DSP

DSP Definition

The DSP (Distributed Storage Protocol) protocol is a new-generation Internet protocol paradigm based on data file encryption, distribution, storage, sharing and other dimensions. The aim of the DSP protocol is to become the core infrastructure of Web3.0.

Why Build DSP Protocol

The history of the rise and development of human civilization is a history of ever-expanding cooperation. Even today, it is difficult to imagine any human problem that cannot be solved through more cooperation—cooperation between people of different political factions, different languages, and completely different ideologies. Many of the achievements we have made as a species, such as the Internet or space travel are all done through global cooperation. The Internet is one of our earliest global cooperation systems. In this system, trust is not primarily based on institutions. At least not at first. Today's HTTP-based protocol of the Internet, becoming a basis for institutions, organizations, companies, and even countries under the framework of trust to run the online world. But this kind of centralized trust system gradually began to reveal a lot of problems.

1. **Global IOT smart devices cannot be truly interconnected**

Today's IOT environment is almost a fragmented closed system. The Wide-area Internet of Things and the local-area Internet of Things cannot interact. It is difficult to connect IT networks based on privatized industrial systems. However, the data of the Internet of Things often requires high consistency and security, this is a problem that any technology under a centralized system is difficult to solve. From the day the IOT was born, it was fragmented. The result is that shared devices of different brands need to be opened with different applications. These nodes are not equivalent, and there is no standard, and they are fragmented. Devices of the same type cannot communicate, let alone different devices.

2. Low network transmission efficiency

In the current mode mainly based on centralized servers or service clusters, when a hot resource appears in the network or when a large number of terminals access the same file resource at certain times, there will be network congestion and cannot access files effectively. HTTP such relying on the backbone network for centralized resource access will also bring high server operating costs and broadband costs, which means that a large number of branch network broadband is wasted. In addition to the traditional asymmetry of the bandwidth provided by your ISP often have to upload and download the phenomenon exists.

3. Data redundancy under split data storage

At present, in the case where the major Internet platforms have built or rented independent storage centers, countless centralized databases are scattered around the world, but because the data network is not interoperable, the same file content is stored. For example, the same movie is authorized to be played on different video platforms, which means that the same video file will exist in the storage center of different video platforms. At the same time, for the optimization of user access speed in different regions, the same file is even Multi-regional storage will be performed under the same platform. In this way, a large amount of redundant data is formed in the entire network because of the independence of the storage network, and the invisible increase also brings huge storage costs.

4. **Data ownership issues**

In today's Internet technology, for individual users, in fact, our data, documents, and digital identities are not in our own hands, but all are in the hands of Internet oligarchs. When we need our identity and our data files, we need to apply to these giants and institutions through cumbersome procedures, in other words, our data does not belong to itself, but belongs to the data technology giant. Any of our operations on data requires authorization from a centralized organization. Even, These data are actually allowed to use them to monopolize their empire.

5. **Data loss**

Two thousand years ago, the Alexandria Library was burned down, and thousands of precious documents in history have permanently disappeared from the world. Everyone believes that this is a human tragedy. However, this kind of thing happens on the Internet every day. According to statistics, the average life cycle of a website is about 100 days, and 2% of the network links will last forever. When you think about how important the web is in culture and information today, you realize that these numbers are devastating, such as the Internet Archive Foundation, a non-profit organization is trying to back up the entire web, but in fact, according to the current web development and hourly rate , This is an impossible task.

6. File index failure

At present, based on the HTTP protocol, finding a certain data in the network is based on the location of the file, rather than identifying the data by content. This limitation means that if the file location is only remembered, once the file is moved, the file cannot be found again, even if the file still exists on the network.

A recent Harvard study shows that all hyperlinks US Supreme Court opinion cited in 49% no longer worked. The opinion points to a location where the correct content was saved at some time in the past, but the content is no longer available at that location. Such data storage is fragile and inefficient.

7. Data Leakage

As early as 2017, "The Economist" announced that the data is "new oil". Nowadays, people are competing to obtain as much data as possible on the Internet, and at the same time, the user's data can bring better advertising. Unfortunately, these data are stored on these centralized servers, and these servers are the favorite targets of hackers. Even if companies like Facebook and Google hire the best engineers in the world, the leakage incident has happened many times. Even companies like Facebook and Google that hire the best engineers in the world have had multiple data leakage. Once you start storing data in a centralized server, there is an incentive to steal the data, and every system can be cracked.

8. Monitored and over-censored

In fact, many companies even include the government-built network environment and applications. The information we browse and store on the Internet is actually forced to be monitored and filtered, and we cannot even access most of the content we need to access. Any browsing behavior In the centralized record storage and even judgment, the enterprise will carry out profit-driven push behavior through personal data. This is actually a very worrying direction.

9. Internet link status is unstable

In the current network world, a link that is abnormal due to a single point of failure can easily cause abnormal or slow data transmission from one node to another, and once it occurs, in the case of non-human intervention, almost

impossible to recover. For example, the Tesla server had a problem for a short time, which caused the owners of the world to rely on the app to open the vehicle at the same time to be unable to open and drive the vehicle and caused some social problems.

10. Developer is not friendly

Due to the central differentiation of traditional Internet organizations, in the traditional network protocol, to build applications like Uber, you need to install many components, such as payment, storage, and identity management. When another company builds an application like DiDi, it also needs to rebuild these components, a lot of repetitive work..

With the emergence of these problems, we are also approaching the limits of this trust system. In fact, we need to start helping the Internet to transfer the trust, from trust in institutions to trust in technology.

It is gratifying that these problems have gradually been noticed and people have begun to try to solve these problems, such as the IPFS protocol project. However, from the current situation, there are still a lot of problems still to be solved in this project. E.g.:

1. **IPFS lacks end-to-end encryption.** All browsing activities that occur in IPFS are public to the operator. When and what content IPFS users see and publish, they can be fully recorded. If the government wants to list the IP addresses of the readers of an article, it is not difficult to introduce the specific

people behind these addresses, and at the same time such data will become a stolen data for everyone.

2. I **PFS does not design a data persistence guarantee mechanism.** IPFS cannot automatically push data to avoid things like pushing pornographic images to women or religious taboo content to inappropriate people. But if this is encrypted data, it is a ciphertext of random numbers that can't see the content, so there is no such problem. IPFS is not encrypted and is directly related to IPFS's lack of ciphertext deduplication technology. Without this technology, IPFS can only choose between encryption and deduplication. IPFS chose deduplication and sacrificed encryption, which led to the lack of IPFS data persistence guarantee mechanism.

3. **IPFS lacks a content deletion mechanism.** Each IPFS user must manually check the content and assume full legal risk. Even if an article is found to be defamatory or infringing through legal procedures, there is no mechanism for the node to automatically offline content Instead, each node needs to be manually deleted. If it is not actively deleted, it will continue to spread the content until it is cleaned up. The IPFS node stores downloaded content by default. If a person reads illegal content, that person will also provide illegal content to other nodes, that is, from a consumer of illegal content to a communicator.

4. **IPFS lacks end-to-end sharing capabilities.** In essence, it is still client-to-server transmission. It does not implement the direct data distribution

and sharing capabilities between clients through the protocol itself. In this way, for a light node such as a client, data storage, distribution, and sharing operations cannot be achieved.

5. **IPFS lacks shared permission control.** For files in the IPFS network, any file has the same access rights to everyone and is currently completely open. The same file cannot be controlled for different users by different access rights, which greatly limits many content distributions and sharing scenarios.

DSP Protocol Vision

What we need is a new set of free and open network protocols - based on this set of protocols, people can achieve true data freedom and freedom of interconnection. At the same time, it is necessary to form a user data file sharing library (database) under a free network. This database is neutral and jointly owned by the users themselves. This means that no human institution can control it. Of course, building such a database is not easy, because someone has to put in a lot of work to maintain its authenticity. Traditionally, we have relied on intermediaries like technology giants, as well as banks and governments to ensure database security. And now, what we want is database trust based on mathematics and physics. In other words, the database must be self-monitoring, self-operating, and self-managing. We should not rely on Google, Facebook to protect us from harm

or data such as how to store or how to use. All of this should return to the individual. The storage, distribution, sharing and authorization of data can be completely controlled by the individual. Only under such a network protocol environment can we have true data freedom. As an open and free network protocol, we need any network-capable device to join the network through the new protocol and become a data point in the network. Not only those devices with special configuration or special authorization, all devices share the same network environment, and the data of all devices are interwoven in the same network.

DSP Lab hopes to accelerate the creation of a distributed, decentralized, and interconnected parallel world network to create a higher-performance, more economical, safer, and more open data sharing transmission protocol and file storage network. As the foundation of the next generation Internet, it provides fast, free and safe interaction of information and data for the entire network, reduces the cost of data transmission and storage, and accelerates the revolution in the era of information freedom.

The Changes that DSP will bring

1. **Internet of Everything**

   DSP will provide a standardized network protocol interface for external network devices. Under the standardized connection based on the DSP- based

network protocol, different devices can be connected to the same network with a standardized protocol, and connect to the network through a standardized protocol interface for data transmission, sharing and storage. And all access devices will have their own identity in the network, so that different types, different companies, different attributes of the equipment will form the interconnection of data, and ultimately form a truly interconnected network world.

2. Copyright Control, Data Distribution

DSP can provide a complete copyright control agreement. Through the copyright control agreement, the same file can be distributed to different copyright holders according to the copyright attribution, so that users on the entire network actually share the same file body, It avoids the need for different copyright holders to maintain the content of a file separately, and avoids the redundancy of data storage to the greatest extent while saving a lot of storage costs..

3. The ownership of the file data is returned to the user

In the network world based on the DSP protocol, all data and files do not belong to any centralized organization, and are completely controlled by the user's own identity. Even with the participation of protocols such as encrypted sharding, even nodes that store file data cannot have the data

itself. The goal of DSP is to completely return the user's own data usage rights to the user. If at any time you want to delete a certain data, it will physically deleted from the network. If you want to authorize certain people to view certain data, you can only view it after being authorized. Users have full control over the data.

4. Personal data cannot be leaked

As a decentralized storage protocol, DSP not only stores file data on different nodes, but at the same time, the same file will also be broken up by sharding and encryption, which means that even if you get all the data of a node, It is also impossible to have a complete document. The nesting of multiple encryption methods fundamentally eliminates the possibility of data leakage.

5. Prevent personal behavior data from being monitored and over-reviewed

In the DSP protocol, all behavior data generated by individuals and file data uploaded and downloaded are targeted for encryption processing, and even path obfuscation processing can be performed through routing protocols, making it impossible to trace the specific sender and receiver. It guarantees the absolute security of personal privacy to the greatest extent and avoids the possibility of being used by the third party.

6. Make sure the data persists

In the network environment based on the DSP protocol, all data are backed up in different nodes according to the protocol. This kind of backup can be

reasonably controlled, and the deduplication operation can be performed under the condition of ensuring data security to reduce the storage network redundancy, or you can choose to like the Zero Network, selectively maximize the number of backups and nodes. Maximize the security of extremely important data. At the same time, all backups will dynamically synchronize the content, while ensuring the security of the data while maintaining the timeliness and consistency of the content. When an external accident causes a backup to fail and cannot be accessed normally, there will be new nodes to perform new file backups. Such a protocol can ensure the permanent existence of files and data in the network. In addition, by changing the addressing mode of the content in the network, from address-based addressing to content-based addressing, this will completely avoid the problem of unable to locate the network file due to the failure of the file address.

7. Network link-like intelligent self-optimization

In the network supported by the DSP protocol, the transmission status between all nodes of the entire network should can be captured in real time and analyzed in real time. The link status of the entire network will be completely transparent and controlled. Through analysis and intelligent adjustment, it can dynamically change the direct path of some nodes to avoid congested or faulty links and use the normal path. This adjustment is an automatic behavior of the network and occurs in real time throughout the network. Through such a protocol, communication and transmission between

all nodes of the entire network can be ensured in an optimized state. The availability of the network will be maximized.

8. Will be more developer friendly

Under the new protocol framework, distributed applications are gradually being created, and various business components, such as identity, storage, payment, etc., will be used. The components of these business modules will be developed and constructed by different organizations and individuals. At the same time, the components will be contributed and placed on the network for reuse, so that each component only needs to be built once, which greatly reduces redundant development work. Will help developers focus more on the business itself.

We see this as the foundation for innovation in the next world, we cannot imagine when the dust settles, the influx of a new wave of innovation will happen in coming, But we've seen the beginning of a process, starting with some technologies that are also working, for example, block chain, edge computation, IOT, zero net, onions protocol， etc. DSP will be a basic new paradigm for computing networks, and the key feature of this new paradigm is never mentioned before, which is the trusted storage, programmable data storage distribution network protocol. All data will no longer flow from institutions to individuals from the bottom up, but will begin to flow from individuals and clients to the entire network from the bottom up.