Universidad Politécnica
de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Grado en Grado en Ingeniería Informática

Trabajo Fin de Grado

Desarrollo de un Sistema de Intercambio Directo de Archivos entre Dispositivos Basado en IPFS

Autor: Nicolás Cossío Miravalles
Tutor(a): Fernando Pérez Costoya

Madrid, Abril - 2023

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Grado en Ingeniería Informática

Título: Desarrollo de un Sistema de Intercambio Directo de Archivos entre Dispositivos Basado en IPFS

Abril - 2023

Autor: Nicolás Cossío Miravalles
Tutor: Fernando Pérez Costoya
Arquitectura Y Tecnología De Sistemas Informáticos
ETSI Informáticos
Universidad Politécnica de Madrid

Resumen

IPFS, también conocido como Protocolo de Sistema de Archivos Interplanetario, es un protocolo de red y un sistema de archivos diseñado para hacer la web más rápida, segura y abierta. Este sistema permite a los usuarios no solo recibir, sino también alojar contenido en una red P2P completamente descentralizada.

IPFS tiene varias ventajas clave. A diferencia de protocolos como HTTP, en IPFS los recursos se identifican por su contenido en lugar de por su ubicación. Esta característica permite a cualquier nodo de la red convertirse en proveedor de contenido dentro de ella, lo que se traduce en una mayor eficiencia, seguridad, escalabilidad y resiliencia para el almacenamiento y distribución de datos. IPFS facilita la creación de aplicaciones descentralizadas (dApps) al proporcionar herramientas como un sistema de almacenamiento de archivos distribuido y un sistema de nombres descentralizado (IPNS) para la web. Al mismo tiempo, promueve el desarrollo de aplicaciones resistentes a la censura y una web verdaderamente abierta y descentralizada.

Este trabajo de fin de grado se divide en dos partes:

La primera consiste en el estudio del ecosistema de IPFS. Se abarca desde su arquitectura, algoritmo de intercambio de bloques, identificación basada en contenido, hasta su estructura de datos. Se analizan ejemplos de casos de uso en la Web3, como la distribución descentralizada de contenido, el almacenamiento de datos en la cadena de bloques y la publicación de datos permanentes.

La segunda parte del trabajo consiste en la creación de un sistema de intercambio de archivos basado en IPFS. Se presenta un posible diseño de una arquitectura centralizada habitual que se usaría para una aplicación de intercambio seguro de archivos. Se profundiza en posibles puntos únicos de falla, preocupaciones de privacidad y problemas de escalabilidad que surgen al depender de una sola autoridad o servidor. Con estos puntos establecidos, se introduce el sistema ideado. Empleando la naturaleza distribuida de IPFS, esta propuesta tiene como objetivo abordar los problemas mencionados y la mejora de la propiedad y la privacidad de los datos.

Algunas características fundamentales de este sistema son:

- Archivado y compresión de archivos y directorios utilizando tar y gzip.
- Encriptación segura de archivos utilizando aes-256-cbc.
- Encriptación de secretos facilitada por JSON Web Encryption (JWE).
- Verificación de autoría mediante el uso de Identificadores Descentralizados (DIDs), en forma de firma de contenido utilizando JSON Web Signatures (JWS).
- Uso de bases de datos descentralizadas impulsadas por OrbitDB, que permiten:

- Silos de usuarios, registro automático y controladores de acceso distribuidos.
- Notificaciones push mediante una cola de mensajes descentralizada.
- Bases de datos locales con persistencia para uso interno de la aplicación.

La aplicación desarrollada funciona en sistemas operativos Windows, MacOS y Linux. Mediante una interfaz de comandos de consola los usuarios pueden compartir archivos de manera segura y privada sin la necesidad de depender de servidores centralizados.

Abstract

IPFS, also known as the InterPlanetary File System, is a network protocol and file system designed to make the web faster, more secure and open. This system allows users not only to receive but also to host content on a fully decentralized peer-to-peer network.

IPFS has several key advantages. Unlike protocols like HTTP, in IPFS, files are identified by their content rather than their location. This feature allows any node in the network to become a content provider, resulting in greater efficiency, security, scalability and resilience for data storage and distribution.

IPFS facilitates the creation of decentralized applications (dApps) by providing tools such as a distributed file storage system and a decentralized naming system (IPNS) for the web. At the same time it promotes the development of censorship-resistant applications as well as a truly open and decentralized web.

This undergraduate thesis is divided into two parts:

The first part consists of the study of the IPFS ecosystem. From its architecture, block exchange algorithm, content-based addressing, to its data structure. Examples of use cases in Web3, such as decentralized content distribution, blockchain-based data storage, and permanent data publishing, are also analyzed.

The second part of the thesis involves the creation of a secure and decentralized file-sharing system based on IPFS. The process starts by outlining the design and limitations of a typical centralized architecture for an application of the proposed type. Emphasizing on potential single points of failure, privacy concerns and scalability issues that arise from relying on a single authority or server.
With these points established, the devised system is then introduced. Employing the distributed nature of IPFS, this proposal aims to address the aforementioned issues, while also enhancing data privacy and ownership.

Fundamental features of this system encompass:

- File or directory archiving and compression using tar and gzip.
- Secure file encryption using aes-256-cbc.
- Secrets encryption facilitated by JSON Web Encryption (JWE).
- Authorship verification through the usage of Decentralized Identifiers (DIDs), in the form of content signing using JSON Web Signatures (JWS).
- Usage of decentralized databases powered by OrbitDB which enable:
  - User silos, automatic registration, and distributed access controllers.
  - Push notifications via a decentralized message queue.
  - Local databases with persistence for internal application use.

The developed application supports Windows, MacOS, and Linux operating systems. Through a command-line interface, users can securely and privately share files without relying on centralized servers.

Tabla de contenidos

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

El presente Trabajo de Fin de Grado (TFG) se centra en el desarrollo de un sistema de intercambio de ficheros basado en IPFS (InterPlanetary File System)[1].
A continuación, se describen las motivaciones y necesidades que han llevado a la realización de este proyecto.

1.1. Motivación y necesidad

El desarrollo de un sistema de intercambio de ficheros basado en IPFS se encuentra en la confluencia de varias tendencias tecnológicas y sociales que están dando forma al futuro de la web. En particular, este proyecto se relaciona estrechamente con el avance hacia la *Web3*[2], una visión de un internet más descentralizado, seguro y resistente a la censura. En esta sección, exploraremos cómo un sistema de intercambio de archivos encaja en este nuevo panorama y por qué es relevante para el progreso de la Web3.

Los servicios de almacenamiento y compartición de archivos actuales, como Google Drive, Dropbox, Microsoft OneDrive y otros proveedores de almacenamiento en la nube son servicios centralizados. Pese a ser populares y ampliamente utilizados debido a su facilidad de uso, accesibilidad y confiabilidad, presentan ciertos problemas y limitaciones. Los usuarios dependen de una sola entidad para almacenar y gestionar sus archivos, lo que puede generar problemas si la empresa experimenta fallos técnicos, cambia sus políticas de uso, o se convierte en el objetivo de un ataque cibernéticos maliciosos. Además, esto otorga a estas empresas un gran poder sobre los datos de los usuarios, lo que puede conducir a problemas de privacidad y control de la información.

Otras alternativas como FTP (File Transfer Protocol) ofrecen una mayor autonomía y control sobre los archivos, pero también tienen inconvenientes. FTP es un protocolo que permite la transferencia de archivos entre un cliente y un servidor a través de una red. FTP carece de robustas medidas de seguridad modernas, puede ser vulnerable a ataques y requiere un mayor conocimiento técnico y esfuerzo para su configuración y mantenimiento.
En resumen, a pesar de la mayor autonomía y control directo que FTP puede ofrecer, no es comparable con un servicio en la nube en términos de seguridad, facilidad de uso y eficiencia de costos. Esto es teniendo en cuenta los conocimientos y requisitos del usuario promedio de un servicio de estas características.

La arquitectura detrás de este tipo de servicios se basa en el modelo cliente-servidor. En este modelo, un servidor central almacena la información sobre la lista de nodos y recursos disponibles en la red y es vital para el funcionamiento del sistema. Esto facilita encontrar rápidamente los nodos o recursos disponibles, pero el sistema es relativamente vulnerable en términos de fallos o ataques y la escalabilidad está limitada debido a la presión sobre el elemento central [3].

La alternativa a estos servicios centralizados es el uso de tecnologías *peer-to-peer* (de igual a igual en español). Una aplicación peer-to-peer (p2p) es un tipo de red donde no existen clientes ni servidores fijos, sino una serie de nodos que actúan como iguales y pueden funcionar tanto como clientes como servidores entre sí.

Existen varias tecnologías p2p que permiten compartir archivos entre usuarios sin necesidad de un proveedor central, el más famoso y conocido siendo BitTorrent[4]. Sin embargo, estas tecnologías no son adecuadas para el intercambio de archivos entre usuarios no conocidos, ya que requieren que los usuarios confíen en que los archivos que se comparten son los que se anuncian.

Esto es algo que resuelve el Inter Planetary File System (IPFS). El Sistema de Archivos Interplanetario es un sistema de archivos distribuido que busca conectar todos los dispositivos al mismo sistema de archivos. En cierto modo, IPFS es similar a la Web, aunque podría verse como una sola red BitTorrent, intercambiando objetos dentro de un repositorio Git.
En otras palabras, IPFS permite guardar y acceder a bloques de datos identificados por su contenido, no por su ubicación, y que se pueden transferir rápidamente entre los nodos. Además, IPFS usa estos bloques para crear enlaces que también se basan en el contenido, no en una dirección que apunta a una ubicación donde se puede encontrar el contenido. Esto forma un grafo dirigido acíclico generalizado de Merkle (Merkle DAG), una estructura de datos sobre la que se puede construir sistemas de archivos versionados, cadenas de bloques e incluso una Web Permanente. IPFS combina una tabla hash distribuida, intercambio de bloques incentivado y espacio de nombres autocertificante, sin puntos únicos de falla ni necesidad de confianza entre los nodos que la forman[5].

En este proyecto se usará IPFS como bloque central, sobre el que construirá el sistema previamente descrito.

1.2. Objetivos y alcance del proyecto

El objetivo principal de este proyecto es el desarrollo de un sistema de intercambio de ficheros basado en IPFS, mediante una aplicación de escritorio. Este sistema debe permitir a los usuarios compartir archivos de forma segura y confiable, sin necesidad de ningún proveedor central de ningún tipo.
Debe integrar capacidades de encriptación y control de acceso para garantizar la seguridad de los archivos compartidos. La integración de cuentas de usuario, con la posibilidad de hacer grupos, elegir contactos con los que compartir, se propone como algo imprescindible para lograr un sistema autocontenido y sin necesidad de herramientas externas para su uso. Por último se debe integrar un sistema de notificaciones para el que los usuarios puedan recibir avisos de nuevos archivos compartidos, o de cambios en los archivos compartidos.

Para lograr esto se han cumplido los siguientes objetivos:

- Investigar sobre IPFS y su funcionamiento para entender cómo funciona el protocolo y cómo se puede utilizar para el sistema propuesto.
- Investigar sobre el ecosistema en torno a IPFS, con objetivo de comprender la madurez y viabilidad de esta tecnología, así como de las herramientas basadas en esta que se pueden utilizar para el sistema propuesto.
- Diseñar una arquitectura para el sistema de intercambio en torno a las tecnologías y herramientas seleccionadas.
- Implementación de un prototipo funcional del sistema propuesto.
- Analizar la viabilidad de IPFS en base a la experiencia obtenida en el desarrollo del prototipo.
- Analizar posibles mejoras y ampliaciones del sistema propuesto.

Por tanto pese a que el objetivo principal es el desarrollo de un sistema de intercambio de ficheros basado en IPFS, también se realizará una labor de divulgativa sobre IPFS y su ecosistema, con el objetivo de comprender esta tecnología y su viabilidad como alternativa a muchas de las tecnologías actuales.

1.3. Estructura de la memoria

En este capítulo se ha introducido el proyecto, explicando las motivaciones y necesidades que han llevado a su realización.

En el capítulo 2: 'Contexto' se pone en situación el estado actual de tecnologías relacionadas con el proyecto, tanto alternativas como otras implementaciones que usen IPFS u otras tecnologías similares que cumplan parcial o completamente con los objetivos del proyecto. Al comienzo de este capítulo también se explica brevemente la historia de internet y su evolución hasta el presente. La razón de ser de esta sección se debe a la necesidad de poner en situación el porqué detrás de la dominancia de ciertos protocolos que han guiado el modelo de internet actual, y que han llevado a la necesidad de alternativas como IPFS.

Dentro de este capítulo se explica el funcionamiento de IPFS, tratando los siguientes temas: arquitectura interna, funcionamiento, ecosistema y herramientas relacionadas. Con esta sección se busca dar una visión general de esta tecnología y su ecosistema para poder entender el sistema propuesto.

En el capítulo 3: 'Estado del arte' se lleva a cabo un breve análisis de algunas otras implementaciones que usan IPFS o tecnologías similares, así como de otras alternativas a IPFS que cumplen parcial o completamente con los objetivos del proyecto.

El capítulo 4: 'Desarrollo de IPFShare' se centra en el desarrollo del sistema propuesto. Este se ha estructurado en en:

- **Requisitos del sistema:** se explica el funcionamiento deseado del sistema.
- **Diseño del sistema:** se presenta la arquitectura y diseño propuestos en este proyecto, así como las herramientas utilizadas.
- **Implementación:** se explica la implementación realizada, así como las decisiones tomadas durante el desarrollo. Esta sección incluye partes de código relevantes para entender la implementación realizada.

En el capítulo 5: 'Resultados y conclusiones' se realiza una serie de pruebas del sistema desarrollado. Para ello se ha creado un escenario de uso real con distintos usuarios en varios lugares del mundo.

El capítulo 6: 'Resultados y conclusiones' se analiza el resultado obtenido del desarrollo del proyecto. Se contrastará el resultado con los objetivos propuestos y con servicios de transferencia de archivos centralizados.

Sobre el alcance del proyecto, en el capitulo 7: 'Trabajos futuros' se explora las posibles vías de expansión y mejoras para el proyecto en el futuro. También se expresan las esperanzas y expectativas para el crecimiento y posible impacto del mismo.

Capítulo 2

Contexto

En esta sección se intentarán poner en perspectiva, de una forma no exhaustiva, las distintas razones históricas que dan lugar a la necesidad de crear un sistema de almacenamiento descentralizado y distribuido como IPFS. Para ello, se hará un breve repaso histórico de la evolución de Internet y de los protocolos que lo han ido conformando. Posteriormente, se explicará la tendencia centralista del sistema actual, existente a un nivel intrínseco y estructural, además de otros problemas que se derivan de esta situación. Finalmente, se expondrá la propuesta de solución que IPFS ofrece para solventar estos problemas, sobre la que se profundizará en la sección 2.2: 'IPFS como alternativa a HTTP'.

2.1. Breve historia de Internet

2.1.1. Predominancia de los protocolos TCP/IP

La historia de internet está marcada por la competencia entre distintos protocolos de comunicación que buscaban establecerse como el estándar para intercambiar información entre diferentes redes y sistemas. Uno de los episodios más relevantes de esta competencia fue la llamada *"Guerra de los protocolos"* [6], en la que el conjunto de protocolos TCP/IP, creado entre los años 1973 y 1974 por Vint Cerf y Robert Kah, se enfrentó a otras propuestas como OSI, X.25 o SNA[1].

TCP/IP logró imponerse a la competencia debido a las siguientes características principalmente:

- **Interoperabilidad** : La capacidad de TCP/IP para conectarse fácilmente con diferentes tipos de ordenadores y sistemas operativos le otorgaba una ventaja sobre otros protocolos que eran más específicos o limitados en su compatibilidad. Esta característica permitía que diversas tecnologías y plataformas pudieran comunicarse entre sí sin problemas, lo cual era esencial para crear una red global como internet.
- **Flexibilidad** : TCP/IP podía adaptarse a distintos medios de transmisión, como cables de cobre, fibra óptica o incluso enlaces inalámbricos, lo que facilitaba su implementación en una amplia variedad de entornos y situaciones. Otros

[1] En la figura 1 de la página 42 se muestra un resumen de las principales características de cada uno de estos protocolos.

protocolos, en cambio, podrían haber requerido modificaciones o adaptaciones específicas para funcionar en diferentes tipos de medios de transmisión.

- **Resistencia** frente a fallos: TCP/IP fue diseñado para ser robusto en caso de fallos en la red, permitiendo que los paquetes de datos pudieran ser retransmitidos y encontrar rutas alternativas en caso de problemas. Esta capacidad de recuperación era fundamental para garantizar la continuidad y fiabilidad de las comunicaciones en una red global con múltiples nodos y enlaces.
- **Escalabilidad** : TCP/IP podía soportar el crecimiento de la red al permitir la incorporación de nuevos nodos y enlaces sin afectar negativamente su rendimiento. Su diseño jerárquico y descentralizado facilitaba la expansión de la red y evitaba los cuellos de botella que podrían haberse producido con otros protocolos menos escalables.

Estas ventajas hicieron que TCP/IP se convirtiera en la opción preferida frente a otros protocolos, al ser una solución más versátil, resistente y escalable para la creciente demanda de interconexión entre sistemas y redes en todo el mundo. Cabe destacar también que era una solución con arquitectura abierta, no propietaria y de uso gratuito, es decir, sin necesidad de pagar licencias por su uso [7].

Como en toda guerra también hubo un trasfondo político. Este hecho suele ser ignorado al abordarse este tema desde un punto de vista puramente tecnológico. Y es que en 1980, el Departamento de Defensa de Estados Unidos declaró TCP/IP como el estándar para todas las redes militares [8]. A esto se sumaron numerosas comunidades de investigación y universidades que adoptaron TCP/IP como su protocolo de comunicación, como por ejemplo, Stanford University, donde Vint Cerf colaboró con Robert Kahn en el diseño del protocolo [8]; University of California, Los Angeles (UCLA), que participó en el desarrollo temprano y las pruebas de TCP/IP [8]; y University College London (UCL), donde el profesor Peter Kirstein promovió el uso de TCP/IP en Europa y su equipo contribuyó al desarrollo y pruebas del protocolo [9].

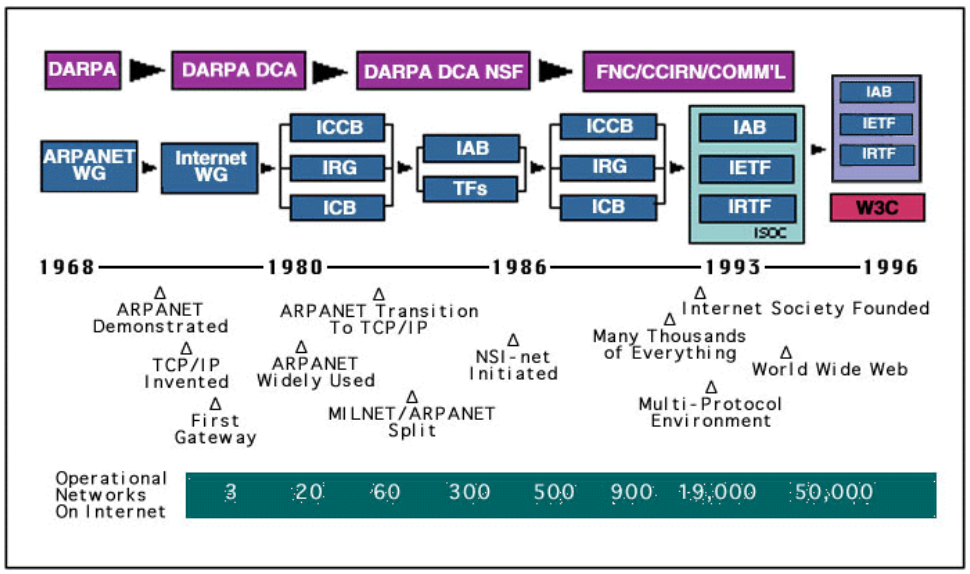

Figura 2.1: Evolución de los protocolos de Internet. Fuente [8]

Esta completa adopción del protocolo se dió por finalizada cuando ARPANET precur-

sor de internet y financiado por la Agencia de Proyectos de Investigación Avanzados de Defensa (DARPA), llevó a cabo la transición exitosa de su antiguo protocolo, el Network Control Program (NCP), a TCP/IP el 1 de enero de 1983 [8].

En resumen, la rápida adopción de la comunidad científica y académica, sumada al respaldo gubernamental consolidaron TCP/IP como el estándar dominante en la industria de las redes de comunicación.

2.1.2. La World Wide Web y HTTP

El modelo TCP/IP asentó una forma de comunicación estándar entre computadores y redes, aunque este estaba limitado principalmente al mundo académico y científico. No fue hasta la creación de la World Wide Web (WWW) cuando el Internet concebido como es en la actualidad se convirtió en un fenómeno global y accesible para todo el mundo.

Antes de la WWW, el acceso a la información en Internet se realizaba a través de los protocolos a nivel de aplicación mostrados en la figura 2.1

Protocolo	**Descripción**
FTP (Protocolo de Transferencia de Archivos)	Utilizado para transferir archivos entre cliente y servidor a través de una red.
Telnet	Basado en texto utilizado para el acceso remoto a computadoras y servidores, permitiendo a los usuarios controlarlos a través de una interfaz de línea de comandos.
Gopher	Diseñado para buscar y recuperar documentos de manera jerárquica, utilizando una interfaz basada en menús.
SMTP (Protocolo Simple de Transferencia de Correo)	Utilizado para enviar mensajes de correo electrónico entre servidores y, finalmente, al cliente de correo del destinatario.
NNTP (Protocolo de Transferencia de Noticias en Red)	Utilizado para la distribución, consulta y recuperación de artículos de noticias en la red Usenet.
POP3 (Protocolo de Oficina de Correos 3)	Utilizado para recuperar mensajes de correo electrónico desde un servidor de correo remoto hasta un cliente de correo local.
IMAP (Protocolo de Acceso a Mensajes de Internet)	Permite a los usuarios acceder y administrar sus mensajes de correo electrónico en un servidor de correo, sin descargarlos a un cliente de correo local.

Cuadro 2.1: Protocolos de capa de aplicación antes de HTTP

Estos servicios se encuentran en nivel de aplicación dentro del *stack* TCP/IP, como se muestra en la figura 2.2. Algunos de ellos se siguen usando hoy en día, o tienen su caso de uso (IMAP, POP3, FTP), pero en lo referente a archivos, ofrecían métodos básicos de navegación y compartición. Carecían de la capacidad de inter-conectar documentos de manera intuitiva y visual. [2]

[2]Cabe destacar que en esta época, los documentos eran principalmente texto plano, sin formato, y

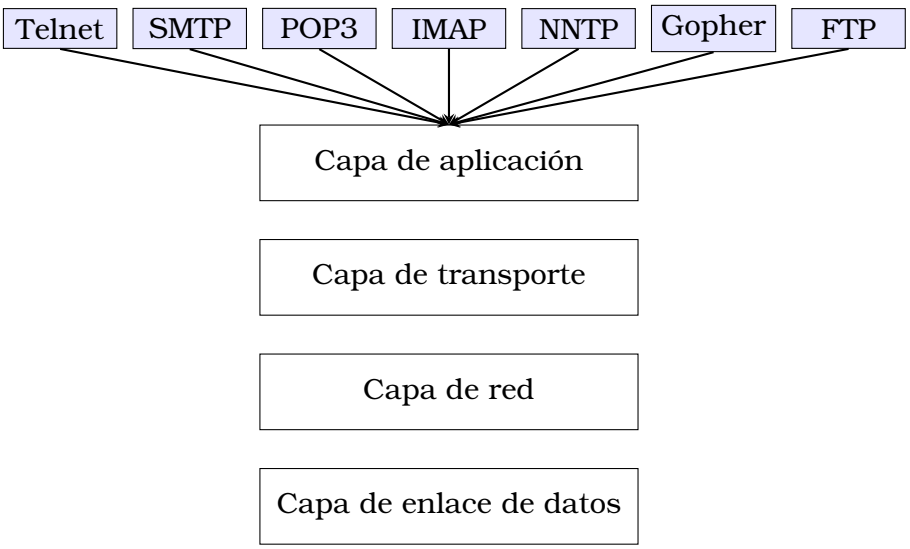

Figura 2.2: Capas del protocolo TCP/IP mostrando algunos protocolos de la capa de aplicación

En 1989, el científico británico Tim Berners-Lee propuso la creación de la WWW, un sistema de información global que permitiría a los usuarios navegar y acceder a documentos interconectados mediante enlaces. Estos documentos, conocidos como páginas web, se almacenarían en computadoras conectadas a la red y podrían ser accedidos a través de un programa especial llamado navegador web, que interpretaría el código de las páginas y mostraría su contenido al usuario.
HTML (Hyper Text Markup Language) es el lenguaje que describe estos documentos. Permite enlazar documentos entre sí mediante hipervínculos. Un hipervínculo es una referencia unidireccional en un documento electrónico que entrelaza diferentes documentos o secciones entre sí. Los usuarios tienen la oportunidad de seguir estos enlaces con tan solo un clic en el texto ancla (texto enlazado) para navegar a los documentos o las secciones correspondientes[10]. Aunque es un concepto simple y con el que cualquier persona en la actualidad esta familiarizada este factor dictamina la forma en la que se usa internet en la actualidad. Los usuarios de internet interactúan con el contenido en internet mediante estos enlaces.
La WWW se basó en tres tecnologías clave: HTML, un lenguaje de marcado para crear páginas web; HTTP, un protocolo para solicitar y transferir recursos a través de la web; y URL, un sistema de direcciones para localizar recursos en la web[8]. Y es este último el que genera una gran problemática que resuelve IPFS.

URL significa Uniform Resource Locator, que se traduce al español como Localizador Uniforme de Recursos. Es un sistema de direcciones utilizado en la web para localizar de manera única recursos como páginas web, imágenes, videos y otros archivos. Una URL consta de varios componentes, incluyendo el esquema (como 'http://' o 'https://'), el nombre de dominio (como 'www.ejemplo.com'), la ruta del recurso y otros parámetros opcionales.

Sin embargo, a medida que la web ha crecido en tamaño y complejidad, el enfoque de direccionamiento basado en la ubicación física de los servidores puede presen-

no existía la posibilidad de incluir imágenes o vídeos.

tar limitaciones. Por ejemplo, si un recurs' se encuentra en una URL específica y esa URL cambia o el servidor deja de estar disponible, el acceso al recurso se verá comprometido.

IPFS aborda este problema mediante el uso de un sistema de direccionamiento basado en el contenido, en lugar de la ubicación. En IPFS, cada archivo y bloque de datos se identifican mediante su contenido, utilizando una función hash criptográfica. Esto permite que los archivos y bloques se puedan encontrar y acceder de forma fiable, independientemente de su ubicación física.

Esto permite a IPFS ofrecer una serie de ventajas sobre el sistema de direccionamiento basado en la ubicación de la web tradicional, como la resistencia a la censura, la persistencia de los datos y la verificabilidad del contenido. En la siguiente sección se profundizará en estas ventajas y en cómo IPFS las hace posibles.

2.2. IPFS como alternativa a HTTP

2.2.1. Introducción

IPFS fue presentado al mundo en 2014 por Juan Benet, en un informe técnico titulado *IPFS - Content Addressed, Versioned, P2P File System*[5]. Benet presenta el concepto de IPFS y su proposición de crear un sistema de archivos distribuido y descentralizado que permita a los usuarios almacenar y compartir archivos de forma segura y confiable.

Benet es también el fundador de Protocol Labs[11], una empresa dedicada a la creación de protocolos de código abierto para la Web3. IPFS es un proyecto de código abierto y, pese a que Protocol Labs está detrás de este, no es el único contribuidor a su desarrollo. Esto es otro de los puntos fuertes de IPFS, la comunidad que lo rodea. En la sección 2.3: 'Ecosistema en torno a IPFS' se profundiza en este aspecto.

En IPFS, cada archivo se identifica de manera única a través de su contenido mediante un hash criptográfico. Esto significa que cualquier nodo en la red puede actuar como un proveedor de contenido al almacenar y compartir archivos, permitiendo una mayor disponibilidad y un internet verdaderamente descentralizado. En lugar de depender de un único servidor web para acceder a un recurso, los usuarios pueden obtener el contenido de cualquier nodo que tenga ese recurso en particular.

Estos identificadores de contenido se conocen como CID (Content Identifier). Dado que un CID es un puntero que señala a un contenido particular, se puede usar un CID en vez de URL en un enlace. De esta manera se puede acceder a un recurso de manera fiable, independientemente de su ubicación física, mientras haya algún otro nodo de la red en posesión del contenido que buscamos.

2.2.2. Fundamentos

IPFS opera a través de tres principios fundamentales que marcan una diferencia significativa con respecto a los sistemas de archivos convencionales: direccionamiento por contenido, red peer-to-peer y el grafo acíclico dirigido de Merkle (Merkle DAG).

Direccionamiento por Contenido: En IPFS, los archivos no se ubican por su dirección sino por su contenido. Cada archivo posee un identificador único, denominado CID (Content Identifier), generado a partir de un hash criptográfico de su contenido. Esta característica asegura la inmutabilidad de los archivos, es decir, los archivos no pueden ser alterados sin modificar su CID. Adicionalmente, el direccionamiento por contenido favorece la deduplicación, dado que archivos con contenido idéntico compartirán el mismo CID, lo que conlleva a su almacenamiento único dentro de la red.

```
1 $ ipfs add somefile # comando para añadir archivos a ipfs a través de la
  ↪  línea de comandos
2 added QmVtuHo6C7NUonYTYUNbmgGxvSwrTVGXuahJxhxnoSxPpM somefile
```

Figura 2.3: Ejemplo de CID generado por IPFS

Red Peer-to-Peer: IPFS se basa en una red descentralizada en la que cada integrante, o nodo, puede interactuar directamente con cualquier otro nodo, sin la necesidad de intermediarios o servidores centrales. Los nodos funcionan tanto como proveedores como consumidores de contenido, guardando y compartiendo fragmentos de archivos con otros nodos. Esta red peer-to-peer hace que el contenido sea más accesible y resistente a la censura, al evitar la existencia de un único punto de fallo o control.

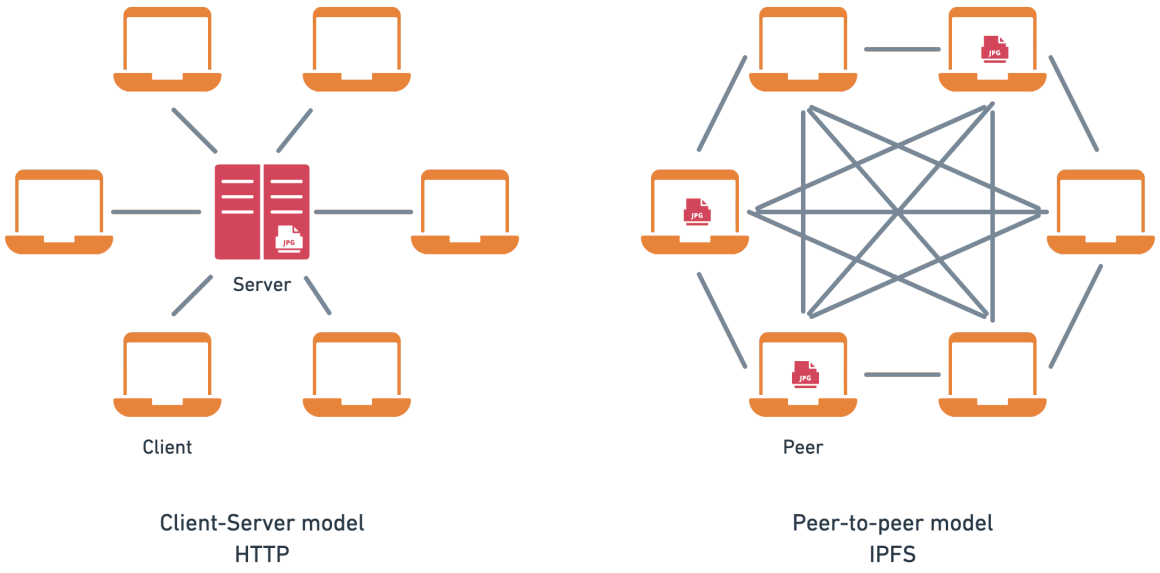

Figura 2.4: Red centralizada en comparación con una red descentralizada

Grafo Acíclico Dirigido de Merkle (Merkle DAG): Los archivos y sus relaciones dentro de IPFS se representan mediante una estructura de datos conocida como Merkle DAG. Un Merkle DAG es un grafo donde cada nodo tiene un identificador único (CID) que se genera a partir de su contenido y el de sus nodos hijos. Los nodos pueden ser hojas o nodos intermedios, dependiendo de si tienen o no nodos hijos. Los nodos hoja contienen datos binarios de los archivos, mientras que los nodos intermedios contienen enlaces a otros nodos. Los nodos intermedios permiten dividir

archivos grandes en bloques más pequeños y formar estructuras jerárquicas, como directorios o sistemas de archivos. El Merkle DAG facilita la verificación de integridad y autenticidad de los archivos, dado que cualquier cambio en el contenido o en los enlaces se refleja en el CID del nodo afectado y sus ancestros.

Cada uno de estos conceptos se profundizará dentro del apartado correspondiente a continuación.

2.2.3. Arquitectura

IPFS es un conjunto de protocolos de código abierto que combina múltiples conceptos existentes de redes peer-to-peer (P2P), datos enlazados y otras áreas para permitir que los participantes intercambien fragmentos de archivos.

Estos conceptos concretados en protocolos forman distintos niveles de abstracción, cada uno de los cuales se puede utilizar de forma independiente y conforman la arquitectura de IPFS, también conocido como el *stack* de protocolos de IPFS. En la figura 2.5 se muestra el stack de protocolos de IPFS.

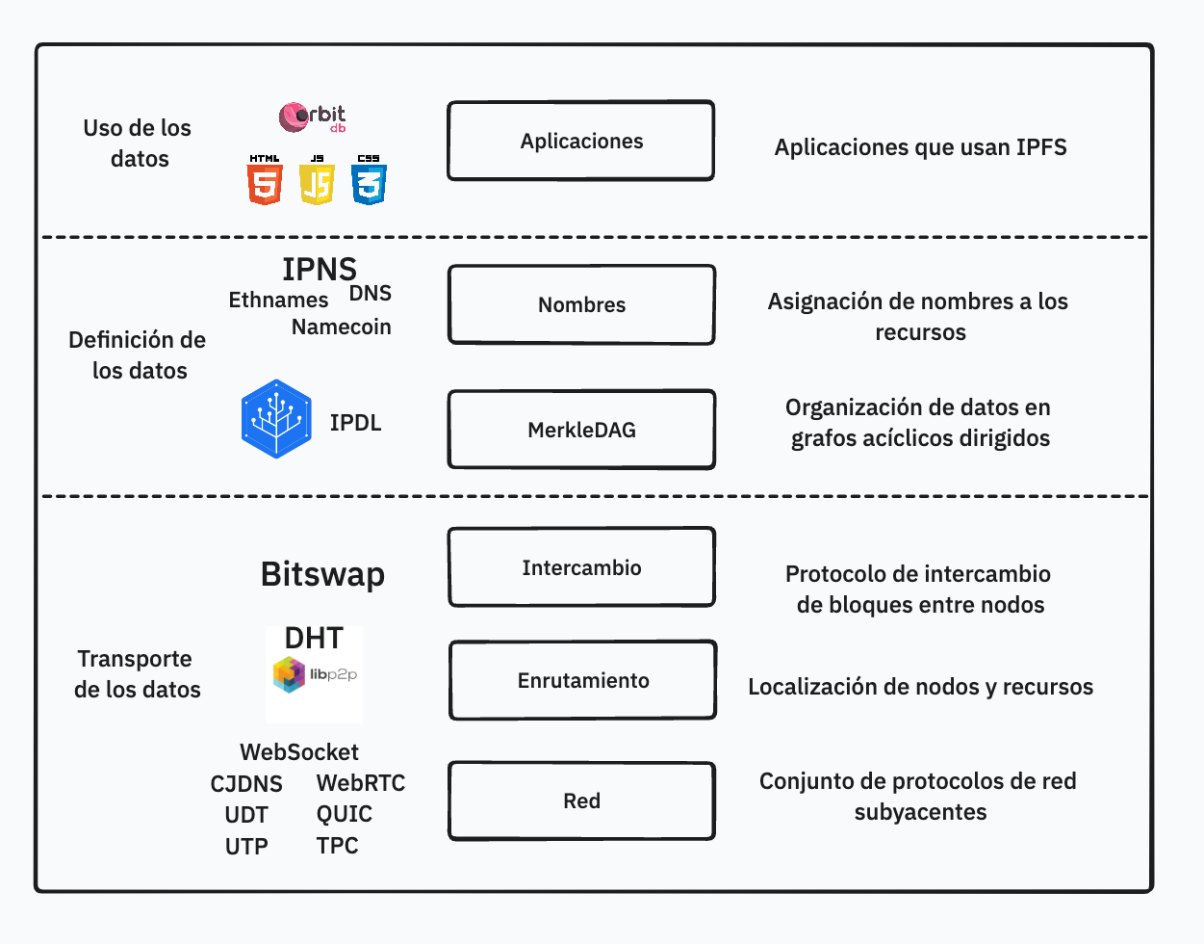

Figura 2.5: Stack de protocolos IPFS

Como se puede observar, este stack se divide tres grupos según la funcionalidad que

brinda cada capa. Este diseño en capas subdivido en componentes independientes permite que estos pueden ser ampliados o reemplazados según se necesite. Esta modularidad en el diseño está respaldada por una biblioteca de redes P2P llamada libp2p[12].

Libp2p es una suite de protocolos y herramientas modulares que permite la creación de sistemas de red peer-to-peer (P2P). Se encarga de gestionar todas las necesidades de red, como la negociación de protocolos, el enrutamiento, la detección de nodos y la transmisión de datos.

2.2.3.1. Capa de red

En la capa de red encontramos los protocolos de transporte de red mediante los cuales los nodos se pueden comunicar. Estos protocolos provienen de libp2p[13] y son los siguientes:

- **TCP**: proporciona una entrega de datos confiable, ordenada y con control de errores sobre redes IP.
- **UDP**: proporciona una entrega de datos simple, sin conexión y no confiable sobre redes IP.
- **QUIC**: Un protocolo de transporte multiplexado y seguro que se ejecuta sobre UDP, proporcionando flujos confiables, de baja latencia y cifrados.
- **WebSockets**: Un protocolo que permite la comunicación bidireccional entre un navegador web y un servidor sobre TCP.
- **WebRTC**: Un protocolo que permite la comunicación en tiempo real entre navegadores web mediante conexiones peer-to-peer.

Debido a esta gran variedad de protocolos de transporte, libp2p proporciona una forma de identificar el transporte que se esté usando mediante direcciones *multiaddr*. Los multiaddresses son una forma de representar las direcciones de red como encapsulaciones de protocolos arbitrarios. Estos multiaddresses admiten direccionar para cualquier protocolo de red. Siguen una sintáxis simple, lo que los hace fáciles de analizar y construir.

En IPFS, se utilizan los multiaddresses para identificar y localizar los nodos en la red. Cada nodo al configurarse por primera vez genera un par de claves pública y privada. La clave pública se utiliza para crear un identificador único del nodo en la red llamado peerID, que es el hash de su esta clave pública. La clave privada se utiliza para firmar mensajes y autenticar la identidad del nodo. Además, los nodos tienen una o más direcciones de red que combinan protocolos y valores para indicar cómo conectarse a ellos. Por ejemplo, una dirección de red podría ser /ip4/1.2.3.4/tcp/4001/ipfs/QmFoo, lo que significa que el nodo QmFoo está escuchando conexiones TCP en el puerto 4001 utilizando la dirección IP 1.2.3.4.

Los multiaddresses se pueden encapsular entre sí para crear capas de transporte más complejas. Por ejemplo, se puede utilizar /dns4/example.com/tcp/1234/tls/ws/tls para indicar una conexión segura con WebSockets sobre TLS utilizando el dominio example.com y el puerto 1234.

La amplia variedad de protocolos de transporte disponibles garantiza la adaptabilidad

de IPFS, ya que los nodos pueden utilizar múltiples protocolos de transporte simultáneamente y cambiar entre ellos según las condiciones de la red. Esto significa que las opciones de transporte de un nodo dependen del entorno en el que se ejecute. Por ejemplo, en un navegador web sólo se pueden utilizar WebSockets y WebRTC, mientras que en otros entornos se pueden utilizar todos los protocolos mencionados anteriormente, siempre que sean compatibles a nivel de sistema operativo y hardware.

2.2.3.2. Enrutamiento y descubrimiento de nodos

Descubrimiento de nodos:
Es el proceso de encontrar y anunciar servicios a otros nodos en una red P2P. Se puede realizar utilizando diversos protocolos, como por ejemplo, la difusión de mensajes a todos los nodos de la red o utilizar una serie nodo de arranque para proporcionar una lista de nodos conocidos.

Esto último se conoce como nodos de arranque (bootstrapping). Es una lista de nodos predefinidos y de confianza que ayudan a los nuevos nodos a unirse a la red y a descubrir otros nodos, facilitando el proceso de construcción y mantenimiento de la red distribuida. La lista de bootsrappers la define cada nodo. La figura 2.6 muestra los bootstrappers por defecto en una instalación de IPFS.

```
"Bootstrap": [
"/dnsaddr/bootstrap.libp2p.io/p2p/QmNnooDu7bfjPFoTZYxMNLWUQJyrVwtbZg5gBMj⌋
↪ TezGAJN",
"/dnsaddr/bootstrap.libp2p.io/p2p/QmQCU2EcMqAqQPR2i9bChDtGNJchTbq5TbXJJ16⌋
↪ u19uLTa",
"/dnsaddr/bootstrap.libp2p.io/p2p/QmbLHAnMoJPWSCR5Zhtx6BHJX9KiKNN6tpvbUcq⌋
↪ anj75Nb",
"/dnsaddr/bootstrap.libp2p.io/p2p/QmcZf59bWwK5XFi76CZX8cbJ4BhTzzA3gU1ZjYZ⌋
↪ cYW3dwt",
"/ip4/104.131.131.82/tcp/4001/p2p/QmaCpDMGvV2BGHeYERUEnRQAwe3N8SzbUtfsmvs⌋
↪ qQLuvuJ",
"/ip4/104.131.131.82/udp/4001/quic/p2p/QmaCpDMGvV2BGHeYERUEnRQAwe3N8SzbUt⌋
↪ fsmvsqQLuvuJ"
],
```

Figura 2.6: Bootstrappers por defecto en una instalación de IPFS

Enrutamiento:
Por otro lado, enrutamiento se refiere a encontrar la ubicación específica de otro nodo de la red. Esto se realiza típicamente mediante el mantenimiento de una tabla de enrutamiento u otra estructura de datos similar que realiza un seguimiento de la topología de la red. En el caso de IPFS se usa una tabla de hash distribuida conocida como DHT (Distributed hash table).

En la práctica, la distinción entre el enrutamiento y el descubrimiento de nodos no siempre está clara, de hecho suelen ocurrir simultáneamente.

Los protocolos principales[3] que usa IPFS y libp2p para este propósito son:

- **ping**: Protocolo de comprobación de disponibilidad. Los nodos pueden utilizarlo para verificar la conectividad y el rendimiento entre ellos.
- **autonat**: Protocolo de detección de NAT. Asiste a los nodos en la identificación de su accesibilidad desde internet, esto es útil para detectar si los nodos que se encuentran ocultos detrás de un NAT o de un firewall [14].
- **identify**: Protocolo para el intercambio de claves y direcciones con otros nodos. Facilita el intercambio de información esencial, como los protocolos soportados, las claves públicas, las direcciones, etc.
- **kademlia**: Protocolo para la implementación de una tabla hash distribuida para el almacenamiento descentralizado y la recuperación de información de nodos y contenidos.
- **mdns**: Protocolo de descubrimiento de nodos locales con cero configuración, usando DNS de multidifusión. Ofrece un mecanismo para que los nodos en la misma red local se descubran entre sí sin configuración previa.
- **Circuit Relays**: Es un protocolo que facilita a los nodos el reenvío de tráfico en nombre de otros nodos que no tienen un acceso directo entre ellos[15].
- **rendezvous**: Un protocolo de encuentro que se utiliza como un punto común entre dos rutas. Los puntos de encuentro son típicamente nodos que están bien conectados y son estables en una red, y pueden manejar grandes cantidades de tráfico y datos. Sirven como un centro para que los nodos se descubran. De los pocos mecanismos centralizados que usa libp2p [16].
- **pubsub**: Es una interfaz PubSub para libp2p, diseñada para establecer una base para la comunicación de mensajes mediante un patrón de publicación y suscripción entre los nodos de la red libp2p. Existen diferentes implementaciones de este protocolo, como FloodSub, GossipSub que proporcionan diferentes ventajas.

Tal como se indicó previamente, libp2p dispone de varias estrategias para el enrutamiento y la detección de nodos. IPFS las utiliza todas en sus diferentes configuraciones. La elección de la combinación de estas estrategias se realiza en función del contexto particular de un nodo respecto de otros nodos y la red.

2.2.3.3. Mecanismo de intercambio de contenido

Bitswap: Es un protocolo de intercambio de contenido que se ejecuta sobre una red P2P. Bitswap permite a los nodos intercambiar bloques de datos entre los nodos que conforman la red. Estos pueden solicitar bloques de datos a otros nodos y compartir los bloques que disponen. Bitswap utiliza un mecanismo de intercambio de deuda para garantizar que los nodos intercambien bloques de datos de manera justa y equitativa. Todo esto es posible gracias a que Bitswap mantiene un registro de los bloques que cada nodo tiene y los bloques que necesita.

La transferencia de datos (bloques) en IPFS está inspirada en BitTorrent, pero no es igual uno a uno comparado con este. Dos características de BitTorrent que utiliza IPFS:

[3]Existen más protocolos de enrutamiento y descubrimiento de nodos en libp2p pero estos son los más utilizados por IPFS.

- Estrategia de tit-for-tat (el que no comparte no recibe).
- Obtén primero las piezas raras (mejora el rendimiento).

Una diferencia notable es que en BitTorrent cada archivo tiene un enjambre (también conocido como *swarm*) separado de nodos (formando una red P2P entre ellos). En cambio, IPFS es una única red de nodos formando un gran swarm. La variedad de BitTorrent en IPFS es el ya mecionado Bitswap.

Bitswap es el algoritmo de intercambio de bloques, pero para realizar este intercambio primero se debe saber qué nodos pueden proveer los bloque que se buscan. Esto es posible gracias a la DHT. En IPFS la DHT se utiliza principalmente con dos funciones:

1. **Enrutamiento:** se ha explicado en la subsección anterior.
2. **Anuncios de provisión/consumición de contenido:** los nodos publican en la DHT los bloques de datos que tienen disponibles. Esto permite a otros nodos saber quién tiene los bloques de datos que están buscando. Esta tabla de hash se distribuye por la red mediante el algoritmo de Kademlia.

Sobre el segundo, que es el concierne dentro del mecanismo de intercambio de contenido: Cada nodo tiene dos lista de bloques (CIDs). Bloques que posee y puede proporcionar, y bloques que desea obtener.

- Al recibir una lista de deseos, una entidad que use Bitswap debería procesarla eventualmente y responder al solicitante con información sobre el bloque o el bloque en sí.
- Al recibir bloques, el nodo consumidor debe enviar una notificación de cancelación al resto de nodos a los que ha pedido estos bloques, señalando que ya no los desea.

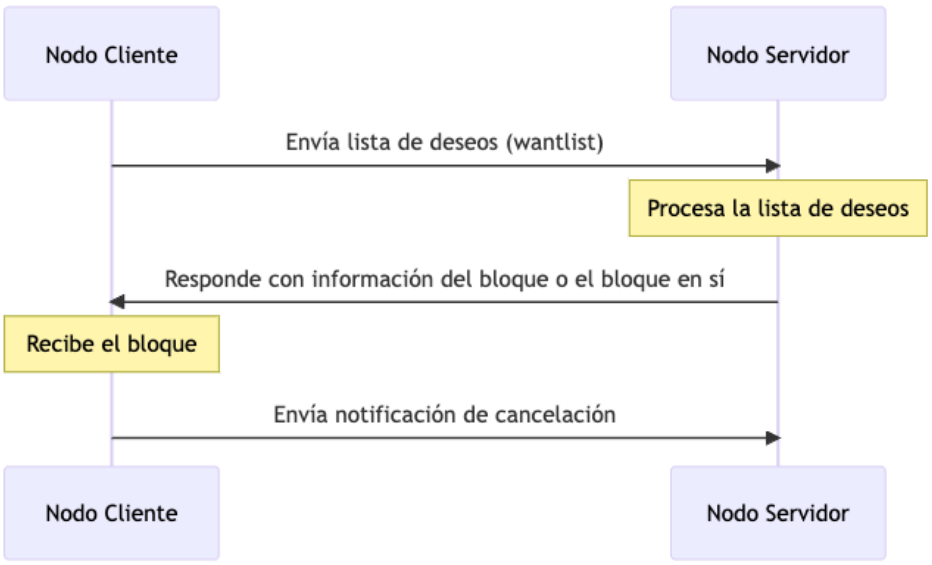

Figura 2.7: Protocolo Bitswap

Por último cabe destacar un concepto muy importante dentro de IPFS en torno al

guardado de bloques: **la recolección de basura.**

Cada nodo tiene una capacidad de almacenamiento, siendo esta el límite de bloques que puede almacenar. A medida que un nodo va obteniendo y compartiendo más contenido a través de IPFS, los bloques que recibe se van ubicando su almacenamiento local. La cantidad de bloques puede aumentar rápidamente y ocupar espacio innecesario. Algunos de estos bloques pueden estar referenciados por objetos obsoletos o que ya no son necesarios, lo que significa que no se utilizan ni se acceden directamente.

La recolección de basura en IPFS es el proceso mediante el cual se eliminan los bloques que ya no son necesarios en un nodo. Sin embargo, IPFS utiliza un sistema de almacenamiento basado en referencias, lo que significa que un bloque puede ser referenciado por múltiples objetos y mantenerse en el sistema aunque no esté directamente en uso.

Para evitar que los bloques necesarios sean eliminados por accidente sin el deseo del usuario poseedor del nodo, IPFS introduce el concepto de *pinning*. Un pin es una instrucción que le indica al nodo que debe mantener un bloque o conjunto de bloques en su almacenamiento local, incluso si no están siendo utilizados directamente por el nodo.

Los pinsets son conjuntos de CIDs que se desean mantener en el nodo. Estos pinsets permiten a los usuarios especificar qué bloques desean mantener de manera persistente, evitando así que sean eliminados durante el proceso de recolección de basura.

En resumen, los pinsets son conjuntos de CIDs que representan bloques que un nodo desea mantener en su almacenamiento local, y utilizan el concepto de 'pins' para asegurarse de que estos bloques no sean eliminados accidentalmente durante la recolección de basura.

2.2.4. Modelo de datos

2.2.4.1. DAG de Merkle

El modelo de datos en IPFS está basado en árboles de Merkle. Es una estructura de datos en la que cada nodo es una representación hash de un conjunto de datos. Los nodos hoja son representaciones (generalmente a través de una función de hash) de bloques de datos, mientras que cada nodo interno es la representación (nuevamente, generalmente a través de una función de hash) de sus nodos hijos. Esto crea un sistema en el que cualquier cambio en los datos originales cambiará los hashes en la ruta hasta la raíz del árbol, proporcionando una forma de verificar la integridad de los datos.

Un DAG de Merkle es una estructura donde cada nodo es un árbol de Merkle, y están conectados formando un grafo acíclico direccionado. Esto significa que los nodos están conectados de tal manera que siempre hay una dirección (de nodos padres a nodos hijos) y no hay ciclos (no puedes empezar en un nodo, seguir las conexiones y volver al nodo original).

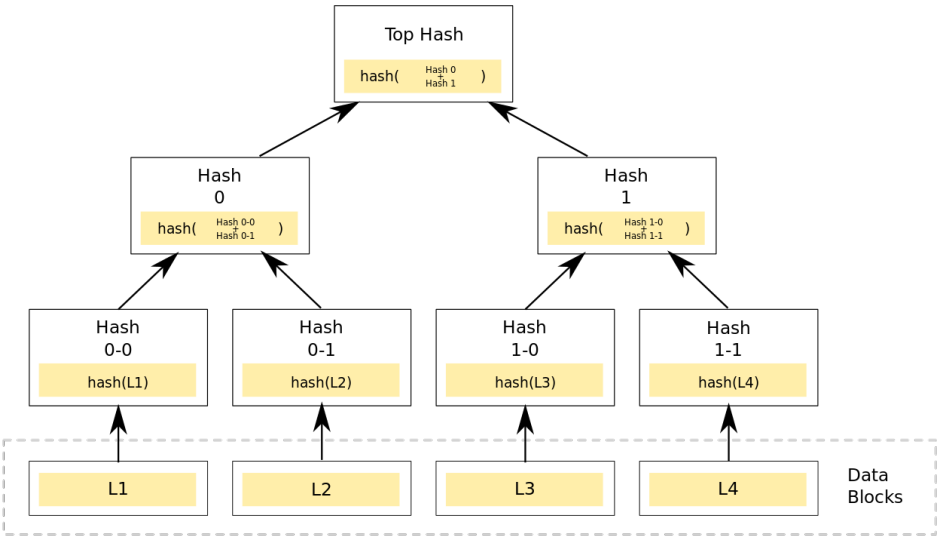

Figura 2.8: Ejemplo de un árbol de Merkle

2.2.4.2. IPLD

IPLD (InterPlanetary Linked Data)[17] es un estructura basada en un DAG de Merkle que permite enlazar datos entre diferentes sistemas distribuidos, como IPFS, Bitcoin o Ethereum. Una estructura de datos inmutable es aquella que no puede ser modificada una vez creada, lo que ofrece ventajas como la seguridad, la consistencia y la ausencia de efectos secundarios. IPLD utiliza estructuras de datos inmutables para representar los bloques de datos que se almacenan y se enlazan entre sí mediante CIDs.

IPLD es una capa de abstracción que permite a los desarrolladores trabajar con datos en diferentes plataformas y protocolos como si estuvieran trabajando con un solo sistema cohesivo. Permite la interoperabilidad a gran escala entre diferentes sistemas de almacenamiento de datos, lo que facilita la creación de aplicaciones y servicios más robustos y resistentes en entornos distribuidos.

El DAG también garantiza a IPFS con característica de control de versiones como Git. Aunque este es un apartado en el que se profundiza poco en IPFS. En el whitepaper, Benet se refiere más bien al hecho de que al igual que Git, IPFS usa Merkle DAGs, y por lo tanto posee características de control de versiones. Los nodos en un Merkle DAG son inmutables. Cualquier cambio en un nodo alteraría su identificador y, por lo tanto, afectaría a todos los ascendentes en el DAG, creando esencialmente un DAG diferente.

En lo que respecta a este trabajo, este hecho permite la propia existencia de OrbitDB, el cual es una pieza clave en el desarrollo de este proyecto. OrbitDB implementa bases de datos distrubuidas en IPFS, se basa en este concepto al ser los datos inmutables y permitir la reconstrucción de la base de datos mediante el intercambio de objetos y la actualización de referencias remotas. TODO poner referencia a orbitdb

2.2.4.3. Códecs de IPLD

Los códecs de IPLD son funciones que transforman el modelo de datos de IPLD en bytes serializados para que puedas enviar y compartir datos, y transforman los bytes serializados de nuevo en el modelo de datos de IPLD para que puedas trabajar con él. Algunos de estos códecs incluyen:

- **DAG-CBOR:** es un formato binario que soporta el modelo de datos de IPLD al completo. Ofrece un excelente rendimiento y es adecuado para cualquier tipo de trabajo.
- **DAG-JSON:** está basado en JSON (Javascript Object Notation). Es un formato más legible, lo que lo hace muy conveniente para la interoperabilidad, el desarrollo y a la hora de depurar código que haga uso de IPLD.
- **DAG-PB:** es otro formato binario usado principalmente para serializar datos en fornmato de unixfsv1 (ir a 2.2.4.4: 'Unixfs' para más información).
- **DAG-JOSE:** El codec dag-jose es un formato para firmar y cifrar objetos JSON.

En este proyecto se ha hecho especial incapié en el uso de DAG-JOSE como códec de los datos en el DAG de Merkle. Esto permite la firma y cifrado de en objetos JWS (JSON Web Signatures) y JWE (JSON Web Encryption) representados en nodos del DAG de IPLD. TODO poner referencia a la parte de JWE y JWS

2.2.4.4. Unixfs

Sobre todo este modelo de datos se establecen otras estructura o abstracciones como Unixfs.

Cuando se agrega un archivo a IPFS, puede que sea demasiado grande para caber en un solo bloque, por lo que se divide en disintos bloques que luego son representados mediante metadatos en una lista de enlaces a estos bloques. UnixFS es un formato usado para describir archivos, directorios y enlaces simbólicos en IPFS. Este formato de datos se usa para representar archivos y todos sus enlaces y metadatos en IPFS. UnixFS crea un bloque (o un árbol de bloques) de objetos enlazados.

```
$ ipfs add Pictures/Wallpapers/ -r
added QmT4pCxAKwjKz58FRkKGZUkuk9dBogzvSqfKaP9aXxRsLh Wallpapers/flow 1.jpg
added QmY1KMKAoHmDK9V3woMGGarVUJ4a1a4HCW1CxRXMtC9KwP Wallpapers/flow 2.jpg

added QmYmdHatxVXfSy9Gjdp8e75cL2mvsGyZ1tT6Qo3ijUV2h5 Wallpapers
```

Al obervar el contenido del CID final de la carpeta subida en el DAG

```
$ ipfs dag get /ipfs/QmYmdHatxVXfSy9Gjdp8e75cL2mvsGyZ1tT6Qo3ijUV2h5 | jq
```

```
{
  "Data": {
    "/": {
      "bytes": "CAE"
    }
  },
```

```
  "Links": [
      {
      "Hash": {
        "/": "QmT4pCxAKwjKz58FRkKGZUkuk9dBogzvSqfKaP9aXxRsLh"
      },
      "Name": "flow 1.jpg",
      "Tsize": 50087191
    },
    {
      "Hash": {
        "/": "QmVRXnTfUDxioWNZ5FbA79xuiZGtkUuxL6rae1MtstGuzu"
      },
      "Name": "flow 2.jpg",
      "Tsize": 32032025
    },
  ]
}
```


Como se puede observar el DAG contiene la estructura de datos que modela un directorio, en este caso de tipo directorio.

```
$ ipfs files stat /ipfs/QmYmdHatxVXfSy9Gjdp8e75cL2mvsGyZ1tT6Qo3ijUV2h5
Size: 0
CumulativeSize: 549538133
ChildBlocks: 22
Type: directory
```

En cambio si se observa el contenido en el DAG de uno de los archivos enlazados:


```
$ ipfs dag get QmT4pCxAKwjKz58FRkKGZUkuk9dBogzvSqfKaP9aXxRsLh | jq
{
  "Data": {
    "/": {
      "bytes": "CAIYnKzwFyCAgOAVIJyskAI"
    }
  },
  "Links": [
    {
      "Hash": {
        "/": "QmbhvAYKs4ERhnvujqDdTVCZdV2Y8UqrnFjGTVver1rzFL"
      },
      "Name": "",
      "Tsize": 45623854
    },
    {
      "Hash": {
        "/": "QmRw866oefrQ98iYb71cvKMmHTLsoTUW5j2cmRR5kkRqLf"
      },
      "Name": "",
      "Tsize": 4463228
    }
  ]
}
```

El archivo ocupa dos bloques cuyos CIDs están en la lista de enlaces del objeto.

```
$ ipfs files stat /ipfs/QmT4pCxAKwjKz58FRkKGZUkuk9dBogzvSqfKaP9aXxRsLh
QmT4pCxAKwjKz58FRkKGZUkuk9dBogzvSqfKaP9aXxRsLh
Size: 50075164
CumulativeSize: 50087191
ChildBlocks: 2
Type: fil
```

2.2.4.5. MFS

Sobre Unixfs existe otra capa más que es la que realmente permite una interacción con IPFS como si de un sistema de ficheros tradicional se tratara. Este componente es denominado *Mutable File System* (MFS) o Sistema de Archivos Mutable. Para hacer esto posible, MFS mantiene un mapa de la estructura de archivos y directorios en IPFS. Cada vez que se realiza una operación en MFS, como crear o mover un archivo, se actualiza este mapa. Sin embargo, los datos subyacentes en IPFS permanecen inalterados. Esto significa que se pueden cambiar la estructura y organización de archivos y directorios en MFS sin tener que copiar o mover los datos reales, manipulando enlaces dentro del DAG.

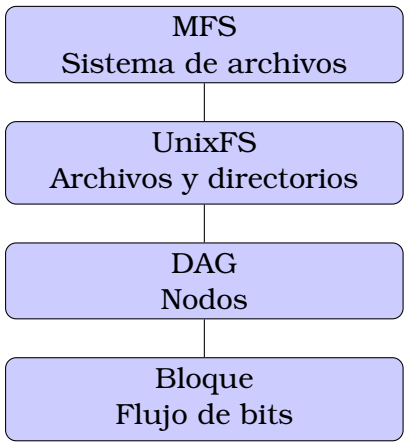

Figura 2.9: Capas de abstracción sobre los datos en IPFS

2.2.5. Sistema de nombres

IPNS, o Sistema de Nombres Interplanetario, es un componente fundamental de IPFS que permite la creación de nombres persistentes para diferentes nodos en la red IPFS. Dado que los contenidos en IPFS son inmutables y se accede a ellos a través de sus CIDs, cualquier cambio en el contenido dará lugar a un nuevo CID. Esto puede resultar inconveniente para los usuarios que necesitan referirse a un contenido específico, incluso si este cambia con el tiempo. Aquí es donde entra en juego IPNS.

IPNS proporciona una capa de indirección que permite a los usuarios referirse a contenidos que pueden cambiar con el tiempo usando un nombre persistente. En lugar de tener que actualizar el CID cada vez que cambia el contenido, los usuarios

pueden referirse al contenido usando un identificador IPNS. Este identificador es una clave criptográfica que se genera cuando se inicializa un nodo IPFS, y es única para cada nodo.

Cuando se desea publicar contenido bajo IPNS, se crea un registro que vincula el identificador IPNS con el CID del contenido. Este registro se firma con la clave privada del nodo, garantizando que solo el propietario del identificador IPNS puede cambiar la vinculación. El registro, contenido en la DHT se propaga luego a través de la red IPFS mediante Kademlia. Cuando otros nodos quieren acceder al contenido, pueden buscar el registro usando el identificador IPNS y obtener el CID correspondiente.

IPNS es de interés particular para este proyecto debido a que permite crear enlaces persistentes a contenido que puede cambiar con el tiempo. Tal y como sucede con una URL que enlaza a un contenido compartido en plataformas de almacenamiento en la nube como Google Drive o Dropbox.

Este es el último apartado en esta explicación de los fundamentos de IPFS. Por supuesto, no se ha cubierto todo lo que IPFS ofrece, pero sí los conceptos fundamentales necesarios para entender el sistema que se propone en el proyecto.

2.3. Ecosistema en torno a IPFS

2.3.1. Introducción

El ecosistema IPFS es un conjunto diverso y creciente de proyectos, herramientas, comunidades e integraciones que trabajan colectivamente para desarrollar, adoptar y evolucionar IPFS. Este ecosistema es fundamental para el éxito y la adopción generalizada de IPFS, ya que proporciona una variedad de recursos y oportunidades para interactuar y construir sobre el mismo.

2.3.2. Proyectos basados en IPFS

Existen numerosos proyectos y productos que se basan en IPFS para ofrecer soluciones innovadoras y disruptivas en diferentes áreas de aplicación. Algunas de estas áreas son:

- Almacenamiento: proyectos que utilizan IPFS para proporcionar servicios de almacenamiento distribuido, persistente y rentable, como Filecoin, Sia, Storj o Textile.
- Alojamiento web: proyectos que utilizan IPFS para alojar sitios web estáticos o dinámicos sin depender de servidores centralizados, como Fleek, Pinata o Unstoppable Domains.
- Distribución de contenido: proyectos que utilizan IPFS para distribuir contenido multimedia, educativo o informativo de forma eficiente y descentralizada, como Audius, Wikipedia Mirror o Origin Protocol.
- Aplicaciones descentralizadas (dApps): proyectos que utilizan IPFS para construir aplicaciones web que funcionan sobre redes peer-to-peer, sin intermedia-

rios ni puntos de fallo, como Brave, Metamask o OpenBazaar.

2.3.3. Herramientas y librerías de IPFS

Existen diversas herramientas y librerías que facilitan el desarrollo y la integración con IPFS, tanto para usuarios finales como para desarrolladores. Algunas de estas herramientas y librerías son:

- Clientes API: herramientas que permiten interactuar con un nodo IPFS a través de una interfaz de programación de aplicaciones (API), como ipfs-http-client, ipfs-js or py-ipfs-http-client.
- Interfaces de línea de comandos: herramientas que permiten interactuar con un nodo IPFS a través de una terminal o consola, como ipfs or ipfs-cluster.
- Marcos de desarrollo: herramientas que facilitan la creación y el despliegue de aplicaciones basadas en IPFS, como Fleek, Textile or 3Box.
- Librerías para integrar IPFS: herramientas que permiten integrar IPFS en otras aplicaciones o plataformas, como ipfs-embed, js-ipfs or go-ipfs.

2.3.4. Comunidades en torno a IPFS

Existen diversas comunidades que se forman alrededor de IPFS, tanto para apoyar el desarrollo y la adopción del proyecto, como para explorar sus posibilidades y beneficios. Algunas de estas comunidades son:

- Comunidades de desarrolladores: comunidades que se dedican a contribuir al código fuente, a reportar errores, a proponer mejoras o a crear nuevas funcionalidades para IPFS, como el equipo principal de IPFS, los colaboradores externos o los grupos locales.
- Comunidades de usuarios: comunidades que se dedican a utilizar IPFS para sus propios fines, a compartir experiencias, a resolver dudas o a dar feedback sobre el proyecto, como los usuarios finales, los creadores de contenido o los operadores de nodos.
- Comunidades de gobernancia: comunidades que se dedican a definir las reglas, los principios y los objetivos del proyecto IPFS, así como a coordinar las acciones y los recursos necesarios para su cumplimiento, como el Protocol Labs, la Fundación Filecoin o el Consejo Asesor.

2.3.5. Integraciones de IPFS

Existen diversas formas en las que IPFS se integra con otras tecnologías y plataformas para ampliar sus capacidades y su alcance. Algunas de estas integraciones son:

- Ethereum: una plataforma de computación descentralizada que permite la creación de contratos inteligentes y aplicaciones descentralizadas. IPFS se integra con Ethereum para almacenar y distribuir los datos asociados a estas aplicacio-

nes, así como para mejorar la escalabilidad y la eficiencia de la red.

- Filecoin: una red de almacenamiento descentralizado que permite a los usuarios alquilar o proveer espacio de disco a cambio de una criptomoneda. IPFS se integra con Filecoin para ofrecer una capa de incentivos económicos y una garantía de disponibilidad y persistencia de los datos almacenados en IPFS.

Como se puede apreciar IPFS es un proyecto con una gran comunidad y un ecosistema muy activo. Dada su naturaleza IPFS es un proyecto que se puede integrar en cantidad de ámbitos y tecnologías, ofreciendo las ventajas ya previamente mencionadas.

Capítulo 3

Estado del arte

En este capítulo se lleva a cabo un breve análisis de algunas implementaciones y plataformas que usan IPFS o tecnologías similares, así como de otras alternativas a IPFS, que cumplen parcial o completamente con los objetivos del proyecto.

Como se ha visto en el apartado anterior existen una gran cantidad de herramientas y proyectos que usan IPFS como base para sus sistemas. Pese a esto en el contexto de archivos compartidos, solo existen algunos proyectos interesantes que merece la pena analizar:

3.1. Peergos

Peergos es con una plataforma web para subir archivos y compartirlos con otros usuarios. Cuenta con un sistema de almacenamiento y comunicación descentralizado y seguro que utiliza criptografía y la red IPFS. Como se puede observar en la figura 3.1 Peergos proporciona una interfaz web para guardar, compartir y editar archivos, fotos, vídeos, mensajes y otros datos de forma privada y sin intermediarios. Peergos garantiza que solo los usuarios autorizados puedan acceder a sus datos, y que nadie pueda espiar o censurar su actividad en línea. Es una plataforma de código abierto y se puede ejecutar en cualquier dispositivo compatible con Java [18].

Este proyecto se encuentra en fase de desarrollo, aunque es un producto completo que además ofrece planes de almacenamiento como si de un proveedor en la nube se tratara. Esto es posible ya que al ser IPFS una red global, cualquier nodo de la red puede mantener *pinneados* (disponibles) los bloques que se deseen, y por tanto ofrecer un servicio de almacenamiento. En este caso Peergos ofrece un servicio de almacenamiento de pago, es decir, que tienen una serie de nodos en la red que mantienen pinneados los bloques de datos de los usuarios que pagan por el servicio.

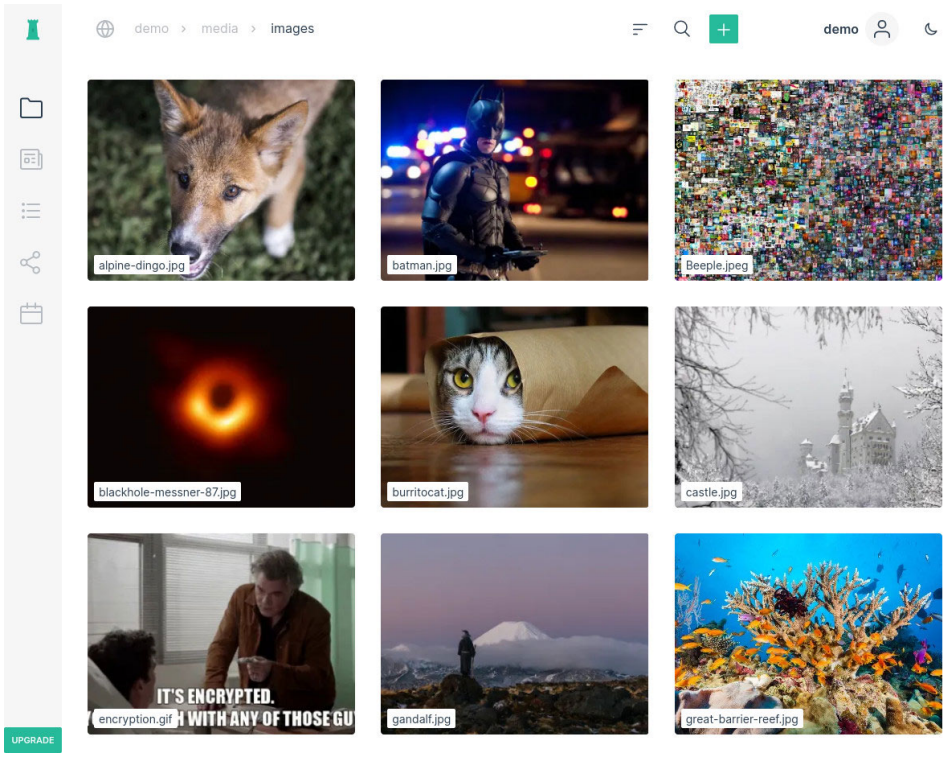

Figura 3.1: Plataforma web de Peergos

Otro aspecto de gran interés es el sistema registro de usuarios. Las claves públicas y los nombres de usuario se almacenan en una estructura de datos global de solo adición, con los nombres asignados por orden de llegada. Esto requiere consenso para garantizar la singularidad de los nombres de usuario. Aquí es también donde se almacena el ID del nodo IPFS del servidor (o servidores) responsables de sincronizar las escrituras del usuario. El problema de la implementación realizada por Peergos es que se necesita de uno o varios servidores centralizado que denominan *Corenode* que mantienen y sirven este registro para los usuarios.

Este registro es en sí la base de datos de usuarios, la autenticación se maneja mediante un sistema de clave pública-privada guardada en el navegador del usuario. Se añade una contraseña elegida por el usuario para acceder a la cuenta.

Peergos es una plataforma muy completa y que ofrece una gran cantidad de funcionalidades que escapan del alcance de este proyecto. Se puede considerar esta plataforma como referencia del potencial de IPFS, y aunque no es perfecta y tiene sus limitaciones, es un buen ejemplo de lo que se puede lograr con esta tecnología.

3.2. Filecoin

Filecoin[19] es una plataforma de almacenamiento en la nube descentralizada basada en blockchain. La plataforma utiliza su propia criptomoneda,llamada Filecoin (FIL), para facilitar e incentivar las transacciones dentro de la red.

Filecoin no está dirigido a consumidores (usuarios de a pie) ya que realmente es un mercado para proveedores de almacenamiento en la nube. Puede llegar a ser

de interés para el futuro de este proyecto ya que se podría integrar el sistema con Filecoin para ofrecer un servicio parecido a un proveedor de almacenamiento en la nube, pero con distintos proveedores que compiten entre sí para ofrecer el mejor servicio.

3.3. Sailplane

Sailplane se describe como una plataforma para *'Compartir archivos de forma colaborativa y punto a punto en el navegador'*.

Sailplane, al igual que este proyecto, usa OrbitDB como el componente central de sus sistema[1]. Como ya se ha explicado previamente, OrbitDB es una base de datos distribuida. Sailplane implementa un backend de almacenamiento conocido como un *store*. Un store en OrbitDB se refiere a una instancia de una base de datos individual que se sincroniza automáticamente con otros stores del mismo tipo mediante la red IPFS.

Sailplane ha implementado su propio store llamado *orbit-db-fsstore*[20]. Este representa un sistema de ficheros montado sobre OrbitDB que se puede sincronizar con otras instancias de la base de datos, lo que permite mantener y compartir un sistema de ficheros sincronizado entre varios usuarios. Este sistema de archivos se puede encriptar aunque no es necesario.

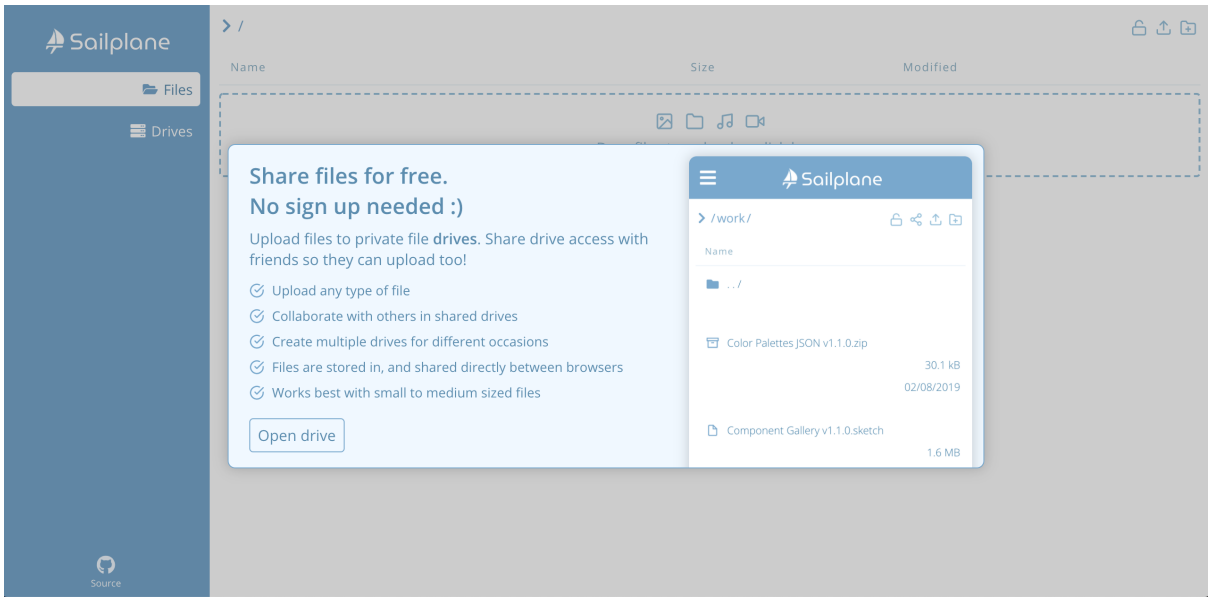

Figura 3.2: Aplicación web de Sailplane

Sailplane ofrece una implementación de un nodo IPFS personalizado llamado *sailplane-node*[21], que expone una interfaz para interactuar directamente con el store. Existe también una web que hace uso de este nodo y ofrece:

[1]Este proyecto fue descubierto hacia el final del desarrollo de este trabajo de fin de grado por lo que este hecho es más bien una casualidad.

- Un sistema de ficheros en el navegador basado en *drives* (discos virtuales) compartidos.
- Un sistema de registro automático y autocontenido. Esto sirve para que la hora de compartir un drive este pueda ver los usuarios con los que puede compartirlo.

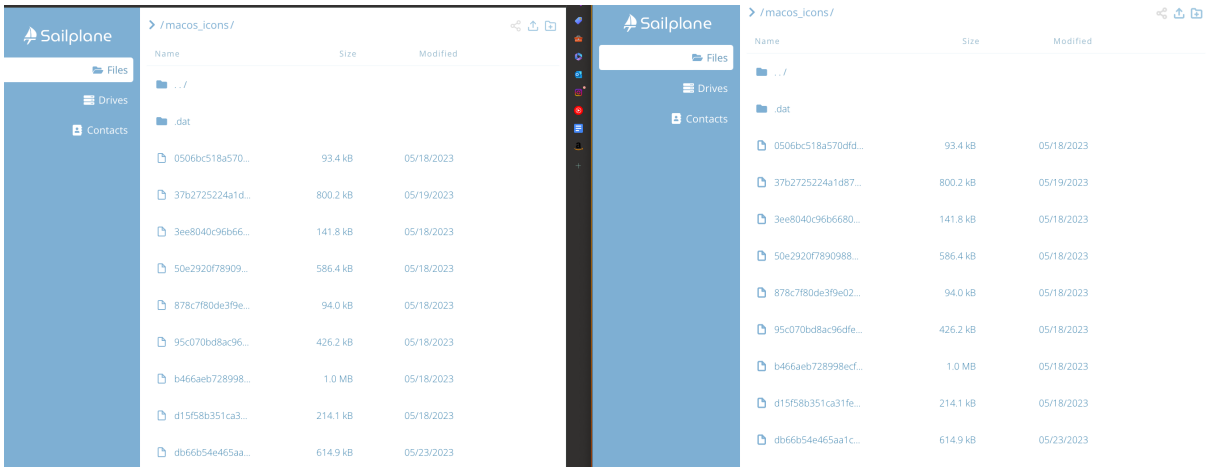

Figura 3.3: Dos nodos Sailplane sincronizando el el mismo drive

Lo que más destaca de Sailplane es la implementación de un sistema de ficheros montado sobre OrbitDB. Esto es algo que se explorará más adelante en la el capítulo 7: 'Trabajos futuros'. El uso de OrbitDB para un sistema de registro automático y autocontenido es algo que también se ha implementado en la propuesta, aunque como se ha comentado previamente, el descubrimiento de este proyecto ocurrió ya habiendo desarrollado esta característica.

3.4. Fileverse

Fileverse es una plataforma de almacenamiento que se integra con IPFS y que permite guardar, compartir y acceder a archivos desde cualquier dispositivo.

Actualmente ofrece dos productos, ambos en formato de aplicación web:

- Fileverse Solo: Una aplicación web para subir y compartir archivos sin autenticación. El usuario sube un archivo y recibe un enlace que puede compartir con otros usuarios. En la figura 3.4 se puede observar un ejemplo de uso.
- Fileverse Portals: Una aplicación web que integra autenticación basada en blockchain. Es un espacio de trabajo para la gestión de archivos sobre blochain (e IPFS) y la creación de contenido. Está en fase de pruebas y no se puede acceder.

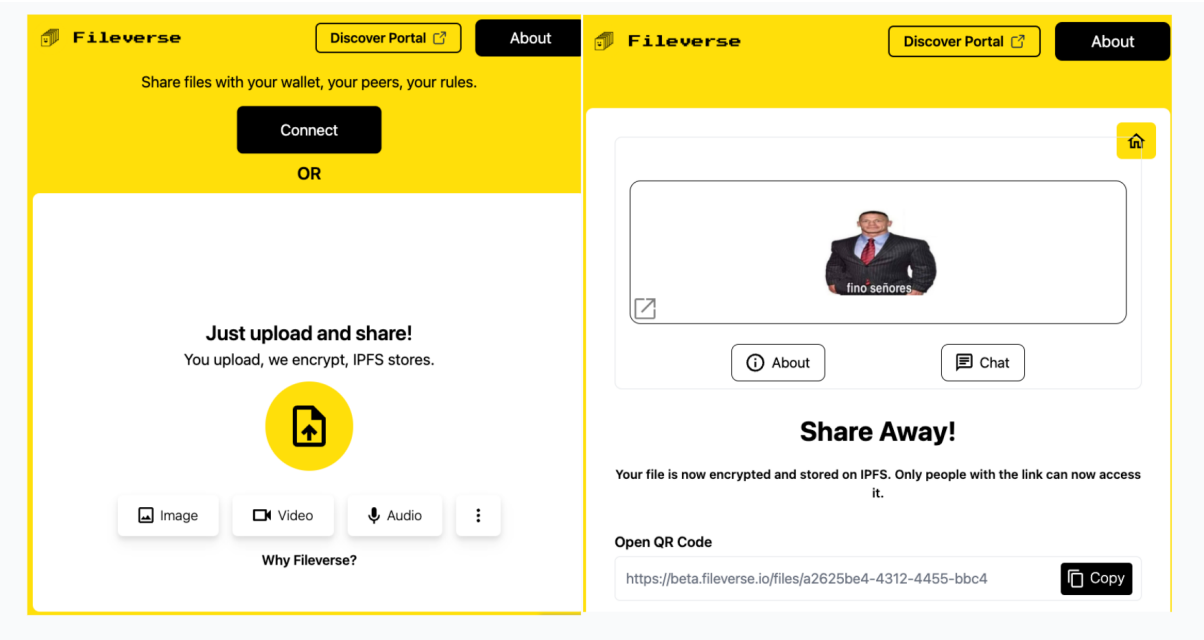

Figura 3.4: Aplicación web de Fileverse para subir archivos

El servicio que ofrece Fileverse Solo es muy simple: el usuario sube un archivo y recibe un enlace que puede pasar a otras personas para descargar dicho archivo. Esta caso de uso es muy similar al que se propone en este proyecto, aunque con algunas diferencias. El uso de una aplicación web resulta más sencillo que una línea de comandos, pero no integra cuentas de usuario ni formas de acceder o gestionar comparticiones pasadas.

3.5. Resumen

Como se puede observar en este capítulo existen varias implementaciones de sistemas de almacenamiento y compartición de archivos basadas en IPFS. Como opinión personal, ninguna presenta un servicio comparable al de los grandes proveedores centralizados de almacenamiento en la nube. Tanto en términos de facilidad de uso, como de funcionalidades y fiabilidad.

Esto se debe a que detrás de estos servicios hay una gran infraestructura e inversión que no es comparable a estos proyectos que tienen un carácter más experimental, investigativo y proviene de equipos de desarrollo mucho más limitados. Aún así, estos proyectos son un buen ejemplo de lo que se puede lograr con IPFS y han sido de gran ayuda a la hora de realizar este proyecto, en particular Sailplane y Peergos.

Capítulo 4

Desarrollo de IPFShare

4.1. Casos de uso

```
import OrbitDB from "orbit-db"
// eslint-disable-next-line @typescript-eslint/ban-ts-comment
// @ts-ignore
import AccessController from
 ↪  "orbit-db-access-controllers/interface"
import DocumentStore from "orbit-db-docstore"
import { IdentityProvider } from "orbit-db-identity-provider"

export interface RegistryEntry {
  peerId: string
  orbitdbIdentity: string // DID
  username: string // alias
}

export abstract class Registry<S, DocType> {
    abstract accessController: AccessController
    abstract store: S
    abstract open(): Promise<void>
    abstract create(): Promise<void>
    abstract replicate(): Promise<void>
    abstract close(): Promise<void>
    abstract addUser(user: DocType): Promise<void>
    abstract getUser(entryId: string): Promise<DocType | undefined>
    abstract updateUser(entryId: string, updates:
     ↪  Partial<DocType>): Promise<void>
    abstract searchUsers(queryFn: (entry: DocType) => boolean):
     ↪  Promise<DocType[]>
    abstract deleteUser(entryId: string): Promise<void>
}

```

4.2. Objetivos

4.3. Requisitos

4.4. Tecnologías

4.4.1. Tecnologías propuestas

4.4.2. Tecnologías usadas

4.5. Arquitectura del sistema

4.6. Implementación

4.6.1. Backend (Electron)

4.6.2. Frontend (React)

Capítulo 5

Resultados y conclusiones

Capítulo 6

Resultados y conclusiones

6.1. Resultados

hola

Capítulo 7

Trabajos futuros

Bibliografía

[1] "IPFS Powers the Distributed Web." [Online]. Available: https://ipfs.tech/

[2] "Web3," *Wikipedia*, Apr. 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Web3&oldid=1148197462

[3] P. Výboch, "Peer-to-peer protocols for file sharing: BitTorrent," Jun. 2017.

[4] "BitTorrent Protocol." [Online]. Available: http://www.bittorrent.org/beps/bep_0003.html

[5] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," Jul. 2014. [Online]. Available: http://arxiv.org/abs/1407.3561

[6] "Protocol Wars," *Wikipedia*, Mar. 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Protocol_Wars&oldid=1145543147

[7] B. Edwards, "The Foundation of the Internet: TCP/IP Turns 40," Sep. 2021. [Online]. Available: https://www.howtogeek.com/751880/the-foundation-of-the-internet-tcpip-turns-40/

[8] B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolf, "A Brief History of the Internet," 1999. [Online]. Available: https://arxiv.org/abs/cs/9901011

[9] E. Mori, "Peter Kirstein obituary," *The Guardian*, Feb. 2020. [Online]. Available: https://www.theguardian.com/technology/2020/feb/09/peter-kirstein-obituary

[10] "Hiperenlace," *Wikipedia, la enciclopedia libre*, Jul. 2023. [Online]. Available: https://es.wikipedia.org/w/index.php?title=Hiperenlace&oldid=152311621

[11] "Protocol Labs." [Online]. Available: https://protocol.ai/

[12] "What is libp2p." [Online]. Available: https://docs.libp2p.io/concepts/introduction/overview/

[13] P. Labs, "Libp2p Connectivity." [Online]. Available: https://connectivity.libp2p.io/microgen.vercel.app

[14] "AutoNAT." [Online]. Available: https://docs.libp2p.io/concepts/nat/autonat/

[15] "Circuit Relay." [Online]. Available: https://docs.libp2p.io/concepts/nat/

circuit-relay/

[16] "Rendezvous." [Online]. Available: https://docs.libp2p.io/concepts/discovery-routing/rendezvous/

[17] "IPLD Docs." [Online]. Available: https://ipld.io/docs/

[18] "Peergos/Peergos: A p2p, secure file storage, social network and application protocol." [Online]. Available: https://github.com/Peergos/Peergos

[19] Filecoin, "A decentralized storage network for humanity's most important information." [Online]. Available: https://filecoin.io/

[20] "Tabcat/orbit-db-fsstore: A custom orbit-db store representing a file system." [Online]. Available: https://github.com/tabcat/orbit-db-fsstore

[21] "Sailplane-node," cypsela, May 2023. [Online]. Available: https://github.com/cypsela/sailplane-node

[22] A. Das, "TCP/IP Protocol Architecture Model - How Does it Work?" Nov. 2022. [Online]. Available: https://geekflare.com/tcp-ip-protocol-architecture-model/

[23] "Protocolo de control de transmisión," *Wikipedia, la enciclopedia libre*, Feb. 2023. [Online]. Available: https://es.wikipedia.org/w/index.php?title=Protocolo_de_control_de_transmisi%C3%B3n&oldid=149440564

[24] "TCP/IP Model vs. OSI Model | Similarities and Differences." [Online]. Available: https://www.fortinet.com/resources/cyberglossary/tcp-ip-model-vs-osi-model

[25] M. Cooney, "SNA and OSI vs. TCP/IP," Oct. 2007. [Online]. Available: https://www.networkworld.com/article/2287941/sna-and-osi-vs--tcp-ip.html

[26] "Layers of OSI Model," Aug. 2017. [Online]. Available: https://www.geeksforgeeks.org/layers-of-osi-model/

Anexo

Característica	**IP/TCP**	**OSI**	**X.25**	**SNA**
Modelo	Suite de protocolos	Modelo de referencia	Protocolo de enlace	Suite de protocolos
Capas	4 (TCP/IP)	7	3	7
Año de lanzamiento	1974 (TCP) / 1981 (IP)	1984	1976	1974
Enfoque	Conmutación de paquetes	Conmutación de paquetes y circuitos	Conmutación de circuitos	Conmutación de paquetes y circuitos
Estándar	IETF	ISO	CCITT (ahora ITU-T)	IBM
Orientación	Red global	Interoperabilidad	Redes de área amplia (WAN)	Redes empresariales
Funcionalidades	Transmisión de datos, enrutamiento, control de flujo, control de congestión, conexión y desconexión	Transmisión de datos, enrutamiento, control de flujo, control de congestión, conexión y desconexión, servicios de presentación y aplicación	Transmisión de datos, control de flujo, conexión y desconexión	Transmisión de datos, enrutamiento, control de flujo, control de congestión, conexión y desconexión, servicios de presentación y aplicación
Uso en los años 90	Muy popular, base del Internet	Intento de reemplazar a TCP/IP, pero fracasó en la adopción generalizada	Utilizado en redes de área amplia (WAN), especialmente en Europa	Utilizado en redes empresariales, especialmente en sistemas mainframe de IBM

Descripción	Un modelo que se basa en la suite de protocolos TCP/IP para transmitir datos por Internet. El modelo es más simple y flexible que el modelo OSI y se usa ampliamente en la actualidad.	Un modelo que se basa en la suite de protocolos OSI para estandarizar la comunicación entre sistemas abiertos. El modelo segmenta múltiples funciones que el modelo IPTCP agrupa en capas únicas y define los servicios e interfaces para cada capa.	Un modelo que se basa en la suite de protocolos X.25 para proporcionar una conexión virtual entre terminales y computadoras a través de una red pública de conmutación de paquetes. El modelo fue uno de los primeros en ofrecer una comunicación confiable entre dispositivos remotos, pero ha sido reemplazado por tecnologías más rápidas y eficientes como Frame Relay e IP.	Un modelo que se basa en la suite de protocolos SNA para integrar los recursos informáticos distribuidos en una red jerárquica. El modelo fue desarrollado por IBM para conectar sus sistemas mainframe y periféricos, pero ha perdido popularidad frente a los modelos basados en IP.

Cuadro 1: Comparación de IP/TCP, OSI, X.25 y SNA en los años 90. Fuentes: [22], [23], [6], [24] [25] [26]