzaron a desarrollar BackRub, que luego se convirtió en Google en 2000.

Herramientas de búsqueda

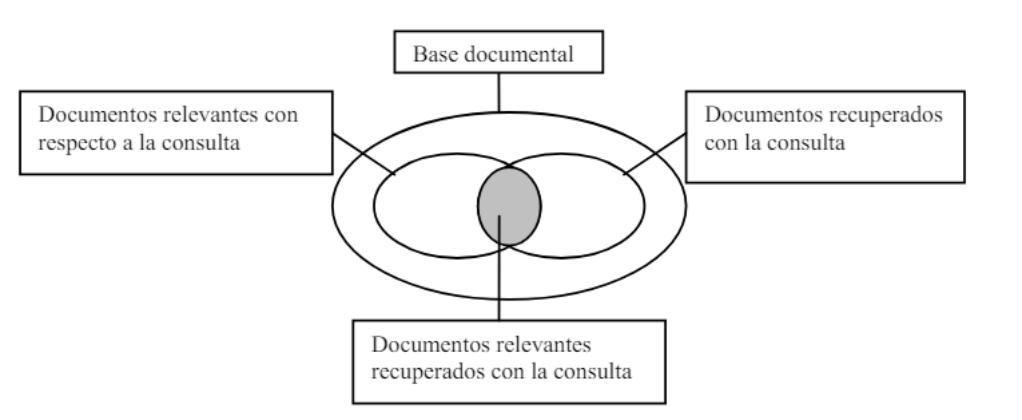
- Los **buscadores** como Google, Altavista, Lycos y Bing son sistemas de recuperación de información que vincula documentos web sin seguir una estructura jerárquica, este tipo de sistemas abarca un mayor número de documentos que los directorios debido al proceso de automatización, además, cada cierto tiempo se comprueba si el documento referenciado no ha sufrido modificaciones.



Para medir la capacidad de un sistema de recuperación de información se suelen usar una serie de métricas estándar. La exhaustividad (recall) y la precisión:

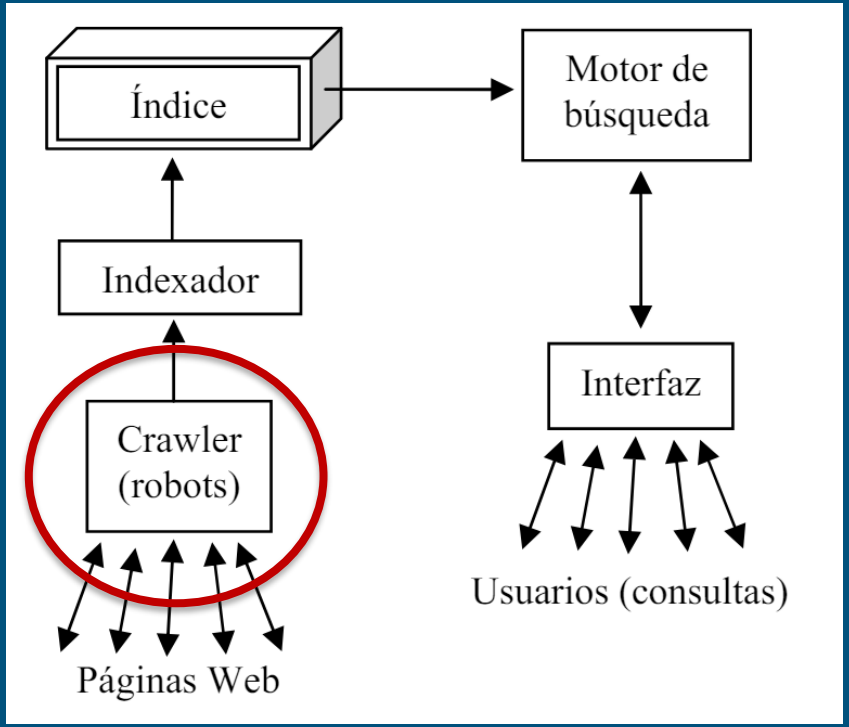
$$recall = \frac{N^{o}\,de.documentos\ .relevantes\ .recuperado\,s}{N^{o}\,de.documentos\ .relevantes\ .de.la.colección}$$

$$precisión = \frac{N^{o}\,de.documentos\ .relevantes\ .recuperado\,s}{N^{o}\,total\ .de.documentos\ .recuperado\,s}$$



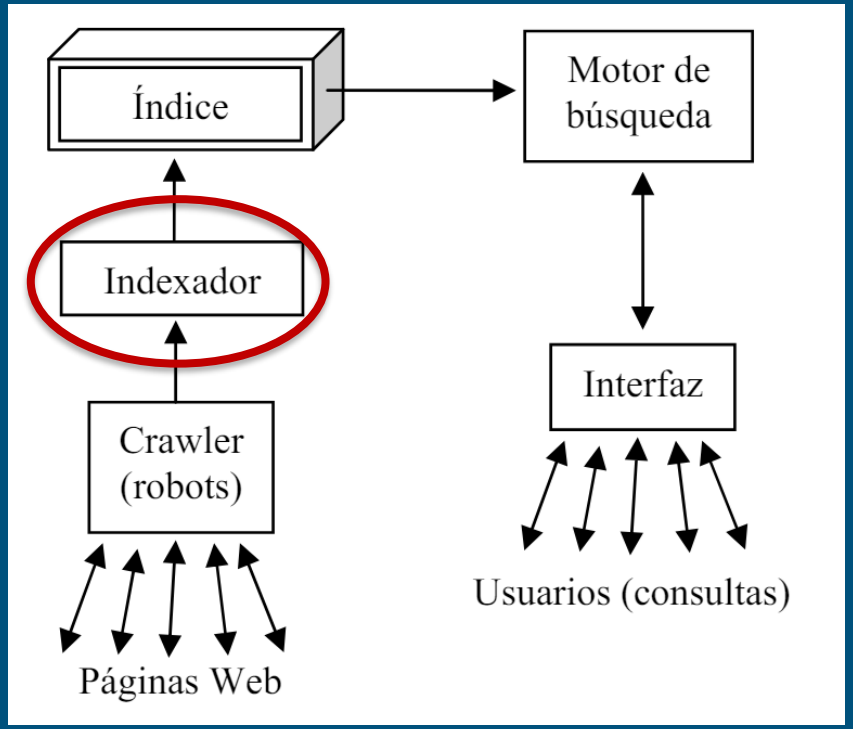
Arquitectura de los buscadores

- **Crawler (Rastreador):**
  - Función: Programa que navega la web, recopila documentos y enlaces.
  - Proceso: Explora páginas web, sigue enlaces y crea una lista de URL.
  - Almacenamiento: Guarda las URL de las páginas recopiladas para su procesamiento.



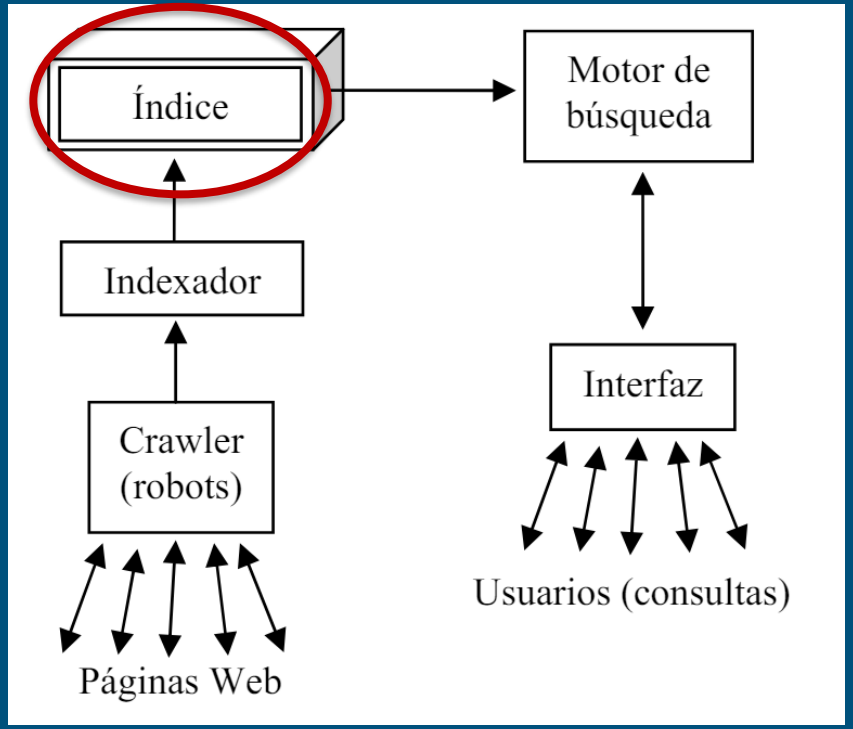
Arquitectura de los buscadores

- **Base de datos o indexador:**
  - Función: Almacena información recopilada por el crawler para búsqueda rápida.
  - Proceso de indexación: Analiza documentos, extrae términos relevantes y los almacena junto con las URL.
  - Algoritmos de indexado: Organiza términos para facilitar la recuperación.



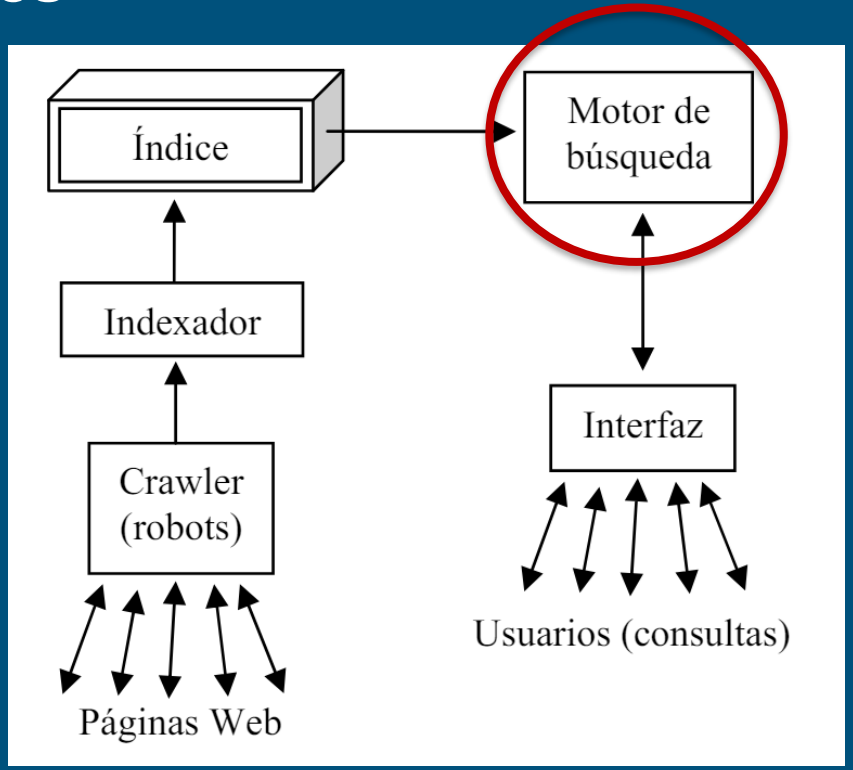
Arquitectura de los buscadores

- **Generador del Índice:**
  - Función: Procesa documentos recopilados por el crawler y los agrega al índice.
  - Proceso de procesamiento: Analiza contenido de documentos y los prepara para su almacenamiento.



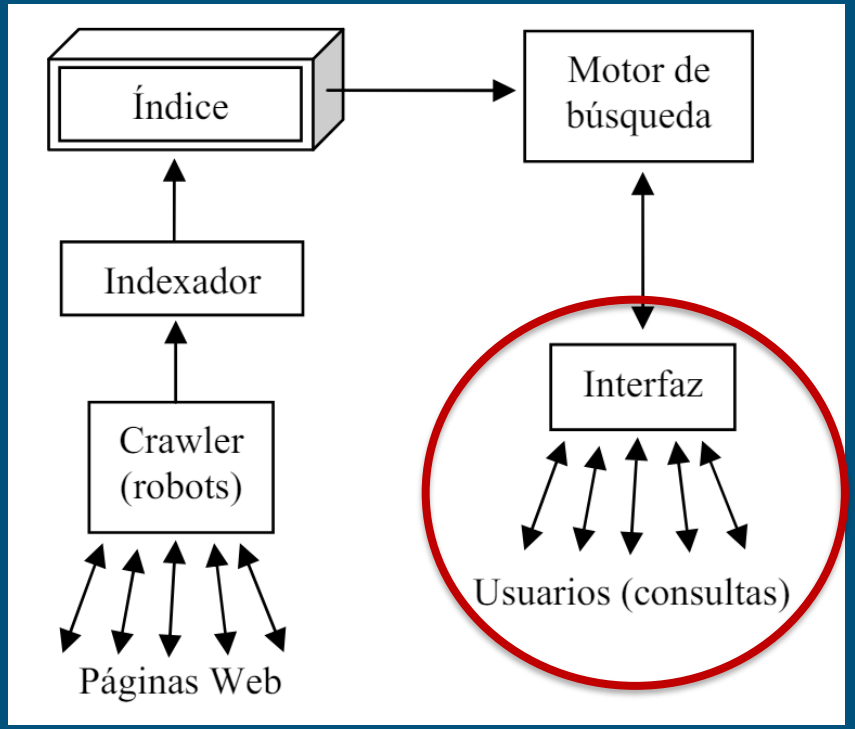
Arquitectura de los buscadores

- **Motor de Búsqueda:**
  - Función: Procesa consultas de usuarios y recupera resultados relevantes de la base de datos.
  - Proceso de recuperación: Encuentra documentos que coincidan con términos de la consulta y los presenta al usuario.

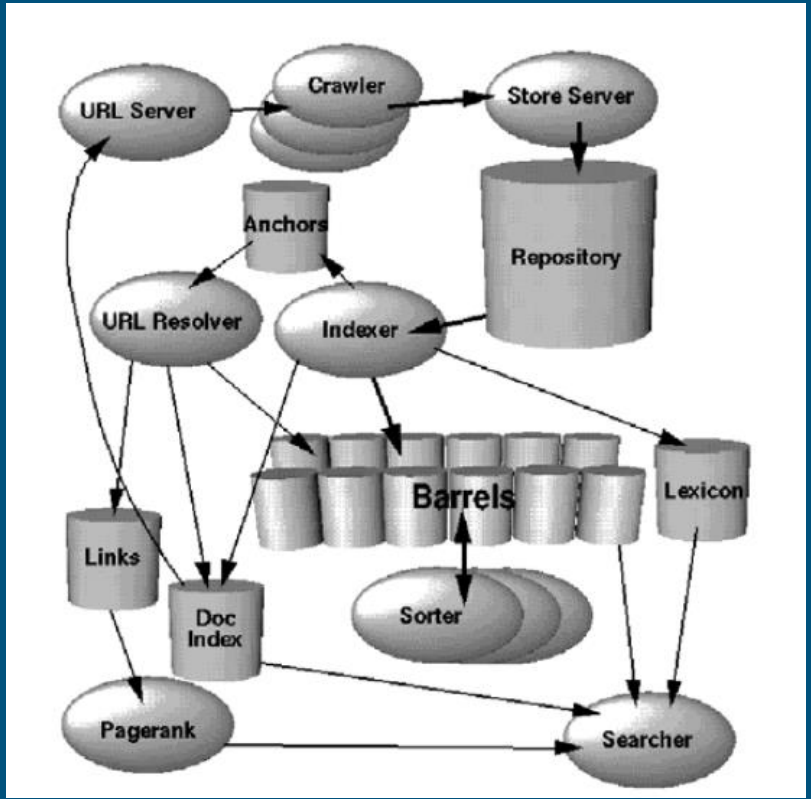


Arquitectura de los buscadores

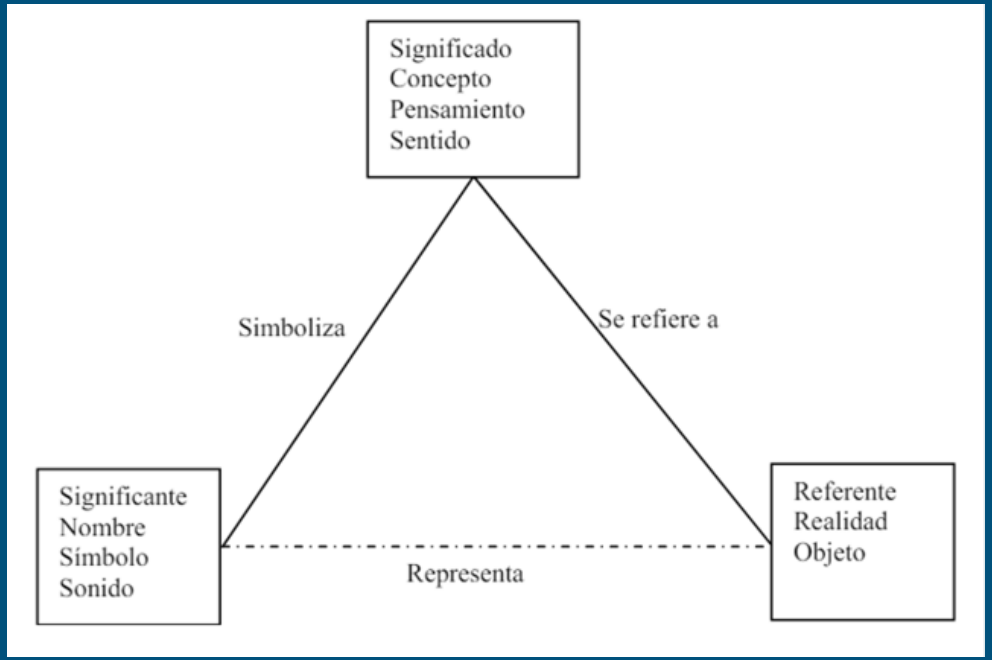
- **Interfaz de Búsqueda:**
  - Función: Proporciona una forma para que los usuarios ingresen consultas y revisen resultados.
  - Proceso de interacción: Presenta una interfaz intuitiva para ingresar consultas y filtrar resultados.



Google utiliza varios crawlers distribuidos para descargar páginas web



Semántica y problemas de búsqueda



La semántica lingüística es esencial para la recuperación de información, las redes semánticas se utilizan para representar relaciones entre conceptos, ya que la distribución geográfica de usuarios y la evolución del vocabulario son desafíos para la búsqueda web.

Desafíos en la búsqueda web

Los desafíos en la búsqueda web son numerosos y reflejan la complejidad del contenido disponible en la red. La distribución, contenido, la presencia de páginas duplicadas, los tipos de datos y la calidad variable del contenido son algunos de los principales problemas.

Para abordar estos desafíos, se pueden implementar diversas soluciones:

- Utilizar técnicas de distribución y replicación de datos
- Implementar algoritmos de duplicación de contenido
- Implementar estrategias de actualización continua de datos
- Emplear técnicas de procesamiento de lenguaje natural y aprendizaje automático.

Finalmente, para garantizar la calidad del contenido, es importante implementar sistemas de filtrado y clasificación de contenido, así como colaborar con expertos para evaluar la veracidad y relevancia de la información.

Seguridad en la web

Seguridad IP

En 1994 se publica “Seguridad en la Arquitectura del Internet”.

Con el fin de conocer los problemas más frecuentes que enfrentamos sin un control no autorizado del tráfico de red.

Ahora se tendría la necesidad de asegurar el tráfico entre usuarios finales utilizando mecanismos de autentificación y cifrado.

Para el año 2001 el CERT informaba sobre más de 52,000 incidentes de seguridad.

Los tipos de ataques más graves incluían los falsos IP.

Intrusos creaban paquetes con direcciones IP falsas y explotaban las aplicaciones que usan autenticación basada en IP.

Y distintas formas de escucha y captura de paquetes, donde los atacantes leen la información transmitida, incluida la de conexión al sistema y la de los contenidos de bases de datos.

Amenazas a la seguridad web

Ataques pasivos

Incluyen escuchas del tráfico de la red entre el navegador y el servidor y la obtención de acceso a información de un sitio web que se supone restringida.

Ataques activos

Incluyen la suplantación de otro usuario, la alteración de mensajes en tránsito entre cliente y servidor y la modificación de información de un sitio web.


IPSEC
INTERNET
PROTOCOL
SECURITY

Aplicaciones

Este protocolo proporciona la capacidad de asegurar las comunicaciones a través de una LAN, de una WAN privada y pública y de internet. Ejemplos:

- Conexión segura entre oficinas sucursales a través de internet.
- Acceso remoto seguro a través de Internet.
- Establecimiento de conexión extranet e intranet con socios.
- Mejora de la seguridad en el comercio electrónico.

Beneficios

- Cuando se implementa en un cortafuegos o un router, proporciona una gran seguridad que se puede aplicar a todo el tráfico que lo cruza.
- Está por debajo de la capa de transporte por ellos es transparente a las aplicaciones.
- Puede ser transparente a usuarios finales. No es necesario entrenar a los usuarios para la utilización de mecanismos de seguridad.
- Puede proporcionar seguridad a usuarios individuales si es necesario, lo cual es útil para trabajadores externos.

Documentos

Request for Comments (RFC) es un documento numérico en el que se describen y definen protocolos, conceptos, métodos y programas de Internet.

- RFC 2401: Descripción general de una arquitectura de seguridad.
- RFC 2402: Descripción de la extensión de autenticación de un paquete a IPv4 e IPv6.
- RFC 2406: Descripción de la extensión de cifrado de un paquete a IPv4 e IPv6.
- RFC 2408: Especificación de las capacidades de gestión de claves.

Permitir estas características es obligatorio para IPv6 y opcional para IPv4.

Además de estos cuatro RFC Security Protocol Working Group ha publicado una serie de borradores adicionales. Los documentos se dividen en siete grupos.

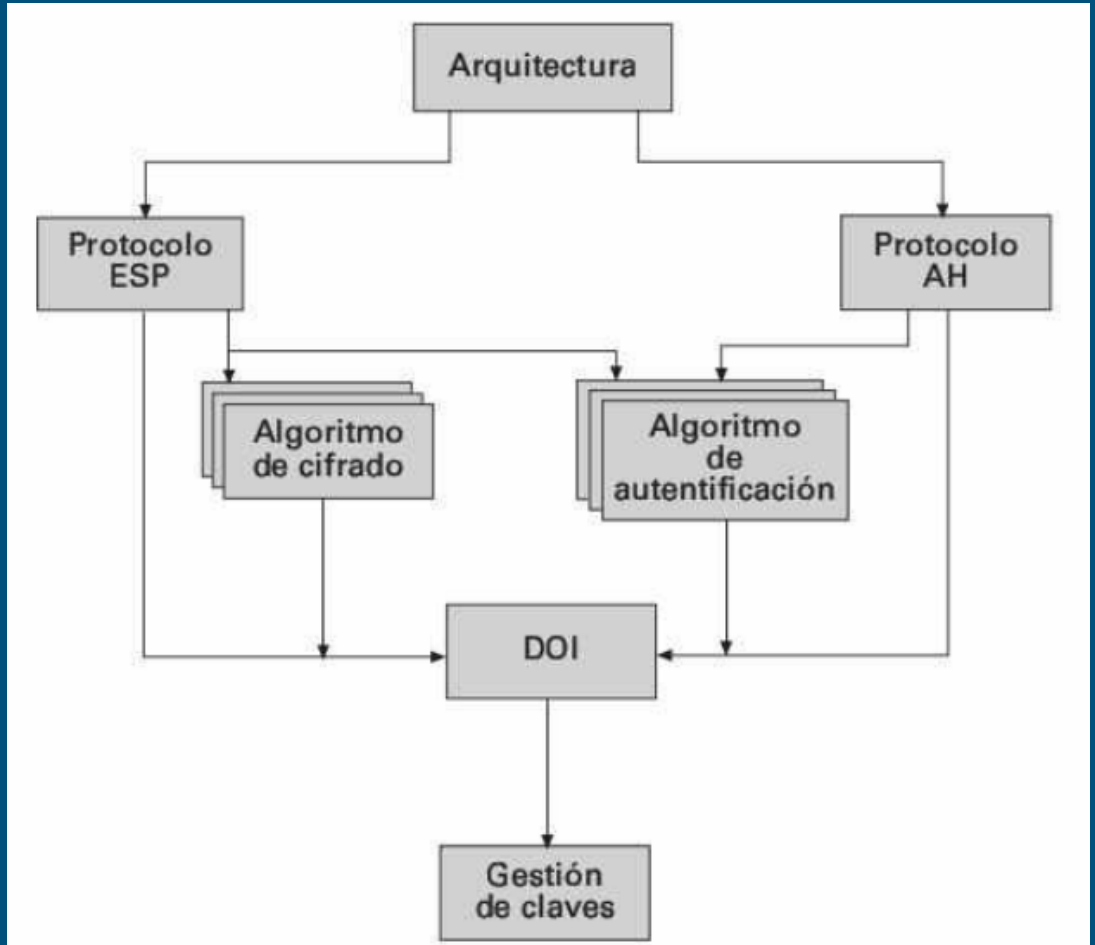


Figura 6.2 Esquema general de documento IPSec

Servicios

Tabla 6.1 Servicios de IPSec

	AH	ESP (sólo cifrado)	ESP (cifrado y autentificación)
Control de acceso	✓	✓	✓
Integridad sin conexión	✓		✓
Autentificación del origen de datos	✓		✓
Rechazo de paquetes reenviados	✓	✓	✓
Confidencialidad		✓	✓
Confidencialidad limitada del flujo de tráfico		✓	✓

Se usan dos protocolos para proporcionar seguridad: un protocolo de

autentificación designado por la cabecera del protocolo, AH, y un protocolo combinado de

cifrado/autenticación designada por el formato del paquete para ese protocolo, ESP.

Asociaciones de seguridad

Una asociación es una relación unidireccional entre un emisor y un receptor que ofrece servicios de seguridad al tráfico que se transporta. Los servicios de seguridad se suministran a una SA para que use AH o ESP, pero no los dos.

Una asociación de seguridad se identifica unívocamente por tres parámetros:

- Índice de parámetros de seguridad.
- Dirección IP de destino.
- Identificador del protocolo de seguridad.

Modo transporte y túnel

Tabla 6.2 Funcionalidad del modo túnel y del modo transporte

	SA en modo transporte	SA en modo túnel
AH	Autentifica la carga útil de IP y las partes seleccionadas de la cabecera IP y de las cabeceras de extensión de IPv6.	Autentifica todo el paquete IP interno (cabecera interior más carga útil de IP) más las partes seleccionadas de la cabecera IP exterior y las cabeceras externas de extensión de IPv6.
ESP	Cifra la carga útil de IP y cualquier cabecera de extensión de IPv6 que siga a la cabecera ESP.	Cifra el paquete IP interior.
ESP con autentificación	Cifra la carga útil de IP y cualquier cabecera de extensión de IPv6 que siga a la cabecera ESP. Autentifica la carga útil de IP, pero no la cabecera IP.	Cifra el paquete IP interior. Autentifica el paquete IP interior.

Secure Sockets Layer

Arquitectura SSL

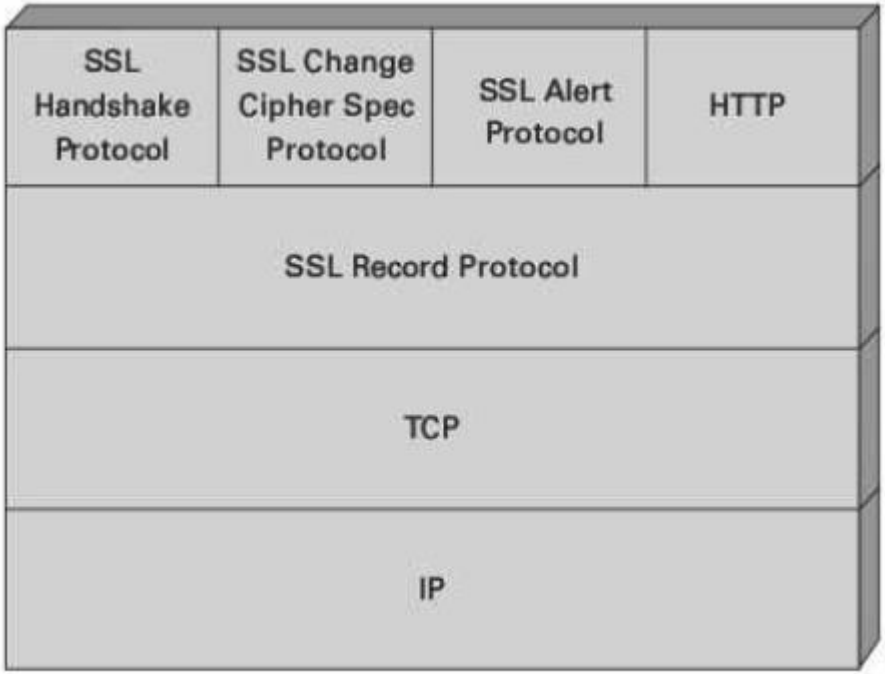


Figura 7.2 Pila de protocolos SSL

SSL está diseñado de forma que utilice TCP para proporcionar un servicio fiable y seguro

extremo a extremo. SSL no es un protocolo simple, sino que tiene dos niveles de protocolos.

Dos conceptos importantes de SSL son la sesión SSL y la conexión SSL, definidos en las especificaciones:

- Conexión: Para SSL cada conexión está asociada con una sesión.
- Sesión: Una sesión SSL es una asociación entre un cliente y un servidor. Las sesiones las crea el protocolo Handshake.

Una fase de sesión se define por los siguientes parámetros:

- Identificador de sesión.
- Certificado de la entidad par.
- Método de compresión.
- Especificación de cifrado.
- Clave maestra.
- Es reanudable.

Un estado de conexión se define por los siguientes parámetros:

- Valores aleatorios del servidor y del cliente.
- Clave secreta para MAC de escritura del servidor.
- Clave secreta para MAC de escritura del cliente.
- Clave de escritura del servidor.
- Clave de escritura del cliente.
- Vector de inicialización.
- Números de secuencia.

Protocolo record de SSL

Proporciona dos servicios a las conexiones SSL:

- Confidencialidad:

  El protocolo Handshake define una clave secreta compartida que se usa para cifrado convencional de la carga útil de SSL.

- Integridad de mensajes:

  El protocolo Handshake también define una clave secreta compartida que se usa para formar un código de autentificación de mensajes (MAC).

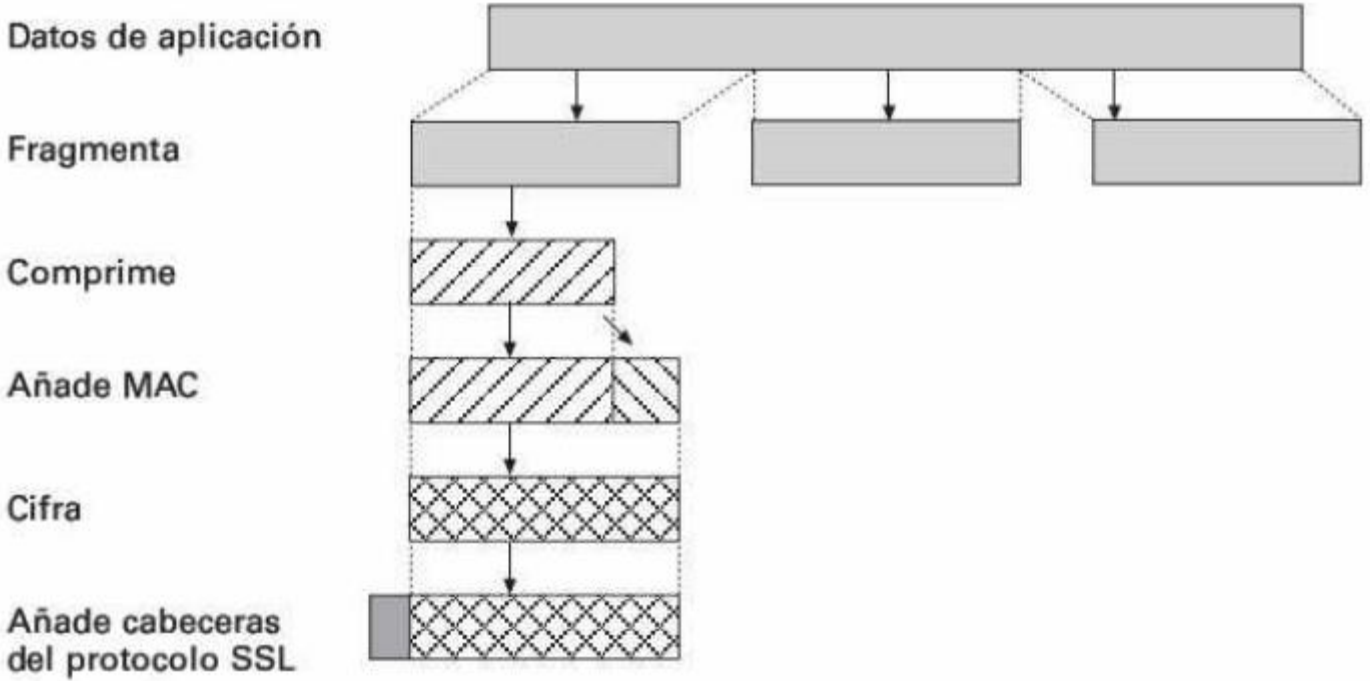
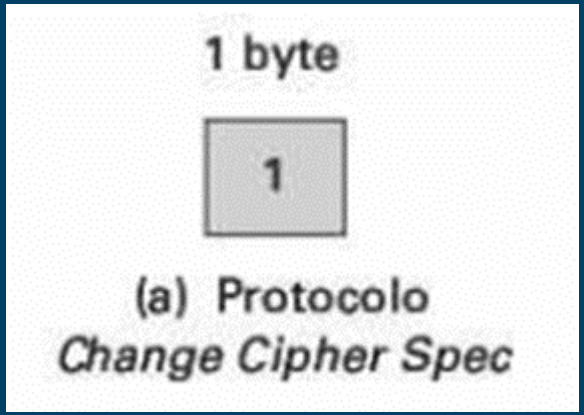


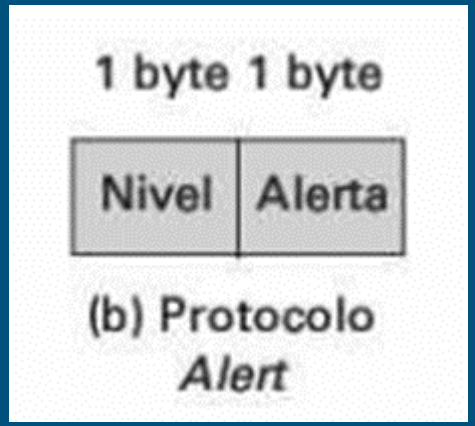
Figura 7.3 Operación del protocolo *Record* de SSL

Protocolo Change Cipher Spec

Este es uno de los tres protocolos específicos de SSL que utilizan el protocolo Record y, además, es el más simple. Este protocolo se compone de un único mensaje, que contiene un único byte con el valor 1.



Protocolo Alert



El protocolo Alert se usa para transmitir las alertas relacionadas con SSL a la entidad par,

los mensajes de alerta se comprimen y se cifran, según se especifica en el estado en

operativo.

Protocolo Handshake

La parte más compleja de SSL es el protocolo Handshake. Este protocolo permite la autentificación mutua de servidor y cliente y negociar un algoritmo de cifrado y de cálculo del MAC y las claves criptográficas que se utilizarán para proteger los datos enviados en un registro SSL. Consiste en una serie de mensajes intercambiados entre servidor y cliente.

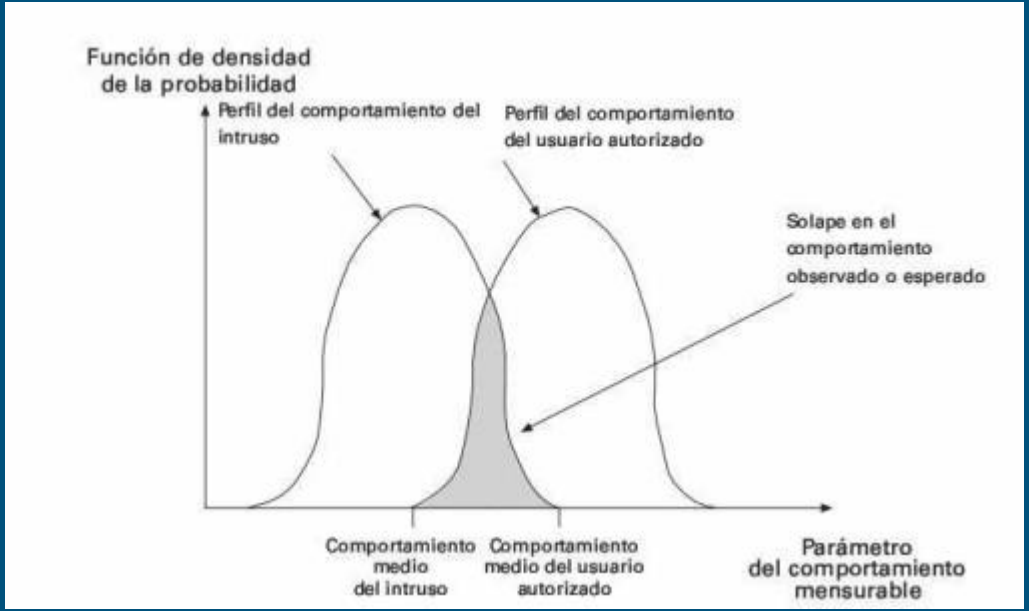
(c) Protocolo *Handshake*

Tipos de Intruso

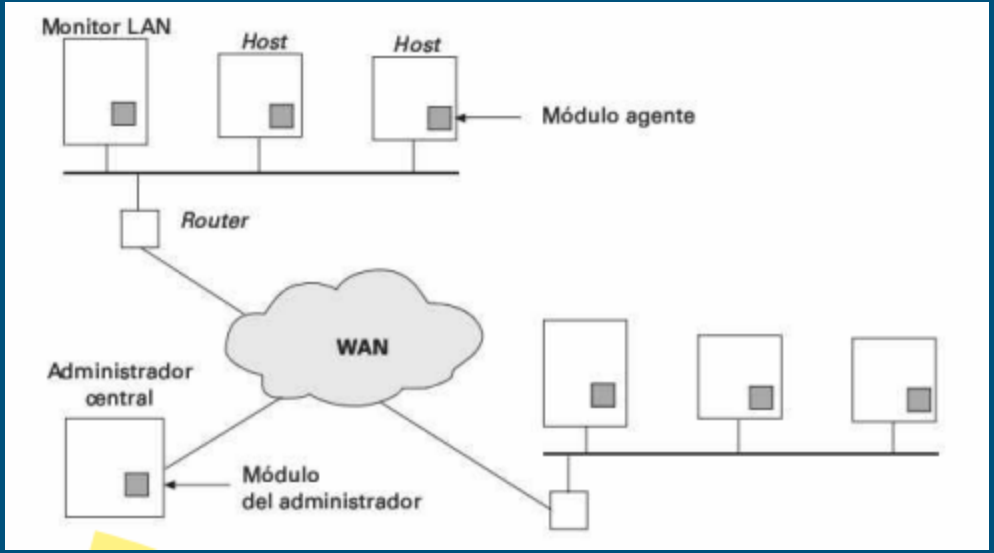
- Suplantadores
- Usuarios clandestinos
- Variedad de propósitos y niveles de riesgo
- Ataques: desde exploración de redes hasta manipulación/interrupción de sistemas

Técnicas de intrusos

- Descubrimiento de contraseñas (pruebas exhaustivas, información personal)
- Importancia de la detección de intrusos como segunda línea de defensa
- Métodos: estadísticos, reglas predefinidas para identificar actividades sospechosas



Registros de auditoría



- Papel crucial en la detección de intrusos
- Registro de actividades de usuarios para análisis posterior
- Recopilar y analizar datos de manera eficiente

Honeypots

- Atraer y recopilar información sobre posibles atacantes
- Simulación de entornos atractivos pero falsos
- Permite rastrear y responder a ataques sin comprometer sistemas reales

Sistemas de contraseñas

- Primera línea de defensa en sistemas multiusuario
- Autenticación de identidad y privilegios de usuarios
- Problema de contraseñas débiles (fáciles de adivinar)

Tipo de contraseña	Tamaño de la búsqueda	Número de coincidencias	Porcentaje de contraseñas coincidentes (%)	Ratio coste/ beneficio[a]
Nombre de la cuenta o de usuario	130	368	2,7	2,830
Secuencias de caracteres	866	22	0,2	0,025
Números	427	9	0,1	0,021
Chino	392	56	0,4	0,143
Nombres de lugares	628	82	0,6	0,131
Nombres comunes	2239	548	4,0	0,245
Nombres femeninos	4280	161	1,2	0,038
Nombres masculinos	2866	140	1,0	0,049
Nombres poco comunes	4955	130	0,9	0,026
Mitos y leyendas	1246	66	0,5	0,053
Sobre Shakespeare	473	11	0,1	0,023
Deportes	238	32	0,2	0,134
Ciencia ficción	691	59	0,4	0,085
Películas y actores	99	12	0,1	0,121
Dibujos animados	92	9	0,1	0,098
Gente famosa	290	55	0,4	0,190
Frases y patrones	933	253	1,8	0,271
Apellidos	33	9	0,1	0,273

Enfoque de antivirus

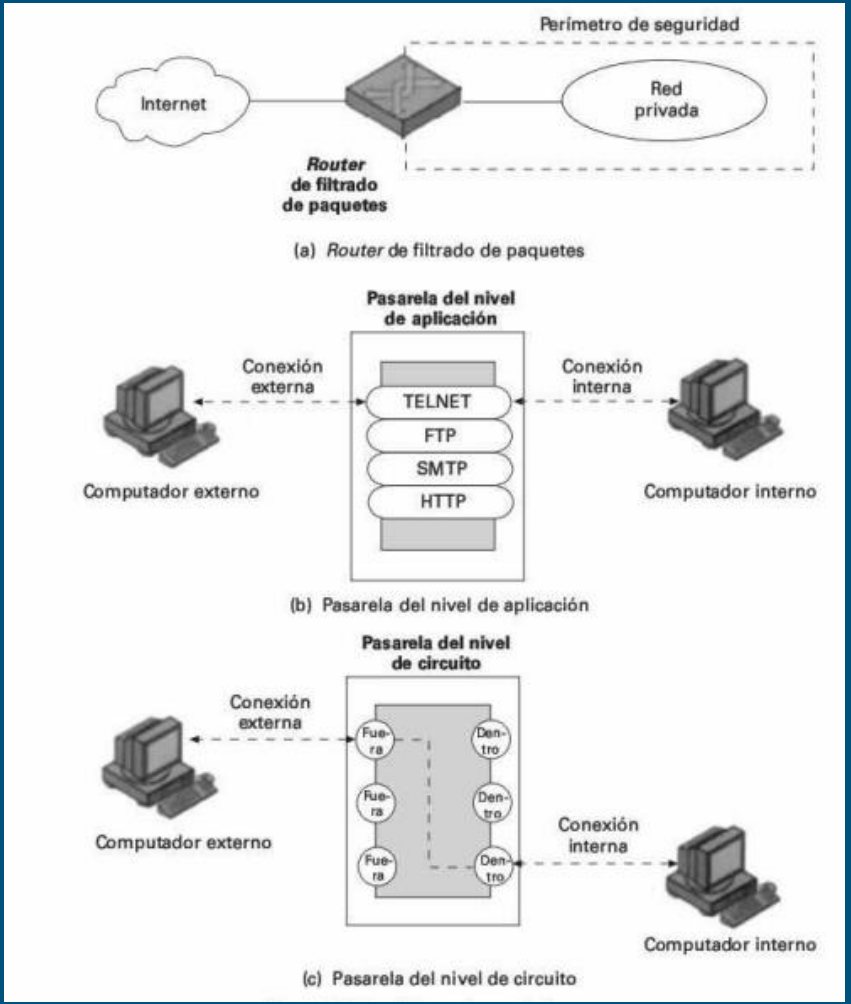
- Prevención, detección, identificación y eliminación de virus
- Exploradores de firmas específicas (primera generación)
- Reglas heurísticas (segunda generación)
- Programas residentes en memoria (tercera generación)
- Paquetes combinados (cuarta generación)

Software de bloqueo de acciones

- Monitoreo en tiempo real del funcionamiento del programa
- Bloqueo de acciones potencialmente dañinas
- Protección contra modificaciones no autorizadas

Cortafuegos

- Barrera de seguridad entre la red corporativa e Internet
- Controla el tráfico y aplica medidas de seguridad
- Tipos: router de filtrado de paquetes, pasarela del nivel de aplicación, pasarela del nivel de circuito



Conclusión

El desarrollo de herramientas de búsqueda y protocolos de seguridad ha transformado nuestra interacción con la información y la protección de datos en línea. La evolución desde Archie hasta Google ha mejorado significativamente la indexación y recuperación de información. Al mismo tiempo, protocolos como IPSec y SSL han fortalecido la seguridad de las comunicaciones en redes públicas y privadas. Para enfrentar los desafíos crecientes, es esencial continuar innovando en la mejora de la precisión de las búsquedas y en la implementación de medidas de seguridad robustas, garantizando un entorno digital seguro y eficiente para todos.

Referencias

Libro 1 William Stallings, (2004), FUNDAMENTOS DE SEGURIDAD EN REDES. APLICACIONES Y ESTÁNDARES. Segunda edición, España, Pearson educación, [Versión electrónica]. Recuperado de: https://web.instipp.edu.ec/Libreria/libro/Fundamentos_de_seguridad_en_redes_Aplica.pdf, (pp. 29-51, 61-75).

Libro 2 José Angel Olivas Varela, (2011), Búsqueda eficaz de información en la Web, (1er edición), Argentina, Editorial de la Universidad Nacional de La Plata (Edulp), [Versión electrónica]. Recuperado de: https://libros.unlp.edu.ar/index.php/unlp/catalog/view/322/302/984-1, (pp. 177-201, 223-246, 306-320, 322-329, 353-355,357,363-366,370-371).